



Zum Einsatz des Deutschen Sprachdiploms der Kultusministerkonferenz (erste Stufe) an Schulen in Deutschland: Ein Problemaufriss

Katrin Wisniewski, Robin Ide, Matthias Schwendemann

Abstract: Neuerdings kommt das für Deutschlernende an Schulen im Ausland entwickelte *Deutsche Sprachdiplom* (DSD) der Kultusministerkonferenz (erste Stufe) auch im Inland zum Einsatz. Allerdings wurde die Validität dieser Testverwendung noch nicht nachgewiesen. Basierend auf einem verbreiteten Validierungsmodell diskutiert dieser Artikel Potenziale und Herausforderungen des DSD-Einsatzes im Inland. Es wird gezeigt, dass positive Konsequenzen erhofft werden, deren Nachweis jedoch noch aussteht; nicht intendierte Folgen werden kaum thematisiert. Ferner zeigt sich eine große Heterogenität der Einsatzzwecke. Der Beitrag bespricht dann das im Spannungsfeld von kommunikativer und bildungssprachlicher Kompetenz angesiedelte Testkonstrukt und diskutiert die Rolle des *Gemeinsamen europäischen Referenzrahmens*. Modelltestanalysen zeigen ferner, dass Testinhalte teils ethisch problematisch sind. Insgesamt verweist der Beitrag auf den Bedarf einer umfassenden Validierungsanstrengung für die neue Verwendung des DSD I.

Recently, the *Deutsches Sprachdiplom* (DSD) I, a language test originally developed for learners of German at schools outside of Germany, has increasingly been used for newly immigrated pupils inside Germany. However, almost no validity evidence justifying this novel test use is available. Based on a well-proven *assessment use argument*, this contribution delineates potentials and possible drawbacks of this test use. For example, we show that while beneficial consequences are expected, there is a lack of evidence proving these effects; unintended consequences have not yet received any attention. Furthermore, the DSD I is being used for a wide range of heterogeneous test purposes. In summary, the contribution recommends a considerable validation effort for the new DSD I use.

Schlagwörter: Deutsches Sprachdiplom I, Validierung/Validität, neu Zugewanderte, Sprachdiagnostik, Sprachtest; Deutsches Sprachdiplom I, validation/validity, newly migrated pupils, language assessment, language testing

1 Einleitung

Schulgebundene sprachdiagnostische¹ Verfahren für Jugendliche, die erst nach dem Schuleintrittsalter nach Deutschland gekommen sind (sogenannte „neu Zugewanderte“ oder auch, spezifischer, „Seiteneinsteiger*innen²“) und relativ spät das Deutsche als Zweitsprache (DaZ) erwerben, sind rar. In Ermangelung passgenauer, d.h. für die Zielgruppe, zielgruppenspezifische Testzwecke und -konstrukte entwickelter standardisierter Testverfahren kommt seit einigen Jahren vermehrt das *Deutsche Sprachdiplom der Kultusministerkonferenz – Erste Stufe (DSD I)* (im Folgenden DSD I) der Zentralstelle für Auslandsschulwesen (ZfA), das ursprünglich für Lernende des Deutschen in klassischen *Fremdsprachenlernkontexten* an Schulen im Ausland entwickelt wurde, auch im Inland zum Einsatz. Schulen in 12 Bundesländern (Stand: Mai 2020) führen das allgemeinsprachliche DSD I in unveränderter Form bei der neuen Zielgruppe zu von originalen Testzielen (s. Abschnitt 3) abweichenden Zwecken durch. Laut Angabe der Kultusministerkonferenz (KMK) nahmen allein im Jahr 2019 bereits 10.700 Schülerinnen und Schüler (SuS) an Prüfungen des Deutschen Sprachdiploms in Deutschland teil (vgl. KMK 2019).

Trotz dieser bemerkenswerten Zahlen wurde die Validität der neuen DSD I-Verwendung bislang noch nicht nachgewiesen. Dies ist ein dringendes Desiderat, denn die Zulässigkeit der Übertragung eines Testverfahrens auf andere Zielgruppen und mit anderen Testzwecken ist alles andere als selbstverständlich. Kein noch so qualitativ hochwertiger Test ist „an sich“ valide (vgl. etwa Messick 1989; Kane 2001). Validität ergibt sich vielmehr erst aus dem jeweiligen Gebrauch eines Tests und bezeichnet das Ausmaß, in dem die Folgen, die sich aus dem Testergebnis für eine/n Testkandidaten/in ergeben (z.B. die (Nicht-)Zulassung in eine bestimmte Bildungseinrichtung), gerechtfertigt werden können (Messick 1989: 13). Testvalidierung muss also als Prozess verstanden werden, der den vollständigen Testentwicklungszyklus (vgl. Bachman/Palmer 1996) inklusive der Testzwecke und der Konsequenzen von Testeinsätzen betrifft. Zu bedenken ist zudem, dass die Zwecke, zu denen

¹ Den Begriff „Sprachdiagnostik“ verwenden wir in einem allgemeinen Sinne als Leistungsmessung einigermaßen gemäß dem englischen Ausdruck „assessment“. Sprachdiagnostik umfasst dementsprechend ein breites Spektrum an Methoden (z.B. Portfolios, Sprachtests, Beobachtungen...), mit dem verschiedenste Aspekte der sprachlichen Kompetenzen bzw. des Sprachstands (vgl. Ehlich 2007: 25) von Lernenden zu verschiedenen Zwecken evaluiert werden (vgl. Lengyel 2012: 8). Damit verstehen wir Sprachdiagnostik als übergeordneten Begriff (in Abgrenzung vom „diagnostic assessment“ bei Alderson, Brunfaut und Harding 2015). Eine solche Sprachdiagnostik kann also formellen oder informellen Charakter haben. Auch geht es in unserem Beitrag um Individualdiagnostik (i.U. etwa zu PISA, vgl. Pižorn/Huhta 2016). Sprachtests (die wir synonym zu Sprachprüfungen setzen) sind eine Form von Sprachdiagnostik, bei der Leistungen in Form eines messbaren Ergebnisses in häufig numerischer Form evaluiert werden.

² Zur Einschulung keine oder fast keine Deutschkenntnisse, sehr wahrscheinlich schon Schulerfahrung vorhanden (vgl. Ahrenholz/Ohm/Brede 2017: 258).

ein Test eingesetzt werden soll sowie dessen intendierte Konsequenzen das gesamte Testdesign bestimmen und nicht umgekehrt (vgl. Pižorn/Huhta 2016: 244). Angesichts der Tatsache, dass Tests nicht nur beabsichtigte, sondern auch unbeabsichtigte Konsequenzen haben können, wird deutlich, dass für eine solche Neuverwendung eine transparente Validierung zentral ist und zwar insbesondere dann, wenn von den Testergebnissen Entscheidendes abhängt. Derart folgenreiche Testverfahren kommen in letzter Zeit allgemein in Migrationskontexten vermehrt zum Einsatz (so genannte *high stakes*-Tests, die etwa über die Berechtigung zur Wahrnehmung von Bildungsangeboten, Aufenthaltstitel o.ä. entscheiden, vgl. Bachman/Palmer 2010: 195; siehe Pochon-Berger/Lenz 2014), und auch beim Einsatz des DSD I im Inland sind offenbar mit dem Abschneiden im Test zumindest teilweise entgegen der ursprünglichen Verwendungsabsicht (KMK/ZfA, persönliche Kommunikation) gravierende Folgen verbunden (s. Abschnitt 7).

Ziel des vorliegenden Beitrags ist die Diskussion einiger Aspekte der DSD I-Verwendung in innerdeutschen Kontexten. Eine Validierung erfolgt hier jedoch nicht. Vielmehr legen wir entlang einiger ausgewählter validitätsbezogener Fragestellungen Potenziale und Herausforderungen des Einsatzes des DSD I an Schulen im Inland im Abriss dar und verweisen auf den Bedarf einer umfassenden Validierungsanstrengung. Ziel ist explizit keine Evaluation des DSD I, sondern eine Problematisierung möglicherweise überdenkenswerter validitätsbezogener Aspekte der neuen Testverwendung in Inland. Dabei nehmen wir ausschließlich auf das allgemeinsprachliche DSD I Bezug.

In drei Hintergrundkapiteln gehen wir zunächst auf die aktuelle Diagnostiklandschaft sprachlicher Kompetenzen bei neu Zugewanderten in der Sekundarstufe ein (Abschnitt 2) und beschreiben das DSD I in Grundzügen (Abschnitt 3). Im Anschluss stellen wir ein im Sprachtestbereich verbreitetes Validierungsframework vor, an dem wir uns im Folgenden grob orientieren (Bachman/Palmer 2010: *Assessment Use Argument*, AUA; Abschnitt 4). Die nachfolgenden Kapitel widmen sich Aspekten der aktuellen Verwendung des DSD I aus Validitätsperspektive. Zunächst stellen wir dazu einige Spezifika der neuen DSD I-Zielgruppen heraus (Abschnitt 5) und widmen uns dann den erwarteten und möglichen unerwarteten Konsequenzen des neuen DSD-Einsatzes (Abschnitt 6). Eng damit verknüpft ist eine Darstellung der Heterogenität der Einsatzkontexte und v.a. der Entscheidungen, die auf Grundlage der Testresultate getroffen werden (Abschnitt 7). Die drei letzten Kapitel befassen sich schließlich mit dem DSD I selbst, indem die Frage der Übertragbarkeit und Angemessenheit der Testkonstrukte (Stichwörter Bildungssprache und kommunikative Kompetenz, Abschnitt 8) sowie der Verwendung des Gemeinsamen europäischen Referenzrahmens diskutiert wird (GER, Europarat 2001, 2018;

Abschnitt 9) und schließlich anhand einer Analyse von sieben Modelltests die Problematik der Testübertragung mithilfe ausgewählter Aufgaben auf inhaltlich-thematischer Ebene gezeigt wird (Abschnitt 10).

2 Sprachdiagnostik bei „Seiteneinsteiger*innen“ in der Sekundarstufe: Überblick

Zur zielgruppenangemessenen, heterogenitätssensiblen Erfassung sprachlicher Kompetenzen bei SuS mit nicht ausschließlich deutscher Muttersprache fehlen immer noch diagnostische Verfahren (vgl. Heppt/Paetsch 2017: 130; Hoffmann/Böhme/Stanat 2017: 121; Jeuk/Lütke 2019: 164–165), auch wenn in den letzten Jahren, mit zunehmendem Augenmerk auf die Bedeutung der Bildungssprache, erhebliche Anstrengungen zur Verbesserung dieser Situation unternommen wurden. Diese zeigen sich in Ergebnissen der *Forschungsinitiative Sprachdiagnostik und Sprachförderung FISS* (2009–2016), insbesondere bei den Projekten *SPRABILON*, *BiSpra* sowie den Projekten *Teilkomponenten der Schreibkompetenzen* und *ProDi* (Ergebnisüberblick in Redder/Naumann/Tracy 2015). Die in dort entwickelten und erprobten diagnostischen Verfahren befinden sich teils in Transferphasen (z.B. *BiSpra-Aufgaben* (Universität Bamberg 2018)); darüber hinaus werden viele Tools im Forschungs- und Entwicklungsprogramm *BISS (Bildung durch Sprache und Schrift* (BISS-Trägerkonsortium 2019)) evaluiert. Für die Sekundarstufe ist hier insbesondere das Verbundvorhaben *EVA-Sek (Formative Prozessevaluation in der Sekundarstufe. Seiteneinsteiger und Sprache im Fach)* zu nennen, innerhalb dessen das DSD I eine wichtige Rolle spielt (vgl. Ahrenholz/Ohm/Ricart Brede 2017 und Ahrenholz/Birnbaum/Ohm/Ricart Brede 2018). Auch kleinere Initiativen setzen sich zunehmend insbesondere mit bildungssprachlichen Konstrukten auseinander (vgl. etwa die Beiträge auf dem Bochumer Symposium *Sprache.Testen* 2018, Drackert/Mainzer-Murrenhoff/Soltyska/Timukova im Druck).

Vor allem für Bildungsetappen nach der Primarstufe fehlen jedoch Instrumente zur zuverlässigen Einschätzung des Sprachstands (vgl. Heppt/Paetsch 2017: 118; vgl. Schuth/Heppt/Köhne/Weinert/Stanat 2015). Dringlich ist der Bedarf auch wegen der Heterogenität der Sprachaneignungsprozesse bei DaZ-SuS im Licht der jüngsten Migration nach Deutschland. Viele der sogenannten „Seiteneinsteiger*innen“ kommen erst nach dem Schuleintritt, teils erst in der Sekundarstufe, mit dem Deutschen in Kontakt und sind teils bereits (gegebenenfalls in anderen Schriftsystemen) literalisiert (vgl. Zepter 2016: 100). U.a. deshalb muss davon ausgegangen werden, dass diese auch in sich natürlich heterogene Gruppe Sprachaneignungsverläufe aufweist, die prinzipiell von den in diagnostischen Verfahren meistens avisierten in

Deutschland geborenen mehrsprachigen Kindern, etwa mit bilinguaem oder frühem DaZ-Erwerb, abweicht (vgl. Czinglar 2018 oder die Studie von Griebhaber 2018). Darüber hinaus stellt sich die Frage nach diagnostischen Bedarfen in späteren Bildungsetappen neu. Die Fachgruppe „Neu Zugewanderte“ in BISS (2018: 2) konstatiert: „Für die spezifische Zielgruppe neu zugewanderter und mehrsprachiger Schülerinnen und Schüler gibt es bis auf die „Niveaustufen DaZ“ noch keine wissenschaftlich geprüften Instrumente“.

Zu den wichtigen vorliegenden Diagnosetools für die Sekundarstufe (Überblick in Heppt/Paetsch 2017; Jeuk/Lütke 2019) zählen also die nicht normierten Beobachtungsverfahren *Niveaubeschreibungen DaZ für die Sekundarstufe I und II* (Döll/Reich 2013; Döll/Reich 2015), ferner die profilanalytisch konzipierten Verfahren *Tulpenbeet* (Reich/Roth/Gantefort 2008) und *Bumerang* (Reich/Roth/Döll 2009). Im Projekt FISA (*Förderdiagnostik sprachlicher Entwicklung im Schulalter*, Jeuk 2013) wurde das *Praxismaterial Förderdiagnostik* entwickelt, das u.a. einen Raschskalierten C-Test enthält (Junk-Deppenmeier/Jeuk 2015). Ein auf dem FISS-Projekt *Teilkomponenten der Schreibkompetenzen* beruhender standardisierter Schreibtest ist noch in der Entwicklung (Hennes/Schmidt/Zepnik/Linnemann/Jost/Becker-Mrotzek/Rietz/Schabmann 2018). In Baden-Württemberg wurde jüngst spezifisch für neu Zugewanderte das Verfahren *Potenziale2P* entwickelt, dessen Deutsch-Baustein GER-bezogen ist und die Zertifizierung eines A2-Niveaus beinhaltet. Genauere Informationen zu diesem erst Ende 2018 fertiggestellten Verfahren liegen allerdings noch nicht vor (vgl. Saleh 2019).

Diesen und anderen sprachdiagnostischen Verfahren wird insgesamt oft eine geringe Testgüte attestiert. So sind individualdiagnostische Verfahren im Schulkontext idealiter an einer großen (vgl. Neugebauer/Becker-Mrotzek 2013: 26; mindestens $N=200$), der Zielgruppe möglichst ähnlichen Bezugsnormgruppe orientiert. Praktikable Normierungen, die der Heterogenität der Zielgruppen einigermaßen gerecht werden (unter Beachtung von Faktoren wie u.a. dem Erwerbsalter, Muttersprachen oder dem Lebensalter, vgl. Lüdtke/Kallmeyer 2007), stellen jedoch eine große Herausforderung dar. In vorliegenden Verfahren wird demgegenüber oft mit gar keinen, veralteten, zu kleinen oder ungeeigneten Normen gearbeitet (vgl. Hoffmann/Böhme 2017: 144; Hoffmann et al. 2017: 121; vgl. Heppt/Paetsch 2017). Testvalidierungen werden selten publiziert. Zudem wird in der Praxis ein teils problematischer Umgang mit sprachdiagnostischen Instrumenten beobachtet (vgl. Geist 2013; Hoffmann et al. 2017; Hoffmann/Böhme 2017). Heppt/Paetsch (2017: 134) kommen zu dem Schluss, dass „eine umfassende Diagnostik schulrelevanter sprachlicher Kompetenzen [...] bislang nicht zu leisten ist.“

Vor dem Hintergrund dieses Mangels an validierten, standardisierten diagnostischen Verfahren, die auch außerhalb eines begrenzten Expert*innenkreises aussagekräftig wären, erscheint der Einsatz des DSD I an deutschen Schulen im Inland zunächst also als eine plausible Option. Es wird sich jedoch zeigen, dass dieses Vorgehen mit etlichen, verschiedene Validitätsaspekte betreffenden Herausforderungen einhergeht.

3 Das Deutsche Sprachdiplom I im Überblick

Im Folgenden wird zunächst das DSD I der KMK in seinem ursprünglichen, auf den Einsatz im Ausland zielenden Verwendungskontext im Abriss vorgestellt. Das DSD I ist eine für den schulischen Kontext konzipierte Deutschprüfung, die von Bund und Ländern gemeinsam verantwortet (KMK 2019) und seit 1974 für 14- bis 16-jährige Deutschlernende an Schulen im Ausland durchgeführt wird (vgl. Meyer-Engling 2014: 32). Die Prüfungen des DSD testen allgemeinsprachliche Fertigkeiten und „eine über den Spracherwerb hinausgehende wissenschaftspropädeutische Bildung und Studierfähigkeit“, die durch den kontinuierlichen Besuch des Unterrichts an einer Schule, die zur Durchführung des DSD berechtigt ist, erworben werden (ZfA 2018a: 3).³ Die Durchführung der Prüfungen zum DSD werden in der *Prüfungsordnung* (ZfA 2018a) und den *Ausführungsbestimmungen* zur Prüfungsordnung (ZfA 2018|2019) geregelt.⁴ Außerdem stellt die ZfA Handreichungen zur schriftlichen und zur mündlichen Kommunikation im DSD I zur Verfügung, die das Testkonstrukt, Aufgabenformate, Bewertungsverfahren und die Operationalisierung von Bewertungskriterien ausführen (vgl. ZfA 2018b; 2019).

Das DSD I orientiert sich dabei gleichzeitig an den Niveaubeschreibungen des GER (Europarat 2001, 2018) und an „innerdeutschen schulischen Standards für moderne Fremdsprachen“ (ZfA 2019: 3; s. Abschnitte 8 und 9). Die Verantwortung für die Durchführung und Entwicklung des DSD und die Auswahl von Schulen, die zur Durchführung des DSDs berechtigt sind (etwa deutsche Auslandsschulen oder Deutsch-Profil-Schulen), liegt beim zentralen Ausschuss für das Deutsche Sprachdiplom, dem sowohl Ländervertreter*innen als auch Vertreter*innen des Bundes angehören. Ziel dieses zentralen Gremiums ist die Schaffung und Sicherstellung

³ Das *Deutsche Sprachdiplom der Kultusministerkonferenz – Zweite Stufe (DSD II)* (Niveaustufen B2/C1, im Folgenden DSD II) und das DSD I Pro als berufsorientierte Variante können hier nicht berücksichtigt werden (https://www.auslandsschulwesen.de/Webs/ZfA/DE/Deutschlernen/DSD/DSD-I-PRO/dsd-i-pro_node.html, 22.08.2019).

⁴ Diese beiden Dokumente und weitere Informationen (z.B. Bewertungskriterien für die einzelnen Prüfungsteile) stehen online zur Verfügung unter: <https://www.kmk.org/service/servicebereich-auslandsschulen/informationen-zum-deutschen-sprachdiplom.html>, 8.11.2019.

einheitlicher Standards und die weltweite Vergleichbarkeit der abgelegten Prüfungen (KMK 2019b). Pädagogisch betreut und entwickelt wird das DSD von der ZfA.

Das DSD ist zentralisiert und wird im Ausland jährlich zweimal zu festgelegten Terminen durchgeführt (KMK 2019b). Im Ausland kann das DSD I nur ablegen, wer mehrere Jahre Deutschunterricht an der Schule besucht hat (600–800 Stunden, ZfA 2018|2019: 13). In Deutschland wird das DSD I je nach Integrationsmodell unterschiedlich, jedoch in der Regel zum Ende des Besuchs von Vorbereitungs-/Willkommensklassen abgelegt (s. dazu Abschnitt 7).

Das DSD I ist ein skaliertes Sprachtest, der die Niveaustufen A2/B1 des GER erfasst. Im Fall einer mit B1 bestandenen Prüfung ermöglicht es im originalen Einsatz entweder den Zugang zu einem studienvorbereitenden Studienkolleg in Deutschland oder zum (bilingualen) Sach- und Fachunterricht an der besuchten Schule im Ausland, ist dafür jedoch keine zwingende Voraussetzung (vgl. ZfA 2009: 39; zu den Einsatzzwecken des DSD I im Inland s. Abschnitt 7).

Im DSD I muss eine Reihe verschiedener Aufgabentypen wie Lückentexte, Multiple Choice- Aufgaben, Zuordnungsaufgaben oder Richtig-Falsch-Aufgaben gelöst werden. Teilnehmende werden in den Teilbereichen Leseverstehen (LV), Hörverstehen (HV), Schriftliche Kommunikation (SK) und Mündliche Kommunikation (MK) geprüft. Der LV- und der HV-Teil beinhalten je fünf Aufgaben geschlossenen bzw. halboffenen Formats (Übersicht in ZfA 2018|2019: 4–5), während als produktive Schreibleistung anhand von Leitfragen ein mitteilungsbezogener Text zu verfassen ist (ZfA 2019). Im Prüfungsteil mündliche Kommunikation wird ein Einzelgespräch zwischen Prüfer*in und Prüfling geführt, und die Testkandidat*innen halten einen Vortrag (mit anschließendem Gespräch), der über ein halbes Jahr im, aber auch außerhalb des Unterrichts vorbereitet wird (ZfA 2018b). Beide produktionsbezogenen Prüfungsteile beinhalten die Sprachhandlungen des Beschreibens, Berichtens, der Darstellung und der Begründung und erfordern das Schildern einer eigenen Meinung (ZfA 2019: 4; ZfA 2018b: 2).

In den vier Teilbereichen können jeweils 24 Punkte erreicht werden, wobei bei der Gesamtnote alle Fertigkeiten mit 25% gewichtet werden. Die Bestehensgrenzen sind variabel und werden bei der Erprobung neu erstellter Testsätze jeweils neu ermittelt (vgl. ZfA 2018a: 10). Durchgeführt werden die Prüfungen i.d.R. an den jeweiligen Schulen, die je einen Prüfungsausschuss für den Teil Mündliche Kommunikation einsetzen. Für diesen Prüfungsteil müssen die Prüfenden speziell geschult sein, wobei die Organisation von Prüferschulungen ebenfalls in den Aufgabenbereich der ZfA fällt. Meist ist die Lehrkraft, die die Teilnehmenden zuvor unterrichtet hat, ebenfalls Teil des Prüfungsausschusses. Die Teile LV, HV und SK werden zentral und elektronisch in Deutschland ausgewertet. Der Teil SK wird von

geschulten Bewerberinnen und Bewertern ausgewertet und zusätzlich testmethodisch abgesichert (vgl. ZfA 2018a: 10).

Seit 2012 wird das DSD I nach einer zweijährigen Erprobungsphase in Hamburg auch an Schulen anderer (insgesamt 12) Bundesländer im Inland eingesetzt (vgl. KMK 2019). Dabei richtet sich das DSD I in Deutschland an SuS in der sprachlichen Erstintegration, wobei davon ausgegangen wird, dass es sich bei dieser Gruppe mehrheitlich um jugendliche Geflüchtete und junge Migrant*innen handelt (zur Zielgruppe s. Abschnitt 5) (vgl. KMK 2019).

4 Validierung von Sprachtests: Ein Framework

Auch wenn der vorliegende Beitrag keine Validierungsstudie darstellt, soll im Folgenden auf grundsätzliche Modellkonzeptionen und Spannungslinien der Validierungsforschung zumindest kurz eingegangen werden. Hier haben sich Validierungsframeworks etabliert, die unterschiedliche Aspekte der Entwicklung bzw. des Einsatzes von Tests fokussieren (z.B. Bachman/Palmer 1996, 2010; Kane 2006, 2013; Mislevy/Almond/Russel 2003) und die je mehrmethodisch quantitative und qualitative argumentative Nachweise dafür zu erbringen versuchen, einen Testeinsatz zu einem bestimmten Zweck und für eine bestimmte Zielgruppe als valide zu rechtfertigen. Der vorliegende Artikel orientiert sich lose an der Struktur des Frameworks von Bachman/Palmers *assessment use argument* (AUA, Bachman/Palmer 2010).

Ein AUA (vgl. Abb. 1) beinhaltet Behauptungen (*claims*) bezüglich der Validität von Sprachtests, die in einer argumentativen Kette mithilfe verschiedener Daten und Nachweise aufeinander bezogen und gestützt werden müssen (durch Schlussregeln (*warrants*) sowie möglichst Stützungen (*backing*) und gegebenenfalls Ausnahmebedingungen (*rebuttals*)). Alle „klassischen“ Testgütekriterien wie z.B. Reliabilitätsaspekte fallen mit in ein AUA. Ein wichtiger Anspruch des AUA sind die Transparenz der Argumentation und die Notwendigkeit empirischer Nachweise. Beispielhaft wird das AUA im Bericht von Pochon-Berger/Lenz (2014) angewandt, der eine Bestandsaufnahme der Testpraxis und der diesbezüglichen Validierungsbemühungen zu integrations- bzw. migrationsbezogenen Sprachtests zur Verfügung stellt, dabei aber nicht auf schulische Kontexte rekurriert.

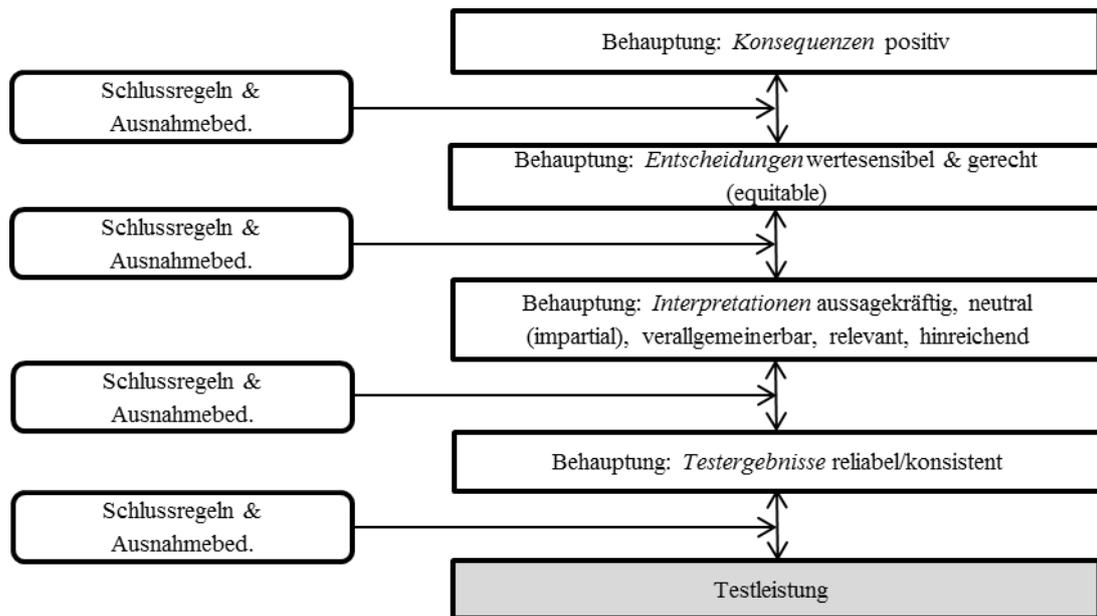


Abb. 1: AUA-Grundstruktur, nach Bachman/Palmer 2010

Wichtige Bausteine des AUA (vgl. Abb. 1) sind die beabsichtigten *Konsequenzen*, die ein Test auf alle involvierten Personen, hauptsächlich die Testkandidat*innen, hat: Diese sollten positiv (*beneficial*) sein (vgl. Bachman/Palmer 2010:178–192; Abb. 1; s. Abschnitt 6). *Entscheidungen*, die aus Testergebnissen resultieren (z.B. eine sprachliche Zulassung zum Hochschulstudium), müssen werteesensibel und gerecht sein (vgl. Bachman/Palmer 2010: 111; s. Abschnitt 7). Dies ist möglich, wenn die *Interpretationen* der getesteten Sprachkompetenzen bestimmte Bedingungen (Aussagekraft, Verallgemeinerbarkeit, Relevanz, Suffizienz, Neutralität (*impartiality*)), erfüllen, die vor allem die Konstruktvalidität betreffen (vgl. Bachman/Palmer 2010: 114; s. Abschnitte 9 und 10). Hier geht es darum, was über die getesteten Kompetenzen (das Konstrukt) bekannt sein muss, um die jeweiligen Entscheidungen treffen zu können. Die *Testergebnisse* selbst müssen reliabel sein (vgl. Bachman/Palmer 2010: 124). Im AUA wird Validität also sozusagen „von hinten“ aufgezügelt, nämlich von den beabsichtigten Konsequenzen über die getroffenen Entscheidungen zu den Interpretationen von Testergebnissen und schließlich zu letzteren selbst.

Bislang kann jedoch im Sprachtestbereich in der Realität insgesamt eine kaum ausgeprägte „**Validierungskultur**“ beobachtet werden (Pochon-Berger/Lenz 2014: 34, vgl. ähnlich zum Beispiel auch Chapelle 2011; Pižorn/Huhta 2016). Zu vielen weit verbreiteten Sprachtests (nicht nur des Deutschen) liegt nur eine spärliche diesbezügliche, öffentlich zugängliche Dokumentation vor. Erschwerend hinzu kommt eine prinzipiell unklare **Verantwortung** für Testvalidierungen. Pižorn/Huhta (2016: 244) führen an, dass Sprachtestentwickler*innen nur selten überhaupt in die Verlegenheit kämen, gegenüber Entscheidungsträger*innen aus nichtsprachlichen

Bereichen die Validität eines Tests rechtfertigen zu müssen. Unklar scheinen also bereits die Adressat*innen/Stakeholder von Sprachtestvalidierungen. Die Verantwortung für die Durchführung von Validierungen kann prinzipiell den testentwickelnden Institutionen/Firmen oder den testnutzenden (schulischen) Institutionen zugeschrieben werden. Oft erstellen jedoch Sprachtestexpert*innen AUA für Test(programme), für die sie selbst (mit)verantwortlich sind, so dass die Objektivität fraglich ist (vgl. Chapelle 2011: 26); Kane (2011: 4) spricht von einem „confirmationist bias“.

Besonders rar sind Studien, die sich an die heiklen, teils (sprachen-)politikbezogenen **Wirkungsdimensionen**⁵ von Validität heranwagen würden (vgl. Chalhoub-Deville 2016: 466 vgl. bereits Messick 1989). Gerade wenn Tests zu anderen als den originalen Zwecken eingesetzt werden, sind solche Wirkungsstudien jedoch von entscheidender Bedeutung.

Zum einen werden die **Zuständigkeiten** für solche Validierungen von Testkonsequenzen besonders intensiv debattiert. So betonen etwa die einflussreichen *Standards for Educational and Psychological Testing* (AERA 2014: 24, Hvhbg. der Autoren), es sei auch bei der Verwendung eines Tests für einen noch nicht validierten Zweck „incumbent on the **user** to justify the new use, collecting new evidence if necessary“. Gleichzeitig findet sich in der Literatur der Hinweis, auch **Testentwickler*innen** trügen Verantwortung für die Konsequenzen ihrer Tests (vgl. Kane 2013: 58). Sei ein Test längere Zeit erfolgreich im Einsatz, komme es häufig zu Neuinterpretationen und -verwendungen. Dann könnten sich Testentwickler*innen (im Fall des DSD I: die ZfA, s. Abschnitt 6) nicht darauf berufen, dass diese ursprünglich nicht Testzweck gewesen seien, sondern sie seien im Gegenteil mitverantwortlich für die neue „common practice“ (Kane 2013: 58). Chalhoub-Deville wirbt für eine flexiblere Handhabung der Rollen bei der Verantwortung: „Scholarship is called for to systematize practices for defining stakeholder groups, delineating roles, apportioning responsibilities, and elaborating communication/negotiation systems“ (2016: 468).

Zum anderen ist umstritten, inwieweit solche Konsequenzen überhaupt Gegenstand von Sprachtestvalidierungen sind (vgl. Brennan 2006: 8). Bachman (2013: 22–23) vermutet, dass Testentwickler*innen auch deshalb einer Berücksichtigung von Konsequenzen skeptisch gegenüberstehen, weil „the consequences of a test use is a huge unknown and that they do not want to in effect write a blank check for consequences they do not understand“. Demgegenüber geht das AUA aus der Perspektive des *Critical Language Testing* (z.B. McNamara/Roever 2006;

⁵ Die Begriffe Auswirkung, Wirkung, Effekt und Konsequenzen werden hier synonym verwendet.

McNamara/Ryan 2011; Shohamy 2001) noch nicht weit genug bei der Berücksichtigung von Testfolgen. Das *Critical Language Testing* problematisiert insbesondere machtpolitische Aspekte von Tests als Kontroll- und Steuerungsinstrumente und thematisiert vornehmlich *nicht* intendierte Konsequenzen von Tests. Studien in dieser Ausrichtung führen häufig fairness- und gerechtigkeits-theoretische Zweifel an und rekurrieren dabei auf sozialwissenschaftliche Bezugsdisziplinen. In oft ideologiekritischer Weise werden z.B. implizite Normierungen in Testkonstrukten und versteckte Gatekeepingfunktionen von Tests herausgestellt (vgl. Pochon-Berger/Lenz 2014: 19). Ferner werden (sprachen-)politische Aspekte des Testens betont: Shohamy (2007) spricht von Sprachtests als mächtigen sprachpolitischen Werkzeugen bzw. als „de facto language policy“ (vgl. auch Menken 2008).

Auch Chalhoub-Deville (2016: 460) kritisiert, dass weder das AUA noch neuere Ansätze von Kane (vgl. Kane 2013) die Konsequenzen von Sprachtests als soziale und politische Praxis ausreichend berücksichtigen würden. Sie schlägt stattdessen eine Einbettung in philosophische Handlungstheorien (*theory of action*, TOA) vor, um die komplexen Verwebungen politischer Mandate, der Testforschung und sozopolitischer Konsequenzen erfassen zu können (vgl. Chalhoub-Deville 2016: 456, 461; vgl. Bennett/Kane/Bridgeman 2011; Moss 2013). In TOAs spielen politische Aspekte (*policy*) somit eine große Rolle. Coburn/Turner (2011) entwickeln ein Framework, das insbesondere veränderte institutionelle Praktiken (*changes in practice*) aufgrund von Dateninterpretationen fokussiert und sich auch für die Untersuchung der Folgen des DSD I-Einsatzes gegebenenfalls eignen würde.

Neben diesen vielversprechenden neuen Validierungsansätzen, die im deutschsprachigen Raum noch jeglicher Anwendung harren, werden in der Diskussion um Testkonsequenzen seit langem auch die Begriffe *Washback* und *Impact* verwendet (vgl. Alderson/Wall 1993: 115). *Washback* bezieht sich vor allem auf Auswirkungen auf den Unterricht (vgl. Taylor 2005: 154); *Impact* gilt als Weiterentwicklung des *Washback*-Begriffs und wird als übergeordnetes Konzept verstanden, das Konsequenzen von Testverwendungen in ihren bildungs- und gesellschaftlichen Kontexten untersucht (vgl. Saville 2012: 4; Saville/Khalifa 2016: 77). In der Testforschung gewinnen diese Konzepte neuerdings wieder an Bedeutung, „um die Auswirkungen von Tests in ihren bildungs- und sozialpolitischen Kontexten zu untersuchen“ (Harsch 2016: 214). Chalhoub-Deville (2016: 458) zufolge sind die terminologischen Abgrenzungen jedoch kaum haltbar und *Impact* (eher in L2-Testkontexten üblich), *Washback* (eher in Lehrkontexten verbreitet) und *Konsequenzen* (typisch in bildungswissenschaftlichen Zusammenhängen) ließen sich synonym verwenden.

5 Zur Zielgruppenangemessenheit des DSD I für neu zugewanderte SuS

Die Übertragung eines Testverfahrens auf eine neue Zielgruppe stellt eine potenzielle Bedrohung der Validität dar, weil etwa bestimmte Testinhalte für die neue Zielgruppe nicht mehr angemessen sein könnten (vgl. AERA 2014: 54–55). Die möglichst detaillierte Beschreibung der für einen Test vorgesehenen Zielgruppe(n) (z.B. Altersgruppen inkl. kognitiver Entwicklung; Bildungshintergründe; sprachlern- und sprachbezogene Qualifizierungsbedarfe usw.) ist eine wichtige Voraussetzung für alle Validierungsstufen des AUA.

Das DSD I wird unverändert nicht mehr nur für SuS an Schulen im Ausland im Alter von 14-16 Jahren, die ca. 600-800 Stunden Deutschunterricht hatten (vgl. ZfA 2019: 4), eingesetzt, sondern auch für „neu angekommene Schülerinnen und Schüler mit Migrationshintergrund“, und zwar „vor allem jugendliche Flüchtlinge und junge Migranten, die sich erst kurze Zeit in Deutschland aufhalten und noch vor der Integration in den schulischen Regelunterricht“ (KMK 2019) stehen.

Sowohl die ursprünglichen als auch die neuen Zielgruppen des DSD I sind durch eine beträchtliche sprachliche Heterogenität gekennzeichnet. Jede Herausstellung einzelner Merkmale stellt deshalb eine Verkürzung dar und wird auf viele Lernende eben auch nicht zutreffen. Deshalb sind die folgenden Ausführungen nur als rahmende Vorüberlegungen ohne Anspruch auf Vollständigkeit zu verstehen, die Gemeinsamkeiten und Unterschiede bezüglich des Lernens und Gebrauchs der deutschen Sprache in unterschiedlichen Kontexten fokussieren.

Während SuS im Ausland Deutsch als Fremdsprache (DaF) überwiegend in institutionellen Kontexten gesteuert im Unterricht lernen, befinden sich die Lernenden im Inland in einer Situation migrationsbedingter lebensweltlicher Mehrsprachigkeit (vgl. Gogolin 1988; Lüdi/Py 2009). Sie haben deutlich mehr Gelegenheiten zum Sprachenkontakt, und eignen sich das Deutsche individuell unterschiedlich in Abhängigkeit einer Unzahl potenzieller individueller und anderer Einflussfaktoren auf hybride Weise und in sich über die Zeit dynamisch verschiebenden Lernkontexten an. Dies geschieht zumindest teilweise auch ungesteuert. Quantität und Qualität des DaF/DaZ-Unterrichts unterscheiden sich im Einzelfall, auch in Abhängigkeit des schulischen Integrationsmodells, erheblich (von additivem Deutschunterricht in Sprachklassen (Willkommens-/Vorbereitungsklassen) bis hin zur frühen Integration in den Regelunterricht, vgl. Massumi/von Dewitz/Grießbach/Terhart/Wagner/Hippmann/Altinay 2015: 45).

Zudem unterscheiden sich die „alten“ und „neuen“ DSD I-Zielgruppen hinsichtlich diverser Faktoren, die aus der je unterschiedlichen Rolle des Deutschen in institutionellen bzw. gesellschaftlichen Mehrsprachigkeitsgefügen resultieren. Die Debatte zur migrationsbedingten Mehrsprachigkeit (etwa zu Mehrsprachigkeit in der Schule, zur Mehrschriftlichkeit, zu interkulturellen Aspekten usw.) kann hier zwar nicht nachvollzogen werden (vgl. die aktuellen Beiträge etwa von Dirim 2015; Mecheril 2010; Riehl/Blanco López 2019; Sievers 2019; Woerfel/Riehl 2016). Jedoch sei betont, dass für die SuS im Inland das Deutsche die gesellschaftliche Mehrheitssprache und gleichzeitig meist die einzige im Unterricht verwendete Sprache ist (vgl. „monolingualer Habitus“, Gogolin 1994). Der sprachliche Integrationsdruck ist hier dementsprechend deutlich größer. Komplexe Sprachrepertoire-Konstellationen (nicht nur Kombinationen mehrerer Sprachen, auch Diglossiesituationen bzw. verschiedene sprachliche Register innerhalb der beteiligten Sprachen inklusive bereits gelernter Fremdsprachen und der Herkunftssprache(n)) sind anzunehmen.

Demgegenüber ist das Deutsche, bei aller Heterogenität der auch dort auftretenden individuellen Mehrsprachigkeitsrepertoires, für SuS im Ausland in aller Regel nicht die (überwiegende/alleinige) Schul- bzw. Unterrichtssprache, nicht die erste gelernte Fremdsprache und auch nicht die gesellschaftlich mehrheitlich gesprochene Sprache. Soll das schulische Deutschlernen dort außerunterrichtlich gefördert und gefestigt werden, muss es aktiv in institutionelle (schulische) Mehrsprachigkeitsdiskurse eingebunden werden, bspw. durch bilingualen/deutschsprachigen Sachfachunterricht. Dadurch erlangen im Ausland die weiteren individuellen und institutionellen Sprachen ein größeres Gewicht.

Neben diesen wichtigen Unterschieden bestehen auch Gemeinsamkeiten zwischen DSD-Zielgruppen im In- und Ausland. Diese betreffen bspw. das anzunehmende Alter beim Erstkontakt mit dem Deutschen. Dieses liegt bei Jugendlichen mit Fluchthintergrund, die zur neuen DSD I-Zielgruppe gehören (14–16 Jahre) mit großen interindividuellen Schwankungen grob geschätzt bei mindestens 10 Jahren. Damit ähnelt der Zeitpunkt des Erwerbsbeginns eher den SuS im Ausland als dem (früheren) bei in Deutschland geborenen Kindern mit Migrationshintergrund (zum L2-Erwerb bei Jugendlichen vgl. Czinglar 2014, 2018; Massumi et al. 2015).

Die neuen DSD I-Zielgruppen erwerben das Deutsche also relativ spät und in variierenden Anteilen gesteuert/ungesteuert. Dies hat Konsequenzen auch für die im Schulkontext entscheidende Ausbildung literaler Kompetenzen. Während „klassische“ DaZ-Lernende (Deutscherwerb spätestens ab der Einschulung) oft in ihren Herkunftssprachen nur über äußerst eingeschränkte bildungssprachliche Kompetenzen verfügen (vgl. Ahrenholz 2010: 23), ist dies bei neu zugewanderten SuS mit erstsprachlicher Literalisierung nicht notwendig der Fall (ebenso wenig wie bei den

ursprünglichen Zielgruppen). Hier sind sehr heterogene Literalisierungswege und -szenarien denkbar. Allerdings gilt: selbst wenn Literalisierungserfahrungen angenommen werden, beginnt das Deutschlernen von Seiteneinsteiger*innen i.d.R., wenn die Herkunftssprachen noch nicht in allen Registern ausgebaut sind (vgl. Ahrenholz 2010: 26), so dass die Jugendlichen am Anfang einer sich ausbildenden Mehrschriftlichkeit stehen (vgl. Woerfel/Riehl 2016).

Weitere Spezifika der neuen Zielgruppe betreffen die deutlich größere herkunftssprachliche und Alters-Heterogenität der SuS im Inland. Außerdem wird Deutsch im Ausland häufig von einer Minderheit von SuS als zweite oder dritte Fremdsprache gelernt und geht dann mit hohen familiären Bildungsaspirationen und vermutlich auch einem überdurchschnittlichen sozio-ökonomischen Status einher. Gleichzeitig kann sowohl für neue als auch für die ursprünglichen DSD I-Zielgruppen zumindest teilweise von (positiven und negativen) Transfereffekten bereits gelernter Fremdsprachen ausgegangen werden.

Zuletzt, und mit Nachdruck, sind aus ethischen Gründen psychische Gesichtspunkte zu berücksichtigen. Viele Kinder und Jugendliche mit Fluchthintergrund bringen aufgrund schwer traumatisierender Erlebnisse sehr belastende Verlust Erfahrungen mit. Vor diesem Hintergrund müssten Testaufgaben besonders sensibel gegenüber dieser Zielgruppe sein. Im Ausland adäquate Aufgaben können im harmlosesten Fall die Lebenswelt der Kinder und Jugendlichen in Deutschland ignorieren; unethisch jedoch wären Aufgaben, die erlebte Traumata berühren und die Teilnehmenden dadurch zusätzlichen psychischen Belastungen aussetzen (s. Abschnitt 9 für diesbezügliche Beispiele).

Die o.g. Spezifika der „neuen“ DSD I-Zielgruppen bedeuten nun keineswegs, dass das Testverfahren prinzipiell nicht für die neuen Zielgruppen anwendbar wäre. Vielmehr wird ein Validierungsdesiderat deutlich: Ausstehend ist der argumentativ und empirisch belegte Nachweis dafür, dass der Sprachtest ausreichend genau auch auf Besonderheiten der SuS im Inland zugeschnitten ist.

6 Zu erwarteten (und möglichen unerwarteten) Konsequenzen des DSD I-Einsatzes im Inland

Die nachfolgenden Abschnitte des vorliegenden Beitrags folgen dem Aufbau des oben eingeführten *assessment use argument* (AUA, s. Abschnitt 4, Abb. 1). In diesem Validierungsframework hängt die Validität eines Tests zuallererst entscheidend davon ab, dass „die Konsequenzen, die sich aus der Verwendung und Interpretation eines Tests ergeben, für alle Stakeholder positive Auswirkungen haben und für niemanden schädlich sind“ (Pochon-Berger/Lenz 2014: 19). Auch wenn es

zunächst kontraintuitiv erscheinen mag, beginnt unsere Problematisierung von Validitätsaspekten deshalb bei den beabsichtigten und nicht beabsichtigten Auswirkungen der DSD-Verwendung im Inland (dieser Abschnitt) und widmet sich anschließend den Einsatzzwecken des DSD I (Abschnitt 7). Beide über den Test selbst hinausweisenden Aspekte hängen eng miteinander zusammen (vgl. den Begriff der *consequential validity* bei Messick 1996: 254; Taylor 2005: 154).

Hier befassen wir uns zunächst mit der Ebene der **beabsichtigten** Konsequenzen. Testimplementierende Akteure greifen auf solche intendierten positiven Wirkungen zurück, um für den Einsatz von Testverfahren zu argumentieren. Dies ist auch beim DSD I im Inland der Fall. Dabei lassen sich die antizipierten, allerdings bislang nicht empirisch nachgewiesenen (vgl. jedoch die Studie von Ricart Brede/Schrage 2018) positiven Effekte des DSD I-Inlandseinsatzes in vier Ebenen gliedern.

Von Vertreter*innen der ZfA werden erstens prominent erwünschte positive Effekte auf den Unterricht in den Vorbereitungsklassen angeführt. Dem DSD I wird also ein positiver „Rückkopplungseffekt auf den Deutschunterricht“ (Toledo/Mayer 2014: 14-15; vgl. auch Klein/Küpper/Wagner 2017: 320) zugesprochen, ein klassischer Washback-Effekt. Negative Effekte sollen vermieden und kontrolliert werden, schließlich solle die „Teilnahme am DSD [...] immer ein Nebenprodukt eines kompetenzorientierten Unterrichts bleiben, d. h. nicht *teaching to the test* ist das Ziel, sondern eine nachhaltige Wirkung auf die Didaktik und Methodik des Unterrichts“ (Klein et al. 2017: 315).

Der DSD I-Einsatz im Ausland ist an einen vorangehenden DaF-Unterricht gekoppelt und an diesen curricular gebunden (ZfA 2009). Auch im Inland sei zur Vorbereitung auf die Prüfung ein moderner (handlungsorientierter) Unterricht zwingend erforderlich, argumentieren Klein et al. (vgl. 2017: 319; vgl. Dronske 2016a: 22). So würden die SuS in Kompetenzen geschult, die sie insbesondere bei der Teilnahme an der Regelklasse, aber perspektivisch auch im Berufsleben oder Studium unterstützen können (z.B. beim „Argumentieren und Abwägen von Meinungen“ oder beim „Schreiben längerer Texte“ (Klein et al. 2017: 320)).

Auch wenn das Washback-Postulat hier nicht empirisch geprüft werden kann, ist kritisch anzumerken, dass das DSD I in der Praxis teils auch nach dem Übergang in die Regelklasse bzw. in Integrationsmodellen ohne separaten DaZ-Unterricht eingesetzt wird, wie Ricart Brede/Schrage (2018: 214) verdeutlichen (s. Abschnitt 7). Die erwünschten Washback-Effekte können sich in diesen Fällen nicht zeigen. Zudem besteht eine bekannte Gefahr der Rückwirkungen von Tests auf den Unterricht in einer curricularen Verengung auf Testinhalte und Testformate. Das Eintreten positiver Washbackeffekte durch die Einführung von Sprachtests wiederum ist

alles andere als selbstverständlich („each examination needs a tailor-made research project to investigate its washback“ (Pižorn/Huhta 2016: 249–250)), so dass allem voran der Frage nachgegangen werden müsste, inwieweit der Einsatz eines Tests nachweislich zu einer Modernisierung des Unterrichts führt.

Eng mit den erhofften Effekten auf den Unterricht verknüpft sind die erwarteten positiven Auswirkungen auf die Teilnehmenden des DSD I. Somit wirke laut Vertreter*innen der ZfA die Teilnahme an Vorbereitungsklassen, die mit dem DSD abschließen, sehr motivierend auf die SuS (vgl. Dronske 2016b: 137; Klein et al. 2017: 320–321). In diesem Zusammenhang werden insbesondere die authentische Ausrichtung der Inhalte und Aufgabentypen sowie ein frühes und regelmäßiges Feedback über den Lernfortschritt durch den Einsatz von Modellsätzen als motivierend angeführt (vgl. Klein et al. 2017: 321). Ricart Brede/Schrage (2018) führten eine Interviewstudie mit Lehrenden, Sprachlernberater*innen und Schulleitungen in Niedersachsen durch, die zu dem Schluss kommt, dass beim Inlandseinsatz für die SuS tatsächlich „‘Chancen und Nutzen‘ des DSD im Vergleich zu ‚Grenzen und Risiken‘“ überwiegen (Ricart Brede/Schrage 2018: 223). Für die SuS nehmen diese Stakeholder insbesondere eine Motivationsstärkung an, während größere Skepsis gegenüber „der Vorbereitung auf Regelunterricht und Berufsanforderungen bzw. hinsichtlich der bildungs- und fachsprachlichen Ausrichtung des Testformats“ vorherrscht (Ricart Brede/Schrage 2018: 224). Während für diese Studie die SuS selbst noch nicht befragt wurden, ist dies Bestandteil des rahmenden EVA-Sek-Projekts (vgl. Ahrenholz et al. 2017, 2018). Montanari (2016) greift in einer Begleitstudie zum DSD-Länderprojekt in Niedersachsen diesen Aspekt auf und untersucht in Kooperation mit dem Niedersächsischen Landesinstitut für Qualitätsentwicklung (NLQ) insbesondere den schriftlichen und den mündlichen Teil des DSD I. Auch Montanari berichtet, dass die SuS die Teilnahme an dieser Studie „als Stärkung und Motivation erlebt“ (Montanari 2016: 294) hätten, betont aber auch, dass die Anmeldung zur Prüfung „zu einer Orientierung weg von eigenen sprachlichen Handlungsbedarfen hin zu einer Prüfungsausrichtung des Unterrichts“ (ebd.) für die SuS führe. Insgesamt sei in dieser Fallstudie jedoch ebenfalls zu beobachten, dass die Prüfungsorientierung des Unterrichts und der Lernerfolg der SuS bisher in einem ausgeglichenen Verhältnis zueinanderstehen (ebd.).

Drittens werden positive Rückwirkungen des DSD-Einsatzes auf Lehrkräfte erwartet. Neben einem breiten Angebot an Fortbildungsmöglichkeiten zu den für das DSD I relevanten Kompetenzbereichen sowie beispielsweise zum „kriterienbasierten Bewerten von mündlichen und schriftlichen Leistungen“ (Dronske 2016a: 23) böten diese Veranstaltungen auch die Möglichkeit zur Vernetzung und zum Austausch der beteiligten Akteure (vgl. ebd.).

Zudem wird noch ein vierter Aspekt genannt, der nicht explizit auf die Unterrichtspraxis oder das direkte Umfeld des Tests bezogen ist. Der Inlandseinsatz des DSD I hat, laut ZfA, einen positiven Rückkoppelungseffekt für die Institution selbst, da sie durch den innerdeutschen Einsatz neue Expertisen dazugewinne und den Test in einem gänzlich anderen, nämlich inländischen Einsatzfeld erproben könne (vgl. Dronske 2016a: 24).

Zusammenfassend soll erneut deutlich darauf hingewiesen werden, dass es sich bei diesen intendierten Konsequenzen um Erwartungen handelt und eine „systematische Evaluation der Rückwirkungen des Einsatzes des DSD I samt flankierenden Maßnahmen auf den DaF/DaZ-Unterricht in inländischen Vorbereitungsklassen“ (Klein et al. 2017: 318) noch aussteht. Mit Ausnahme der zitierten Studie von Ricart Brede/Schrage (2018: 224), die zu Vorsicht „vor überhöhten Erwartungen gegenüber dem DSD als ‚Universalheilmittel‘“ warnt, und der ebenfalls zitierten Fallstudie von Montanari (2016), liegen keine Studien zu Konsequenzen des DSD-Einsatzes vor. Wie erwähnt ist es methodisch anspruchsvoll, Nachweise für das Eintreten intendierter Konsequenzen zu erbringen, was einer der Gründe für ihr bislang nahezu vollständiges Fehlen sein mag.

Zudem ist bedenklich, dass bislang zu *nicht intendierten Konsequenzen* des DSD I-Einsatzes nicht nur keine Nachweise, sondern offenbar auch keine systematischen Überlegungen vorliegen. Ein nicht bestandenes DSD I wird bspw. vermutlich ebenso demotivierend wirken wie ein bestandenes motivieren kann (vgl. Ricart Brede/Schrage 2018). Konsequenzen sind jedoch insbesondere dann heikel, wenn das DSD I zu Zuweisungszwecken eingesetzt wird (Abschnitt 7). Gerade hier stellt sich die Frage, inwiefern der Inlandseinsatz des DSD I noch seinem Grundgedanken als „Instrument zur Motivation und zur Förderung von Schülerinnen und Schülern“ (Klein et al. 2017: 321) gerecht wird. Zu bedenken ist dann, dass Sprachtests auch sozial ausgrenzen und systematisch benachteiligen können (vgl. McNamara/Roever 2006; McNamara/Ryan 2011; Shohamy 2001), etwa wenn bei in Bildungsinstitutionen eingesetzten diagnostischen Verfahren nicht alle, sondern nur ausgewählte SuS teilnehmen *dürfen* (vgl. Chik/Besser 2011). Die Einführung einer *high-stakes*-Testverwendung für ausgewählte Gruppen im deutschen Bildungswesen sozusagen „durch die Hintertür“ wirft auch Gerechtigkeitsprobleme auf. Exemplarisch ließe sich fragen, warum nur neu zugewanderte SuS, aber nicht andere mehrsprachige SuS zur Zulassung für bestimmte Schulformen das DSD I bestehen müssen oder warum diese Hürde nur in einzelnen Bundesländern oder sogar in einzelnen Schulen überwunden werden muss.

7 Formale Voraussetzungen und heterogene Einsatzzwecke des DSD I im Inland

Im folgenden Abschnitt befassen wir uns mit Zwecken, zu denen das DSD I im Inland eingesetzt wird. Entsprechend dem AUA-Validierungsframework beziehen wir uns damit auf die Ebene der *Entscheidungen*, die auf Grundlage von Testergebnissen getroffen werden (vgl. „decisions“ im *assessment use argument*, Bachman/Palmer 2010: 194–199; Abb. 1). Diese sind zwar von den Konsequenzen, die sich aus dem Testeinsatz ergeben, zu differenzieren (s. Abschnitt 6), aber eng mit ihnen verknüpft und bei der Testentwicklung von zentraler Bedeutung: „In developing a test, we typically have some purpose in mind, and this purpose guides the development of both the test and the interpretation/use argument.“ (Kane 2013:16). Entscheidungen sind direkte Folgen einer Testteilnahme.

In Abhängigkeit der Entscheidungen, die aufgrund von Testergebnissen über Testkandidat*innen getroffen werden, lassen sich Tests in zwei Haupttypen untergliedern. So dient schulische Sprachdiagnostik meist förderdiagnostischen Zwecken (z.B. der Abbildung von Lernfortschritten), seltener jedoch dazu, Zuweisungsentscheidungen zu treffen („Selektionsdiagnostik“, vgl. Gogolin 2010: 1307; Wilhelm/Kunina 2009: 309). Typische Zwecke für Letzteres wären Entscheidungen über Sprachförderbedarf oder bspw. darüber, ob eine Förderung im Regelunterricht oder in gesonderten Klassen stattfinden soll (vgl. Jeuk/Lütke 2019: 164). Selektierend verwendete Sprachdiagnostik wird in vielen Fachkulturen kritisch gesehen (vgl. Jeuk 2013; Jeuk/Lütke 2019: 167). In Anbetracht der aus ihnen resultierenden Konsequenzen müssen besondere Qualitätsanforderungen an solche zuweisungsorientierten *high stakes*-Verfahren gestellt werden.

Für beide Testzweckbereiche kommen auch zunehmend Verfahren zum Einsatz, die der Zertifizierung sprachlicher Kompetenzen dienen, die dann zu einem bestimmten i. d. R. in die Zukunft verweisenden Zweck als ausreichend bescheinigt werden, etwa zur Aufnahme eines Studiums, zum Beginn einer Ausbildung oder einer beruflichen Tätigkeit (vgl. Jeuk/Lütke 2019: 165–166). Damit gehen solche Verfahren über die reine Feststellung eines Sprachstands hinaus.

Im Fall des Inlandseinsatzes des DSD bildet den rechtlichen Rahmen für mögliche konkrete Einsatzzwecke ein Beschluss der Amtschefkonferenz der KMK aus dem Jahr 2012 (KMK 2012). Dieser legt zunächst allgemein die Zulässigkeit des innerdeutschen Einsatzes des allgemeinsprachlichen DSD I zur „sprachlichen Erstintegration von Schülerinnen und Schülern mit Migrationshintergrund“ (KMK 2012:

1) fest, ohne noch genauere Beschränkungen für spezifische Einsatzzwecke anzugeben.⁶

Die aktualisierten Ausführungsbestimmungen (ZfA 2018|2019: 14) legen dann fest, dass ein Bundesland bei Interesse am Einsatz des DSD I den Zentralen Ausschuss (bestehend aus Vertreter*innen der Länder und des Bundes, d.h. des Auswärtigen Amts und der ZfA) kontaktieren und eine konkretisierende Verwaltungsvereinbarung mit dem Sekretariat der KMK und der ZfA abschließen muss, „wobei hier für einzelne Punkte Vereinbarungen getroffen werden können, die von den allgemeinen Bestimmungen der geltenden Ausführungsbestimmungen abweichen, sofern sie nicht der Prüfungsordnung entgegenstehen, dadurch der Charakter der Prüfung nicht grundsätzlich verändert wird und das Anforderungsniveau der Prüfung bestehen bleibt“ (ZfA 2018|2019: 14). Die Verwaltungsvereinbarungen sind nicht öffentlich zugänglich.

Ferner regelt die im September 2018 neu beschlossene Prüfungsordnung zum DSD der KMK (d.h. zum DSD I und II sowie zum DSD I Pro) nun auch den Einsatz des DSD der ersten Stufe „im Rahmen der schulischen Erstintegration in Deutschland“ (ZfA 2018a: 3, §1 und ZfA 2018a: 5f., §6). Auf Entscheidungsebene legt die Prüfungsordnung bezüglich des DSD I jedoch nur fest, dass dieses die notwendigen Sprachkenntnisse bescheinigt, die für die Zulassung zu einem Studienkolleg nötig sind (ZfA 2018a: 11, §16). Da Studienkollegs nur von bildungsausländischen Studienanwärter*innen besucht werden, betrifft dieser Testzweck nicht die neue, inländische Zielgruppe.

Insgesamt entsteht sowohl für den Auslandseinsatz, aber durch die jeweilige Ausgestaltung der Verwaltungsvereinbarungen der Länder auch für die DSD-Verwendung im Inneren ein Freiraum bei der Festlegung der Entscheidungen, die auf Grundlage des DSD I getroffen werden können.

So sind zunächst im Ausland verschiedene Einsatzzwecke möglich und üblich. An einigen Deutschen Auslandsschulen und DSD-Schulen wird das DSD I als „Zwischenstufe auf dem Weg zum DSD II“ (Klein et al. 2017: 314) eingesetzt, jedoch ist diese Verwendung keinesfalls obligatorisch. Zum anderen bestätigt das Bestehen des DSD I auf B1 die Fähigkeit, am deutschsprachigen Fachunterricht teilzunehmen (Dronske 2016a: 21)⁷ und wird dazu herangezogen, den „Übergang in den deutsch-

6 Ein Auszug aus der Niederschrift der entsprechenden Sitzung (NS. 212. AK) vom 15.11.2012 wurde den Autor/innen dieses Beitrags von der KMK zur Verfügung gestellt.

7 Dazu finden sich im DaF-Rahmenplan (2009) mehr Informationen. Dieser versteht sich als „verbindliches Gerüst“ (ZfA 2009: 3) für den Fremdsprachenunterricht (FSU) Deutsch an DSD-Schulen, der explizit auf das DSD vorbereiten muss (ZfA 2009: 5). Beim Erreichen des Niveaus B1 (vorgesehen für den 3. Bildungsabschnitt, vergleichbar 10. Klasse) sollen demzufolge „SuS

sprachigen Zweig“ (ebda.) zu regeln. Insbesondere dieser Verwendungszweck ähnelt laut Vertreter*innen der ZfA dem Inlandseinsatz des DSD (vgl. Dronske 2016a: 21). Wäre das Bestehen des DSD I zwingende Voraussetzung für die genannten Übergangsszenarien, könnte es über die Zertifizierungsfunktion hinaus als Prüfung mit Zuweisungs- bzw. Selektionscharakteristik beschrieben werden.

Für den Inlandseinsatz geht es in der Lesart des Amtschefbeschlusses der KMK (persönliche Kommunikation) und der ZfA „um die Unterstützung und Optimierung der sprachlichen Integration von neu zugewanderten Jugendlichen beim Übergang in den Regelunterricht“ (Klein et al. 2017: 314; vgl. KMK 2012), während das DSD „keine formale Voraussetzung darstellen darf, um von einer Vorbereitungs-klasse in eine Regelklasse zu wechseln“ (Klein et al. 2017: 321). Selezierenden Charakter sollte das DSD I im Inlandseinsatz damit laut KMK/ZfA (persönliche Kommunikation) nicht haben. Diese Einschränkung ist jedoch nicht auf Bundesebene normierend festgehalten; in Publikationen rund um die Verwendung des DSD I im Inland liegt zwar ein starker Fokus auf den erwünschten *Konsequenzen* (Motivation, Förderung, Washback, Klein et al. 2017: 321; s. Abschnitt 6), nicht jedoch auf den DSD-basierten Entscheidungen.

Demgegenüber sind offenbar die konkreten Testzwecke in der Praxis heterogen, was sicherlich auch auf die Kultushoheit der Bundesländer zurückgeht, die ja selbst für die Einsatzformen, Durchführung und Leitung der Prüfung zuständig sind (vgl. ZfA 2018a: 3). Bislang existieren zwar keine öffentlich zugänglichen, umfassenden Berichte über den Einsatz des DSD I in den Bundesländern bzw. in den Schulen, deren Integrationsmodelle sich teils deutlich voneinander unterscheiden (vgl. Daschner 2017; Massumi et al. 2015). Ricart Brede/Schrage (2018) liefern jedoch einen systematisierenden Überblick über unterschiedliche Einsatzzwecke des DSD. An dieser Stelle kann das Verwendungsspektrum des DSD I nicht detailliert wiedergegeben werden. Bekannt ist aber etwa, dass das DSD I an Schulen teils vor dem vollständigen Übergang von der Vorbereitungsklasse (VK) in die Regelklasse (RK) eingesetzt wird, teils danach; an wiederum anderen Schulen ist eine Beschulung in einer VK gar keine Voraussetzung für eine Aufnahme in die RK und unterscheidet sich dann frappierend vom Auslandseinsatz, wo dies zwingend der Fall ist (vgl. Ricart Brede/Schrage 2018: 213, s. Abschnitt 6). An manchen Schulen absolviert nur ein Teil der betreffenden SuS das DSD I, an anderen alle. In einem weiteren Einsatzmodell wird das DSD vor und/oder nach dem Übergang der SuS in die RK in Kombination mit anschließenden weiteren Förderungsmaßnahmen eingesetzt (vgl. Ricart Brede/Schrage 2018: 214).

erfolgreich am deutschsprachigen Fachunterricht bzw. bilingualen (Sach-)Fachunterricht teilnehmen“ können (ZfA 2009: 15) und zur Teilnahme am Studienkolleg befähigt sein.

Auffällig ist, dass das DSD I auch als Zuweisungsentscheidung zu Schulformen (gymnasialer Zweig oder einfacher/mittlerer Schulabschluss) und als Zugangsvoraussetzung für den Eintritt aus der Vorbereitungsklasse (VK) in die Regelklasse (RK) eingesetzt wird (vgl. Ricart Brede/Schrage 2018: 214), also zu Zuweisungs- bzw. Selektionszwecken. Solche Testzwecke sind als *high-stakes* einzuschätzen, (s. Abschnitt 6), und dementsprechende Entscheidungen sind besonders gut zu begründen. Wird die Teilnahme am DSD I zu Zuweisungszwecken nur ausgewählten SuS ermöglicht, entsteht potenziell eine Schieflage bei der Chancengleichheit bzw. der Bildungsgerechtigkeit (vgl. Pižorn/Huhta 2016: 240). An dieser Stelle soll noch einmal ausdrücklich darauf hingewiesen werden, dass dieser high-stakes-Testzweck seitens KMK und ZfA zu keinem Zeitpunkt intendiert war, was nach Rücksprache mit beiden Institutionen auch bestätigt wurde (persönliche Kommunikation).

Für den Einsatz des DSD I zu Selektionszwecken liegen bislang keine Ergebnisse hinsichtlich der Validität der getroffenen Entscheidungen vor. Sowohl für inner-schulische Übergangszwecke als auch für Schulzugangentscheidungen müsste die Aussagekraft des DSD I diesbezüglich dringend quantitativ und qualitativ mehrperspektivisch analysiert werden (vgl. auch Ricart Brede/Schrage 2018: 214–215). Dabei spielt die sogenannte *prognostische Validität* eine zentrale Rolle. Hier sind zum einen unzuverlässige, falsch negative Entscheidungen zu vermeiden, bei denen aufgrund des Nichtbestehens des DSD I mit B1 einem/r Schüler*in der Zugang zur RK verwehrt wird, obwohl diese*r über dafür ausreichende Sprachkenntnisse verfügt. Falsch positive Entscheidungen hingegen würden solche Zugänge ermöglichen, obwohl ein*e Schüler*in keine ausreichenden Sprachkenntnisse hat. Dabei ist der letztgenannte Fall ethisch weniger problematisch. Wird jedoch die Freiheit von Getesteten grundlos eingeschränkt, ist eine Testverwendung ungerecht (vgl. Deygers/Van den Branden/Van Gorp 2017: 5).

Der Nachweis prognostischer Validität ist nun zum einen methodisch komplex, da oft das Vorhergesagte (bspw. schulischer Erfolg, sprachlich erfolgreiche Teilnahme am Regelunterricht o.ä.) schwer zu operationalisieren und mehrdimensional ist. Sprachliche Kompetenzen sind zudem zwar eine notwendige, aber keine hinreichende Bedingung bspw. für Schulerfolg. Eine im Testbereich notorische Schwierigkeit liegt ferner darin, dass die Verantwortung für den Nachweis prognostischer Validität von Testinstitutionen häufig an die Testnutzer*innen (hier: Schulen; s. auch Abschnitt 4) weitergereicht wird. Dies ist offenbar auch beim DSD I der Fall (vgl. Dronske 2016b: 138), wo die Schulen vor dem Hintergrund ihrer eigenen Integrationsmodelle zu entscheiden hätten, wie das DSD am besten eingesetzt würde. Umgekehrt ergibt sich gerade durch die dezentrale Organisation des DSD I-Einsatzes das Problem, dass die Testnutzenden nicht über die notwendigen finanziellen

Mittel und fachliche Expertise verfügen, um derartige Validierungsstudien durchzuführen. Dieses Dilemma sollte unseres Erachtens dringend angegangen werden, da die momentane Situation sich zuungunsten des schwächsten Glieds in der Entscheidungskette, namentlich der Getesteten, auswirken kann.

8 Konstrukte des DSD I: Zwischen „bildungssprachlichen“ und „kommunikativen“ Kompetenzen

Für die (Sprach-)Diagnostik spielt die Beschreibung des zu Messenden, des „Konstrukts“, eine zentrale Rolle. Das – möglichst theoriebasierte – Konstrukt legt das sprachlich Interessierende fest, das dann in Testaufgaben schlüssig zu operationalisieren ist. Im *assessment use argument* von Bachman/Palmer (2010: 209–242; vgl. Abb. 1) betreffen diesbezügliche Überlegungen die Ebene der *Interpretationen*, die aussagekräftig, verallgemeinerbar, relevant und hinreichend sein sollten.

Das Konstrukt des DSD I tangiert zwei sprachliche Kompetenzbereiche, die zwar sowohl für den Inlands- als auch für den Auslandseinsatz relevant sind, aber dort je unterschiedliche Bedeutung tragen, und zwar zum einen bildungssprachliche, zum anderen kommunikative Sprachkompetenzen. Die Gewichtung dieser Kompetenzbereiche scheint momentan (ohne eine Veränderung des Tests selbst) eine Umdeutung zu erfahren. Dieses wird im Folgenden anhand einer Dokumenten- bzw. Literaturanalyse illustriert. Dazu wird auf die Prüfungsordnung (ZfA 2018a), die Ausführungsbestimmungen (ZfA 2018|2019), die Handreichungen zu den Prüfungen (Sprechen, ZfA 2018b und Schreiben, ZfA 2019) sowie auf den DaF-Rahmenplan (ZfA 2009) rekurriert. Wegen seines engen Bezugs auf das DSD I und der Einbettung des Tests in ein umfassendes „DSD-Programm“ (ZfA 2019a), wird auch der nicht direkt testbezogene Rahmenplan hier berücksichtigt.⁸

In allen Prüfungsbestandteilen (s. Abschnitt 3) legt das DSD traditionell einen starken Fokus auf *kommunikative*, alltagsrelevante sprachliche *Handlungsfähigkeiten* (und referiert dabei klar auf den GER, Abschnitt 9). Ältere, auf den ursprünglichen Auslandseinsatz bezogene Dokumente, u.a. der – einigermaßen veraltete – DaF-Rahmenplan (ZfA 2009, vgl. ZfA 2017: 3), verweisen dementsprechend darauf, die

⁸ Dieser enthält im Übrigen auch ausführliche Hinweise zur Beurteilung und zum Prüfen von Deutschkenntnissen im Unterricht (vgl. ZfA 2009: 50–53). Der 2017 erschienene Anhang zum Rahmenplan Deutsch mit Fokus auf berufsorientiertes Deutsch geht explizit auf das DSD-Inlandsprojekt ein und definiert sich als „Soforthilfe“ (ZfA 2017: 3). Gleichzeitig wird ein neuer Rahmenplan als Desiderat benannt: „Darin müssten die Bedeutung der Projektarbeit oder Literaturunterricht stärker berücksichtigt werden, aktuelle Forschungsliteratur müsste ihren Niederschlag finden, und der berufliche Bereich müsste integriert werden“ (ZfA 2017: 3).

Hauptziele des DaF-Unterrichts seien sprachliche Handlungsfähigkeit, die Erweiterung von Selbst- und Weltverständnis und die Verbesserung der Sprachlernkompetenz und -bereitschaft. Auch die im Rahmenplan empfohlenen didaktischen Prinzipien spiegeln die Nähe zur Lerner-, Handlungs-, Kommunikations- und Mehrsprachigkeitsorientierung und zur Themen-, Inhalts-, Aufgaben- und interkulturellen Orientierung des GER (ZfA 2009: 13–14). Solche Zielsetzungen sind zwar v.a. im Auslandseinsatz plausibel (vgl. Dronske 2016b), aber natürlich auch für Deutschlernende im Inland unabdingbar.

Zweitens ist es vor dem Hintergrund des allgemein drängenden Desiderats DaZ-spezifischer bildungssprachlicher diagnostischer Verfahren (s. Abschnitt 2) für das DSD I als schulbezogene Prüfung nicht nur insbesondere aktuell im Inland, sondern grundsätzlich auch im Ausland von Bedeutung, *bildungssprachliche* Kompetenzen zu erfassen. Dieser zweite sprachliche Kompetenzbereich, der natürlich Überschneidungen mit dem o.g. Bereich kommunikativer Kompetenzen aufweist, wird in letzter Zeit stärker in den Vordergrund gerückt, während er in den oben erwähnten DSD-bezogenen Dokumenten (noch) keine Erwähnung findet. In neueren Publikationen zum DSD I-Einsatz im Inland ist zu lesen, die Prüfung habe eine „klar bildungssprachliche Orientierung“ (Dronske 2016b: 137, vgl. Dronske 2016a). So sei die Schreibaufgabe „deutlich bildungssprachlich“ (Dronske 2016a: 22): „Sie verlangt, dass der Prüfling einen zusammenhängenden Gesamttext verfasst. Dabei muss er als Fremdaussagen markierte Äußerungen anderer Personen wiedergeben, von seinen eigenen Erfahrungen berichten und eine eigene Meinung zu dem vorgegebenen Thema entfalten. Damit geht es um für Bildungssprache relevante Sprachhandlungen in kohärenten Texten [...]“ (ebd.). Auch die mündliche Prüfung sei bildungssprachlich, da sie verlange, ein „[...] selbstständig oder in einer Gruppe recherchiertes Thema vorzutragen und dieses in der anschließenden Diskussion zu vertiefen“ (ebd.). Ferner setze sie „einen projektorientierten unterrichtlichen Ansatz voraus“, da die jeweilige Präsentation „in der Willkommensklasse“ vorbereitet werden müsse (ebd.).

Ob das DSD I in der jetzigen Form beide Konstruktbereiche, die im Übrigen auch die unterschiedlichen Primärziele des L2-bezogenen Deutschunterrichts in Inland (Ziel: „sprachliche Integration in Regelklasse“ zur erfolgreichen Unterrichtsteilnahme) und Ausland („Kommunikationskompetenz für den Alltag“, Dronske 2016b: 139) spiegeln, allein durch Spezifizierungen *ex post* abzudecken vermag, ist fraglich. Insgesamt wäre eine stärkere Berücksichtigung bildungssprachlicher Kompetenzen, die ja wichtige aktuelle Entwicklungen des Fachs aufgreift, für das DSD I durchaus wünschenswert. Dazu wäre es jedoch vonnöten, diesbezügliche Testkonstrukte in einer zu überarbeitenden Testbeschreibung transparent und theoriebasiert zu benennen sowie die Operationalisierung detaillierter nachzuzeichnen,

u.a. auch in den Bewertungskriterien. Solche Modifikationen oder Neu-Perspektivierungen betreffen allerdings sowohl das DSD I im Inland als auch im Ausland.

Ferner ergeben sich Schwierigkeiten auf der Validierungsebene der Interpretationen, die besonders beim Einsatz des DSD I zu Zuweisungszwecken problematisch sind, aber im Prinzip auch bei weniger folgereichen Nutzungen eine Rolle spielen. Das sind v.a. Fragen danach, ob die getestete Sprache für den *target language use*-Bereich (z.B.: Sprachverwendung im Unterricht) relevant, hinreichend aussagekräftig und repräsentativ ist, so dass beabsichtigte Entscheidungen (z.B. Übergang in die Regelklasse) auch getroffen werden können. M.a.W. müsste evidenzbasiert begründet werden können, warum davon ausgegangen wird, dass bei Bestehen des DSD I tatsächlich aus sprachlicher Sicht die Teilnahme an einer Regelklasse möglich ist. Dies würde eine gute Kenntnis des Sprachgebrauchs in Regelklassen voraussetzen und umfangreiche Bedarfsanalysen mit verschiedenen Stakeholdern erfordern. Dieses Beispiel verdeutlicht exemplarisch, dass die Testzwecke sehr eng an die Konstrukte eines Tests zu binden sind und dass es problematisch ist, ein identisches Testverfahren für geänderte Zwecke einzusetzen, ohne auch auf Konstruktebene Modifikationen vorzunehmen.

9 Das DSD I im Inland – Zerreißprobe GER-Bezug?

Der GER wurde seit seinem Erscheinen 2001 zwar vielfach kritisiert, ist gleichzeitig aber zum wichtigsten europäischen Bezugsrahmen bei der Erstellung von Sprachtestverfahren geworden, wobei insbesondere das Niveaustufensystem regelmäßig genutzt (und häufig überstrapaziert) wird.

Auch für das DSD I spielt der GER eine wichtige Rolle. Es setzt GER-Niveaus (A2/B1) als *Kriterium* (d.h. als externen Referenzmaßstab) an und stellt auch zur Konstruktdefinition GER-Bezüge her (s. Abschnitt 8). Während grundsätzliche Kontroversen zum GER hier nicht aufgegriffen werden können, soll es vielmehr darum gehen, seine Angemessenheit als Bezugsgröße für den Einsatz des DSD I im Inland zu reflektieren, indem sowohl problematische Aspekte als auch Potenziale herausgestellt werden. Zu klären ist, inwiefern der GER als *tertium comparationis* für die neue Verwendung des DSD I im Kontext migrationsbezogener Mehrsprachigkeit von SuS geeignet sein kann.⁹ Solche Überlegungen verstehen wir mit De

9 Zu verweisen ist auf die grundlegende Kritik an den Skalen des GER wegen ihrer mangelnden theoretischen und empirischen Validierung (vgl. Alderson 2007; Deygers 2019; Fulcher 2004; Hulstijn 2007, 2014; Hulstijn et al. 2010; Little 2007; Wisniewski 2014, 2019). Aus der Skalierungslogik des GER resultierende Problematiken (exemplarisch vgl. Harsch 2018) werden somit grundsätzlich in GER-bezogenen Testkontexte „mitgenommen“. Dieser Gesichtspunkt wird

Jong/Zheng (2016) als „a priori“-Validierung.¹⁰ Die folgende Darstellung fokussiert zunächst die Passung des GER für die neuen Zielgruppen und Verwendungskontexte des DSD I und zeigt zweitens am Beispiel der interkulturellen Kompetenzen das Desiderat einer Anpassung des GER-Bezugs im DSD I bzw. zugehörigen Dokumenten auf.

Zunächst ist festzustellen, dass sich der GER vornehmlich an erwachsene Fremdsprachenlernende und nicht an jugendliche oder kindliche SuS richtet (vgl. Butler 2016; Krumm 2007; Little 2007; Schneider/Lenz/Forster/Vosicki 2019). Im schulischen Fremdsprachenunterricht spielt er dennoch eine wichtige Rolle, sei es bei der Lernzielbestimmung (vgl. etwa North 2014) oder aber in Sprachenportfolios (vgl. Krumm 2007; Schneider et al. 2019). Auch aus fremdsprachlichen Bildungsstandards (vgl. etwa KMK 2012a) ist das Dokument nicht wegzudenken. Zudem enthält der neu erschienene Begleitband zum GER (Europarat 2018) Skalen für junge Lernende (vgl. kritisch Jeuk/Lütke 2019: 166). Insofern ist der ursprüngliche Bezug des DSD I auf den GER durchaus plausibel.

Nichtsdestotrotz verlangt das Fehlen unterrichtsspezifischer Spezifikationen im Referenzrahmen einen flexiblen, modifizierenden Umgang mit den allgemein gehaltenen GER-Deskriptoren. Ein solcher kontextadaptierender Umgang zeigt sich bereits in den Dokumenten zur *ursprünglichen* Nutzung des DSD I: Zur Beschreibung der kommunikativen Teilkompetenzen werden in bearbeiteten GER-Deskriptoren teilweise **Bezüge auf Unterrichtssituationen** hergestellt, z.B.: „ausführliche Hinweise und Erklärungen (z.B. in Unterrichtssituationen) verstehen und angemessen umsetzen“ (B1-Deskriptor für Hörverstehen, ZfA 2009: 16). Auch die „sprachlichen Mittel“ werden z.T. auf den Unterricht bezogen. Ferner enthalten die Kompetenzerwartungen konkretisierende **Beispiele**, z.B. „Brief an eine unbekannte Gastfamilie, Bewerbung für einen Ferienjob, für ein Praktikum, Lebenslauf“ (ZfA 2009: 23). Einschränkend ist festzustellen, dass diese für die innerdeutsche Zielgruppe wohl nur teilweise anwendbar sind.

Neben dem Verwendungskontext „Schule“ ist jedoch erneut zu betonen, dass neu zugewanderte SuS in Vorbereitungsklassen sich das Deutsche nicht im „klassischen“ Sinne als Fremdsprache aneignen, sondern zumindest teilimmersiv in hybriden Aneignungsszenarien unter den Bedingungen lebensweltlicher Mehrsprachig-

jedoch im Folgenden außer Acht gelassen, da er nicht spezifisch für den DSD-Einsatz in innerdeutschen Schulen ist.

¹⁰ Um Nachweise für die Validität dieser Verbindung zu erbringen, müsste etwa gezeigt werden, dass SuS im DSD I vergleichbare Ergebnisse erzielen wie in einem anderen GER-bezogenen Test. Dazu liegen bislang aber weder Erkenntnisse zum ursprünglichen DSD-Einsatz noch zur Verwendung des Tests bei der neuen Zielgruppe vor.

keit (s. Abschnitt 5). Mag nun der breite Mehrsprachigkeitsbegriff des GER („Plurilingualität“, vgl. Lüdi/Py 2009) prinzipiell auch solche Lernwege umfassen (vgl. Europarat 2001: Kap. 1.3), steht das Lernen von Sprachen in migrationsbezogenen Kontexten mit zumindest partiellen sprachlichen Immersionssettings im GER doch nicht im Fokus (vgl. aber Čatibušić/Little 2014).¹¹ Die zunehmende Bedeutung, die der Europarat solchen Settings beimisst, wird zwar nicht zuletzt an der Anzahl an Initiativen, gebündelt v.a. an der Europarat-Institution ECML (*European Centre for Modern Languages*) (ECML 2019) deutlich, die den GER auch auf SuS mit Migrationshintergrund in schulischen Kontexten beziehen. Dennoch nimmt der GER in DaZ-Kontexten eine eher randständige Position ein (i.U. zu förderdiagnostisch konzipierten DaZ-Niveaubeschreibungen, vgl. Döll 2012; Döll/Reich 2013).

Darüber hinaus wurde wiederholt kritisiert, dass Überlegungen zur **Bildungssprache** im GER außen vor bleiben (z.B. Hulstijn 2014). Hulstijn (2014) warnt zudem davor, dass höhere GER-Niveaus eine Konfundierung mit kognitiven Leistungen nahelegen, die weder bei Kindern noch bei Jugendlichen vorausgesetzt werden könnten (vgl. Butler 2016; Little 2007). Gerade vor dem Hintergrund der Betonung bildungssprachlicher Aspekte für den neuen DSD I-Einsatz (s. Abschnitt 8) ist dies ein gewichtiger Einwand. Aus diesem Grund bezweifelt Dronske die Funktionalität des GER für den Inlandseinsatz: „Die Hierarchisierungen des GER, die sich über die Pole nah – fern, einfach – komplex, konkret – abstrakt strukturieren, sind wenig fungibel für bildungssprachlich fokussierte Sprachlernprozesse mit dem Ziel Regelschulunterricht“ (Dronske 2016b: 139).

Im Folgenden wird anhand des exemplarisch gewählten Bereichs der **interkulturellen Kompetenzen**¹² kurz verdeutlicht, dass eine Übertragung des GER-Bezugs nicht ohne eine kontextuelle Konkretisierung denkbar ist. Interkulturelle Kompetenzen spielen im ZfA-Rahmenplan eine wichtige Rolle. In der DSD-Prüfungsordnung (ZfA 2018: §3) werden sie „durch die Verknüpfung von Unterricht mit Begegnung, Auseinandersetzung mit fremdsprachigen und fremdkulturellen Texten“ erwähnt, sie kommen – als ‚schwer messbare Kompetenzen‘ (vgl. Hu 2008) wenig

¹¹ Die Europaratsinitiative LIAM thematisiert die sprachliche Integration erwachsener Migrant*innen (*Linguistic Integration of Adult Migrants*, <https://www.coe.int/en/web/lang-migrants/guiding-principles>, <https://www.coe.int/de/web/lang-migrants/forms-of-linguistic-integration>) und empfiehlt dafür explizit die Verwendung des GER. Shohamy (2007) kritisiert hingegen den GER-Einsatz in migrations- bzw. integrationsbezogenen Testverfahren scharf.

¹² Angesichts der verschiedenen anzunehmenden Sprachkontaktszenarien von DSD I-Kandidat*innen im In- und Ausland bietet der GER evtl. auch Potenziale im Bereich der Mediation, die wie die interkulturellen Kompetenzen im 2018 erschienenen GER-Begleitband (Europarat 2018; vgl. Deygers 2019; Wisniewski et al. 2019) ausführlicher skaliert wurde als zuvor. Aufgrund der anzunehmenden deutlich höheren Anzahl und Variation mehrsprachiger Kontaktsituationen bei innerdeutschen Zielgruppen böte sich hier die Chance, diesen Aspekt unter Beachtung schulischer, individueller und gesellschaftlicher Mehrsprachigkeit inhaltlich auszubauen und auszudifferenzieren.

überraschend – jedoch weder in den Ausführungsbestimmungen noch in den Bewertungskriterien des DSD vor.

Bei der Lektüre des Rahmenplans wird klar, dass sich Themen- und Erfahrungsfelder beider Testadressat*innengruppen zweifellos überschneiden (z.B. persönliche Lebensgestaltung, Ausbildung, Schule und Beruf, Kulturelles Leben und Medien). Die erwarteten „Einsichten und Erkenntnisse“ [sic] jedoch unterscheiden sich zwangsläufig: mit „Reisemöglichkeiten (z.B. Straßen-, Bahnverkehr, Jugendherbergen/Hotels/Campingplätzen vertraut“ sein (ZfA 2009: 31)) erscheint zumindest schief, als typische „Handlung“ in den doch hoffentlich alltäglichen „Begegnungssituationen“ ist die Auswahl von „Briefkontakten zu Deutschen“ dann absurd (ebd.).

Hu beklagt zudem (2008: 22), dass der GER allgemein das Phänomen „migrationsbedingter Mehrsprachigkeit, kulturell-sprachlicher Mehrfachzugehörigkeiten und die möglichen Konsequenzen daraus für sprachliches Lernen unterbelichtet“ und stattdessen auf der Dichotomie der eigenen versus einer Fremdkultur beharre. Insgesamt ist damit auch über die Frage der neuen Zielgruppenangemessenheit hinaus fraglich, was unter der Deutung von Sachverhalten „aus einer spezifischen kulturellen Perspektive“ (ZfA 2009: 31) zu verstehen sei. Hier bieten weder der originale GER noch die diesbezüglichen Skalen im erweiterten Begleitband eine wirklich schlüssige Hilfestellung (vgl. dazu Studer 2020).

Resümieren lässt sich, dass der GER durch seine sprachenpolitische, zwar programmatische, aber nicht präskriptive Ausrichtung allgemein und **flexibel genug** sein mag, um auch auf die Lernendengruppe neu Zugewanderter anwendbar zu sein. Klar scheint aber auch, dass eine Übernahme des GER-Bezugs eine Reflexion, Anpassung und **Konkretisierung** hinsichtlich der neuen Zielgruppen und Lernkontexte erfordert. Dieses Desiderat der Kontextadäquatheit betreffend ist unter anderem zu kritisieren, dass GER-Deskriptoren keinerlei Hinweise auf schulische Lernkontexte enthalten. Da der Referenzrahmen sich aber klar als offenes Dokument versteht (vgl. Europarat 2018: 21; North 2014: 230), würde eine solche Anpassung aus GER-Perspektive jedoch als Routinevorgehen verstanden werden: Das GER-Konzept sieht vor, dass Deskriptoren einzelner Skalen, die bewusst „kontextfrei“ formuliert sind, für spezifische Verwendungszwecke modifiziert werden können und sollen (vgl. exemplarisch Hasselgreen 2013).

Inwiefern solche Resultate der im GER vorgesehenen Gestaltungs- und Interpretationsfreiheit der Deskriptoren letztendlich noch im Einklang mit dem Referenzdokument stehen, wäre empirisch und theoretisch zu prüfen (kritisch dazu vgl.

Eckes/Althaus 2020). Insbesondere wäre die Belastbarkeit des Bezugs solcher Konkretisierungen auf GER-Niveaustufen empirisch und theoretisch nachzuweisen, wenn sie in Testverfahren implementiert werden.

10 Inhalte und Themen des DSD I: Zwischen Berlinalebesuch und Kriegsszenarien

Die Problematik zahlreicher der o.g. Punkte, insbesondere der interkulturellen Aspekte, zeigt sich nicht so sehr in den konstruktbezogenen, recht allgemein formulierten Dokumenten, sondern eher bei deren Operationalisierung in Testaufgaben. Da ein Washback-Effekt des Tests auf den Unterricht (sl. Abschnitt 6) erwünscht ist, haben gerade Testinhalte potenziell einen großen Einfluss auf Unterrichtsgegenstände, können curriculare Ziele dabei aber auch verengen (vgl. Pižorn/Huhta 2016: 244). Ferner sollte(n) die sprachliche Domäne(n) von Testaufgaben laut Bachman/Palmer (1996) den zielgruppenrelevanten Sprachgebrauch authentisch spiegeln (so genannte *target language use tasks*).

Für den vorliegenden Beitrag wurden die Lese- und Schreibaufgabenthemen der online verfügbaren sieben Modellsätze (MS) analysiert. Dabei zeigte sich zunächst eine potenzielle **lebensweltliche Ferne** vieler Themen für wohl einen Großteil der SuS im Inland (z.B. *Berlinalebesuch*, *Austauschjahr in den USA*, *Schüler machen Golf und Yoga*, *Schülerausflug nach Colmar*), die eher die Perspektive des bildungsbürgerlichen Mittelstands (*Tanzkurse*, *Opernbesuche der Eltern*) spiegeln. Zudem stechen **unvorsichtige Ausdrucksweisen** ins Auge (*Rettungsaktion*, MS7; *gelungene Flucht*, MS4, SuS, „deren Eltern für längere Zeit ins Ausland müssen“, MS1).

Bedenklicher ist jedoch, dass in einigen Aufgaben Themen behandelt werden, die **traumatische Erlebnisse** wachrufen können. So sind in MS4 Kurztexte enthalten, in denen Flucht- und Verlufterlebnisse thematisiert werden (Abb.2).

21	Die Uhr schlägt zwölf. Mitternacht. Die Schritte auf dem Gehweg hinter ihm kommen immer näher. Es ist bitterkalt heute Nacht. Keine Sterne, nur der Mond scheint. Gleich erreicht ihn der Fremde. Was will der Fremde von ihm? Etwas Böses? Er fühlt sich wie ein gejagtes Tier. Er sucht andere Menschen, die ihm zu Hilfe kommen könnten. Vergeblich. Schon spürt er den Atem des Verfolgers auf seiner Haut. Er beginnt zu laufen...	
----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

24	<p>Das kleine Mädchen an der Hand der älteren Frau ist traurig. Warum nur muss sie Frankfurt verlassen und in dieses kleine Dorf gehen? Was soll sie hier? Bei diesem alten Mann? Warum kann sie nicht bei der Tante bleiben? Sie liebt zwar die Berge und die Tiere, aber hier kennt sie niemanden. Schnell erobert das kleine Mädchen jedoch das Herz ihres Großvaters und schließt Freundschaft mit einem Jungen.</p>	
----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Abb.2: Textausschnitte aus dem Modellsatz 4 des DSD I, Leseverstehen

MS6 enthält einen Lesetext zu „Gorillas im Dschungel“, der beschreibt, dass infolge der Kriegssituation in vielen Ländern, in denen noch Gorillas leben, deren Lebensraum beschränkt werde und auch der Tourismus leide. In diesem Szenario („Menschen fliehen vor dem Krieg in den Dschungel“) wird das Schicksal der Gorillas empathisch konstruiert, peripher auch das der Touristen („Gefahr für das Leben der Touristen ist einfach zu groß“) und Parkranger („sind aus Angst geflohen“). Die Soldaten „kümmern sich natürlich nicht um die Affen“. Das Schicksal der kriegsbetroffenen Bevölkerung jedoch steht im Hintergrund. Für eine*n aus einem Kriegsland geflohene*n Testkandidaten*in mag dieser Text nachgerade zynisch erscheinen. Es ist höchst fraglich, ob es ethisch vertretbar ist, solche Texte für die neue Zielgruppe auszuwählen; notwendig ist es sicher nicht.

MS6 enthält ferner eine Schreibaufgabe zum Thema „Gewalt im Fernsehen“ (aus einem Statement, Teil des Prompts, dazu: „Ich habe kein Problem mit Gewalt im Fernsehen. In unserer echten Welt gibt es ja auch Gewalt“). In ihrer Antwort sollen die Testkandidat*innen ausführlich ihre Meinung zum Thema schildern. Eines der Bewertungskriterien betrifft die „eigene Erfahrung“ (Bestandteil des Bewertungskriteriums „Inhalt“, neben „Wiedergabe“ und „eigene Meinung“, ZfA 2019: 9): Hier wird „der Grad der Detailliertheit und Nachvollziehbarkeit der im Aufsatz dargestellten eigenen Erfahrungen bewertet“ (ZfA 2019: 11). SuS sollen also möglichst detailreich und nachvollziehbar diesbezügliche eigene Erfahrungen schildern.

Auch hier liegen ethische Probleme auf der Hand: Dass für Kinder und Jugendliche mit Fluchterfahrungen Gewalt höchstwahrscheinlich kein fiktives Konstrukt aus dem Fernsehen darstellt, wird ignoriert und zusätzlich in Kauf genommen, dass diese sich in einer Prüfungssituation, die sprachliche Kompetenzen prüfen soll, mit potenziell extrem belastenden Erlebnissen aus der eigenen Vergangenheit auseinandersetzen müssen. Aus Sicht der Autor*innen dieses Beitrags sollten die DSD-Aufgaben deshalb in dieser Hinsicht dringend überprüft werden.

11 Schluss

Dieser Beitrag stellt den Versuch eines validitätsbezogenen Problemaufrisses zum Einsatz des DSD I für Seiteneinsteiger*innen an innerdeutschen Schulen dar, ohne den Anspruch zu erheben, eine Validierung selbst bereits zu leisten.

Deutlich wurde, dass ein erhebliches Desiderat an standardisierten, aussagekräftigen Sprachtestverfahren für diese Zielgruppe vorliegt, das den Rückgriff auf das DSD zunächst plausibel erscheinen lässt. Gleichzeitig wurde aber auch klar, dass der Inlandseinsatz des DSD I hinsichtlich verschiedenster Aspekte einer enormen Validierungsanstrengung bedarf: Ganz grundlegend zeigte sich, dass die Deutschlernwege der neue(n) Zielgruppe(n) des Testverfahrens sich in vielen Punkten von denen ausländischer DaF-Lernender unterscheiden, jedoch der argumentative und evidenzbasierte Nachweis dafür, dass das DSD I auch auf die Inlandszielgruppe anwendbar ist, noch aussteht. Unser Problemaufriss folgte lose der Struktur des *assessment use argument* (AUA, Bachman/Palmer 2010) und verwies darauf, dass zwar viele positive *Konsequenzen* durch den DSD-Einsatz im Inland erwartet werden, Nachweise dafür jedoch noch nicht erbracht wurden. Ferner fiel auf, dass mögliche nicht intendierte positive und negative Konsequenzen, nicht nur auf individueller, sondern auch auf institutioneller und soziopolitischer Ebene, bislang nicht systematisch bedacht und untersucht wurden. Außerdem zeigte sich, dass die rechtliche Rahmenlage einen breiten Spielraum bezüglich möglicher heterogener *Einsatzzwecke* des DSD I eröffnet (vgl. Ricart Brede/Schrage 2018). Unter anderem wird es auch als *high stakes*-Selektionsinstrument für Übergangentscheidungen im Bildungswesen genutzt. Gleichzeitig wurde die prognostische Validität solcher Entscheidungen bislang nicht untersucht, und die Konsequenzen insbesondere dieser Verwendungsweisen sind unbekannt.

Abgesehen von diesen über das DSD selbst deutlich hinausweisenden Validitätsdimensionen zeigte dieser Beitrag, dass der (unveränderte) Test auf Konstruktebene eine Reinterpretation (von der kommunikativen, handlungsorientierten Sprachkompetenz hin zur Betonung der bildungssprachlichen Kompetenzen) erfährt. Dieser Spagat gelingt jedoch nur bedingt: Eine überzeugende Argumentation für die auch bildungssprachliche Relevanz des DSD bedürfte ebenfalls umfassender theoriebasierter empirischer Analysen, die noch ausstehen, sowie gegebenenfalls dann Modifizierungen des Tests selbst.

Des Weiteren widmete sich der Beitrag der Rolle des GER für das DSD I. Während Dokumente zum ursprünglichen DSD I-Einsatz auf schulische Kontexte bezogene konkretisierende GER-Bezüge enthalten, bleibt die Frage nach der Passung des ja sehr flexiblen, doch auch nicht ins Beliebige erweiter- und modifizierbaren GER

für schulische zweitsprachliche Lernende in Migrationskontexten offen. Gleichzeitig scheint der neu erschienene Begleitband zum GER (Europarat 2018) aber auch Chancen zu bieten.

Eine Analyse einiger Modellsätze zeigte schließlich, dass die Themen einiger Aufgaben für die neue Zielgruppe teils unpassend, teils sogar ethisch bedenklich sind, da sie potenziell erlebte Traumata wachrufen können.

Insgesamt ist das Fehlen von Validierungsstudien zum neuen DSD I-Einsatz auf allen Ebenen des AUA eklatant und insbesondere in Anbetracht der *high stakes*-Verwendungen zu Zuweisungszwecken auch nicht hinzunehmen. Jedoch ist diese Situation gewissermaßen typisch für die schwach ausgeprägte „Validierungskultur“ (vgl. Pochon-Berger/Lenz 2014: 34) zu Sprachtests des Deutschen. Wie oben erwähnt kommen Pochon-Berger/Lenz (2014) zu dem Schluss, es mangle insgesamt an einer solchen „Validierungskultur“, an Transparenz und an Nachweisen für gerechtfertigte Nutzungszwecke von Tests vor allem von unabhängigen Testexpert*innen (vgl. auch Pižorn/Huhta 2016: 244).

Erschwert wird die Zuweisung eindeutiger Verantwortlichkeiten für Validierungen an einzelne Institutionen (s. Abschnitt 3) beim DSD I weiter durch seine dezentrale und heterogene Nutzung. Eine Lösung schiene möglich, wenn Behörden auf Bundes- und Länderebene, Testentwickler*innen, durchführende Schulen (und idealiter externe Sprachtestexpert*innen) in geteilter Verantwortung ihre Rechenschaftspflicht wahrnehmen. In einer beträchtlichen Validierungsanstrengung würde es dann (mindestens) gelten, Spezifika der neuen Zielgruppe durchgängig zu berücksichtigen, ein passendes, revidiertes/erweitertes Konstrukt transparent zu definieren und dabei gegebenenfalls wissenschaftlich haltbare Bezüge auf die Bildungssprache einzubauen sowie die Bezüge auf den GER zu aktualisieren bzw. zu überdenken. Die Relevanz, Verallgemeinerbarkeit, Aussage- und Erklärungskraft der Konstrukte müsste hinsichtlich der konkreten Einsatzzwecke empirisch überprüft werden (z.B.: Deckt das DSD I die Qualität und Quantität von Sprache ab, die in der Regelklasse auch beobachtet werden kann? Werden im DSD I relevante Sprachhandlungen abgedeckt, die mit dem Übergang in die Regelklasse in Zusammenhang stehen?). Durch die Heterogenität der faktischen Verwendungszwecke ist dies eine sicherlich schwer zu meisternde Aufgabe. Zudem müssten die Testaufgaben hinsichtlich ethisch bedenklicher Inhalte und Themen überprüft werden.

Außerdem werden Studien zur prognostischen Validität des DSD I im Inland dringend gebraucht (z.B. mit der Fragestellung: Welche Hinweise gibt es dafür, dass ein erfolgreiches Absolvieren des DSD I auf B1 mit einem zumindest auf sprachlicher Ebene problemlosen Besuch einer Regelklasse zusammenhängt?) sowie Impactstudien, die erwünschte und unerwünschte Konsequenzen auf verschiedenen

Ebenen analysieren (z.B. individuelle Konsequenzen für die Getesteten; institutionelle Washbackeffekte auf den Unterricht; soziopolitische Konsequenzen).

Literatur

- Ahrenholz, Bernt (2010): Bedingungen des Zweitspracherwerbs in unterschiedlichen Altersstufen. In: Friedrich-Ebert-Stiftung (2010): „*Sprache ist der Schlüssel zur Integration*“ *Bedingungen des Sprachlernens von Menschen mit Migrationshintergrund*. Bonn: Bonner Universitäts-Buchdruckerei, 19–29.
- Ahrenholz, Bernt; Ohm, Udo & Ricart Brede, Julia (2017): Das Projekt „Formative Prozessevaluation in der Sekundarstufe. Seiteneinsteiger und Sprache im Fach“ (EVA-Sek). In: Fuchs, Isabel; Jeuk, Stefan & Knapp, Werner (Hrsg.): *Mehrsprachigkeit: Spracherwerb, Unterrichtsprozesse, Seiteneinstieg. Beiträge zum 11. Workshop „Kinder und Jugendliche mit Migrationshintergrund“*. Stuttgart: Fillibach bei Klett, 241–258.
- Ahrenholz, Bernt; Birnbaum, Theresa; Ohm, Udo & Ricart Brede, Julia (2018): Formative Prozessevaluation in der Sekundarstufe. Seiteneinsteiger und Sprache im Fach. In: Henschel, Sofie; Gentrup, Sarah; Beck, Luna; Stanat, Petra (Hrsg.): *Projektatlas Evaluation. Erste Ergebnisse aus den BISS-Evaluationsprojekten*. Berlin: BISS-Trägerkonsortium, 43–47.
- Alderson, Charles J. (2007): The CEFR and the Need for More Research. *The Modern Language Journal* 91, 658–662.
- Alderson, Charles J. & Wall, Diane (1993): Does Washback Exist? *Applied Linguistics* 14: 2, 115–129.
- Alderson, J. Charles; Brunfaut, Tineke & Harding, Luke (2015): Towards a Theory of Diagnosis in Second and Foreign Language Assessment: Insights from Professional Practice Across Diverse Fields. *Applied Linguistics* 36: 2, 236–260.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education [AERA] (2014): *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Bachman, Lyle & Palmer, Adrian (1996): *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

- Bachman, Lyle & Palmer, Adrian (2010): *Language assessment in practice. Developing language assessments and justifying their use in the real world.* Oxford u.a.: Oxford University Press.
- Bennett, Randy; Kane, Michael & Bridgeman, Brent (2011): *Theory of action and validity argument in the context of through-course summative assessment. Paper presented at invitational Research Symposium on Through Course Summative Assessment, Atlanta, GA.*
https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Paper_Bennett_Kane_Bridgeman.pdf (19.11.2019).
- BISS-Trägerkonsortium (2019): Tool-Dokumentation. <https://biss-sprachbildung.de/angebote-fuer-die-praxis/tool-dokumentation/> (22.11.2019).
- Brennan, Robert (2006): Perspectives on the evolution and future of educational measurement. In: Brennan, Robert (Ed.): *Educational measurement.* Westport: Praeger, 1–16
- Butler, Yuko Goto (2016): Assessing young learners. In: Tsagari, Dina & Banerjee, Jayanti (Eds.): *Handbook of Second Language Assessment.* Berlin: De Gruyter Mouton, 359–376.
- Ćatibušić, Bronagh & Little, David (2014): Immigrant pupils learn English. A CEFR-related empirical study of L2 development (English profile studies). Cambridge: Cambridge University Press.
- Chalhoub-Deville, Micheline (2016): Validity theory: Reform policies, accountability testing, and consequences. *Language Testing* 33: 4, 453–472.
- Chapelle, Carol A. (2011): Validity argument for language assessment: The framework is simple. *Language Testing* 29: 1, 19–27.
- Chik, Alice & Besser, Sharon (2011): International language test taking among young learners: A Hong Kong case study. *Language Assessment Quarterly* 8: 1, 73–91.
- Coburn, Cynthia E. & Turner, Erica O. (2011): Research on Data Use: A Framework and Analysis. *Measurement: Interdisciplinary Research and Perspectives* 9: 4, 173–206.
- Czinger, Christine (2014): Was wir aus Altersunterschieden für die Grammatikvermittlung lernen können. In: Dengerscherz, Sabine; Businger, Martin & Taraskina, Jaroslava (Hrsg.): *Grammatikunterricht zwischen Linguistik und Didaktik: DaF/DaZ lernen und lehren im Spannungsfeld von Sprachwissenschaft, empirischer Unterrichtsforschung und Vermittlungskonzepten.* Tübingen: Narr, 199–216.
- Czinger, Christine (2018): Zweitspracherwerb im Jugendalter. Die Bedeutung des Alters und literaler Kompetenzen von neu zugewanderten Jugendlichen. In: von Dewitz, Nora; Terhart, Henrike & Massumi, Mona (Hrsg.): *Neuzuwanderung und Bildung. Eine interdisziplinäre Perspektive auf Übergänge in das deutsche Bildungssystem.* Weinheim, Basel: Beltz Juventa, 158–173.

- Daschner, Peter (2017): Flüchtlinge in der Schule. Daten, Rahmenbedingungen und Perspektiven. Ein Überblick. In: McElvany, Nele; Jungermann, Anja; Bos, Wilfried & Holtappels, Heinz Günter (Hrsg.): *Ankommen in der Schule. Chancen und Herausforderungen bei der Integration von Kindern und Jugendlichen mit Fluchterfahrung*. Münster, New York: Waxmann, 11–26.
- De Jong, John & Zheng, Ying (2016): Linking to the CEFR using a priori and a posteriori evidence. In: Tsagari, Dina & Banerjee, Jayanti (Eds.): *Contemporary Second Language Assessment*, London: Bloomsbury, 83–100.
- Deygers, Bart (2019): The CEFR Companion Volume: Between Research-Based Policy and Policy-Based Research. *Applied Linguistics*. <https://doi.org/10.1093/applin/amz024> (22.11.2019).
- Deygers, Bart; Van den Branden, Kris & Van Gorp, Koen (2017): University entrance language tests: A matter of justice. *Language Testing* 449–476.
- Dirim, İnci (2015): Umgang mit migrationsbedingter Mehrsprachigkeit in der schulischen Bildung. In: Leiprecht, Rudolf & Steinbach, Anja (Hrsg.): *Schule in der Migrationsgesellschaft. Ein Handbuch* (Band 2). Schwalbach: Debus Pädagogik, 25–48.
- Döll, Marion (2012): *Beobachtung der Aneignung des Deutschen bei mehrsprachigen Kindern und Jugendlichen. Modellierung und Prüfung eines sprachdiagnostischen Beobachtungsverfahrens*. FörMig Edition. Münster, New York: Waxmann.
- Döll, Marion & Reich, Hans H. (2013): *Niveaubeschreibungen Deutsch als Zweitsprache für die Sekundarstufe I*. Radebeul: Sächsisches Bildungsinstitut. <https://publikationen.sachsen.de/bdb/artikel/14477/documents/26554> (20.4.2019).
- Döll, Marion & Reich, Hans H. (2015): *Niveaubeschreibungen Deutsch als Zweitsprache für die Sekundarstufe II*. Radebeul: Sächsisches Bildungsinstitut. <https://publikationen.sachsen.de/bdb/artikel/14477/documents/26554> (20.4.2019).
- Drackert, Anastasia; Mainzer-Murrenhoff, Mirka; Soltyska, Anna & Timukova, Anna (Hrsg.) (im Druck): *Testen bildungssprachlicher Kompetenzen und akademischer Sprachkompetenzen – Zugänge für Schule und Hochschule* Frankfurt u.a.: Lang. https://www.ruhr-uni-bochum.de/sprachetesten/mam/content/programm_1.pdf (20.4.2019).
- Dronske, Ulrich (2016a): Zum Einsatz des Deutschen Sprachdiploms in Vorbereitungsklassen. *Fremdsprache Deutsch*, 21–24.
- Dronske, Ulrich (2016b): Interview: Dr. Ulrich Dronske, Zentralstelle für das Auslandsschulwesen: Das Deutsche Sprachdiplom für neu zugewanderte Kinder und Jugendliche. In: Heinrichs, Petra (Hrsg.): *Unterricht für neu zugewanderte Kinder und Jugendliche in der Sekundarstufe I. Fachliche und fachdidaktische Perspektiven*. Köln: Bezirksregierung Köln. 137–140.

- Eckes, Thomas & Althaus, Hans-Joachim (2020): Language Proficiency Assessments in Higher Education Admissions. In: Oliveri, María Elena & Wendler, Cathy (Eds.): *Higher education admission practices: An international perspective*. Cambridge, UK: Cambridge University Press, 256–275.
- Ehlich, Konrad; Bredel, Ursula; Garne, Brigitta; Komor, Anna; Krumm, Hans-Jürgen; McNamara, Tim; Reich, Hans H.; Schnieders, Guido; ten Thije, Jan D. & van den Bergh, Huub (Hrsg.) (2007): *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*. Bonn, Berlin: BMBF. http://home.edo.tu-dortmund.de/~hoffmann/PDF/bildungsreform_band_elf.pdf (19.11.2019).
- Ehlich, Konrad (2007): Sprachaneignung und deren Feststellung bei Kindern mit und ohne Migrationshintergrund. Was man weiß, was man braucht, was man erwarten kann. In: Konrad Ehlich et al. (Hrsg.): *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*. Bonn, Berlin: BMBF. http://home.edo.tu-dortmund.de/~hoffmann/PDF/bildungsreform_band_elf.pdf (19.11.2019), 11–75.
- Europarat (Hrsg.) (2001): *Gemeinsamer Europäischer Referenzrahmen für Sprachen. Lernen, Lehren, Beurteilen*. Berlin u.a.: Langenscheidt.
- Europarat (Ed.) (2018): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Strasbourg. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989> (07.01.2019).
- European Centre for Modern Languages of the Council of Europe [ECML] (2019): *European Centre for Modern Languages of the Council of Europe*. <https://www.ecml.at/Home/tabid/59/language/en-GB/Default.aspx> (17.7.2019).
- Fachgruppe Sprachliche Bildung neu zugewanderter Kinder und Jugendlicher (2018): *Empfehlungen für den Übergang*. https://landesschulamt.sachsen-anhalt.de/fileadmin/Bibliothek/Landesjournal/Bildung_und_Wissenschaft/Landesschulamt/Migration/Empfehlungen_fuer_den_UEbergang.pdf (02.12.2019).
- Fulcher, Glenn (2004): 'Deluded by Artifices? The Common European Framework and Harmonization.' *Language Assessment Quarterly* 1: 4, 253–266.
- Geist, Barbara (2013): *Sprachdiagnostische Kompetenz von Sprachförderkräften*. Berlin u.a.: De Gruyter.
- Gogolin, Ingrid (1988): *Erziehungsziel Zweisprachigkeit. Konturen eines sprachpädagogischen Konzepts für die multikulturelle Schule*. Hamburg: Bergmann und Helbig.
- Gogolin, Ingrid (1994): *Der monolinguale Habitus der multilingualen Schule*. Münster u.a.: Waxmann.

- Gogolin, Ingrid (2010): Sprachstandsdiagnose. In: Krumm, Hans-Jürgen; Fandrych, Christian, Hufeisen Britta & Riemer, Claudia (Hrsg.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 1. Halbband*. Berlin, New York: De Gruyter Mouton, 1305–1314.
- Grießhaber, Wilhelm (2018): Narrative Texte von Seiteneinsteigerinnen und Seiteneinsteigern im Vergleich mit ein- und mehrsprachigen Schülerinnen und Schülern. In: Ricart Brede, Julia; Maak, Diana; Pliska, Enisa (Hrsg.): *Deutsch als Zweitsprache und Mehrsprachigkeit. Beiträge aus dem „Workshop Deutsch als Zweitsprache, Migration und Mehrsprachigkeit 2016“*. Stuttgart: Fillibach bei Klett, 161–182.
- Harsch, Claudia (2016): Testen. In: Schramm, Karen; Caspari, Daniela; Klippel, Friederike; Legutke, Michael (Hrsg.): *Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch*. Tübingen: Narr Francke Attempto, 205–218.
- Harsch, Claudia (2018): How suitable is the CEFR for setting university entrance standards? *Language Assessment Quarterly* 15: 1, 1–7.
- Hasselgreen, Angela Maria (2013): Adapting the CEFR for the Classroom Assessment of Young Learners' Writing. *Canadian Modern Language Review* 69, 415–435.
- Hennes, Ann-Kathrin; Schmidt, Barbara M.; Zepnik, Sabine; Linnemann, Markus; Jost, Jörg; Becker-Mrotzek, Michael; Rietz, Christian & Schabmann, Alfred (2018): Schreibkompetenz diagnostizieren: Ein standardisiertes Testverfahren für die Klassenstufen 4-9 in der Entwicklung. *Empirische Sonderpädagogik* 3, 294–310.
- Heppt, Birgit & Paetsch, Jennifer (2017): Diagnostik sprachlicher Kompetenzen im Schulbereich. In: Titz, Cora; Geyer, Sabina; Ropeter, Anna; Wagner, Hanna; Weber, Susanne & Hasselhorn, Marcus (Hrsg.): *Konzepte zur Sprach- und Schriftsprachförderung entwickeln*. Stuttgart: Kohlhammer, 117–137.
- Hoffmann, Lars; Böhme, Katrin & Stanat, Petra (2017): Mit welchen diagnostischen Verfahren wird in Grundschulen Sprachförderbedarf festgestellt? *Frühe Bildung* 6: 3, 116–123.
- Hoffmann, Lars & Böhme, Katrin (2017): Wird sprachlicher Förderbedarf in der Grundschule sicher erkannt? Zur Klassifikationsgüte von diagnostischen Entscheidungen. *Zeitschrift für Pädagogische Psychologie* 31: 2, 137–147.
- Hu, Adelheid (2008): Interkulturelle Kompetenz. Ansätze zur Dimensionierung und Evaluation einer Schlüsselkompetenz fremdsprachlichen Lernens. In: Frederking, Volker (Hrsg.): *Schwer messbare Kompetenzen. Herausforderungen für die empirische Fachdidaktik*. Baltmannsweiler: Schneider Hohengehren, 11–35.
- Hulstijn, Jan H. (2014): The Common European Framework of Reference for Languages: A challenge for applied linguistics. *ITL : International Journal of Applied Linguistics* 165, 3–18.

- Hulstijn, Jan H. (2007): The Shaky Ground beneath the CEFR: Quantitative and Qualitative Dimensions of Language Proficiency. *The Modern Language Journal* 91 663–667.
- Hulstijn, Jan H.; Alderson, Charles; Schoonen, Rob (2010): 'Developmental Stages in Second-Language Acquisition and Levels of Second-Language Proficiency: Are there Links between them?' In: Bartning, Inge; Martin, Maisa; Vedder, Ineke (Eds.): *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research*. EuroSLA. http://www.eurosla.org/monographs/EM01/11-20Hulstijn_et_al.pdf (20.11.2019), 5–10.
- Jeuk, Stefan (Hrsg.). (2013): *Sprachförderung und Förderdiagnostik in der Sekundarstufe I*. Stuttgart: Fillibach bei Klett.
- Jeuk, Stefan & Lütke, Beate (2019): Sprachdiagnostik im Primar- und Sekundarbereich. In: Jeuk, Stefan & Settineri, Julia (Hrsg.): *Sprachdiagnostik Deutsch als Zweitsprache: ein Handbuch*. Boston, MA: De Gruyter Mouton, 163–195.
- Junk-Deppenmeier, Alexandra & Jeuk, Stefan (Hrsg.) (2015): *Praxismaterial Förderdiagnostik. Werkzeuge für den Sprachunterricht in der Sekundarstufe I*. Stuttgart: Fillibach bei Klett.
- Kane, Michael (2001): Current concerns in validity theory. *Journal of Educational Measurement* 38, 319–42.
- Kane, Michael (2006): Validation. In: Brennan, Robert L. (Ed.): *Educational measurement*. Westport: Praeger, 17–64.
- Kane, Michael (2011): Validating score interpretations and uses. *Language Testing* 29: 1, 3–17.
- Kane, Michael (2013): Validating the interpretations and uses of test scores. *Journal of Educational Measurement* 50, 1–73.
- Klein, Wassilios; Küpper, Vera & Wagner, Iris (2017): Das Deutsche Sprachdiplom der Kultusministerkonferenz im Kontext der sprachlichen Erstintegration an weiterführenden und beruflichen Schule. In: Middeke, Annegret; Eichstaedt, Annett; Jung, Matthias & Kniffka, Gabriele (Hrsg.): *Wie schaffen wir das? Beiträge zur sprachlichen Integration geflüchteter Menschen*, Materialien Deutsch als Fremd- und Zweitsprache, Band 97. Göttingen: Universitätsverlag, 313–323.
- Kultusministerkonferenz, Amtschefkonferenz [KMK] (2012): *Niederschrift der Sitzung der Amtschefkonferenz am 15.11.2012* (NS. 212. AK).
- Kultusministerkonferenz [KMK] (2012a): *Beschluss der Kultusministerkonferenz vom 18.10.2012: Bildungsstandards für die fortgeführte Fremdsprache (Englisch/Französisch) für die Allgemeine Hochschulreife*.
- Kultusministerkonferenz [KMK] (2019): *Integration und Spracherwerb – Das DSD in den Ländern*. <https://www.kmk.org/themen/deutsches-sprachdiplom-dsd/integration-und-spracherwerb-vor-ort.html> (17.9.2019).

- Kultusministerkonferenz [KMK] (2019a): *Deutsches Sprachdiplom – Kurz und Knapp*. <https://www.kmk.org/themen/deutsches-sprachdiplom-dsd.html> (1.7.2019).
- Kultusministerkonferenz [KMK] (2019b): *Weltweite Prüfungen, zentrale Organisation – Struktur und Organisation des Programms zum Deutschen Sprachdiplom der Kultusministerkonferenz*. <https://www.kmk.org/themen/deutsches-sprachdiplom-dsd/weltweite-pruefungen-zentrale-organisation.html> (15.5.2019).
- Krumm, Hans-Jürgen (2007): Profiles Instead of Levels: The CEFR and its (Ab)Uses in the Context of Migration. *The Modern Language Journal* 91: 4, 667–669.
- Lengyel, Dorit (2012): *Sprachstandsfeststellung bei mehrsprachigen Kindern im Elementarbereich. Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WIFF)*. Frankfurt am Main: Deutsches Jugendinstitut e.V. http://www.poliglotti4.eu/docs/Expertise_Sprachstandsfeststellung.pdf (12.12.2016).
- Little, David. (2007): The Common European Framework of Reference for Languages: Perspectives on the making of supranational language European policy. *The Modern Language Journal* 91: 4, 645–655.
- Lüdi, Georges & Py, Bernard (2009): To be or not to be ... a plurilingual speaker. *International Journal of Multilingualism* 6: 2, 154–167.
- Lüdtke, Ulrike M. & Kallmeyer, Kirsten (2007): Kritische Analyse ausgewählter Sprachstandserhebungsverfahren für Kinder vor Schuleintritt aus Sicht der Linguistik, Diagnostik und Mehrsprachigkeitsforschung. *Die Sprachheilarbeit* 6, 261–278.
- Massumi, Mona; von Dewitz, Nora; Grießbach, Johanna; Terhart, Henrike; Wagner, Katharina; Hippmann, Kathrin & Altinay, Lale (2015): *Neu zugewanderte Kinder und Jugendliche im deutschen Schulsystem. Bestandsaufnahme und Empfehlungen*. Köln: Mercator-Institut.
- McNamara, Tim & Roever, Carsten (2006): *Language testing: the social dimension*. Malden, MA: Blackwell Publishing.
- McNamara, Tim & Ryan, Kerry (2011): Fairness versus justice in language testing: the place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly* 8: 2, 161–178.
- Mecheril, Paul (2010): Migrationspädagogik. Hinführung zu einer Perspektive. In: Mecheril, Paul; Castro-Varela, Maria do Mar; Dirim, İnci; Kalpaka, Annita & Melter, Claus (Hrsg.): *Migrationspädagogik*. Weinheim: Beltz, 7–22.
- Menken, Kate (2008): High-stakes tests as de facto language education policies. In: Shohamy, Elana & Hornberger, Nancy (Eds.): *Encyclopedia of Language and Education Vol. 7: Language testing and assessment*. Dordrecht: Springer, 401–413.

- Messick, Samuel (1989): Validity. In: Linn, Robert L. (Eds.): *Educational Measurement*. New York: Macmillan and American Council on Education, 13–103.
- Messick, Samuel (1996): Validity and washback in language testing. *Language Testing* 13: 3, 241–256.
- Meyer-Engling, Bettina (2014): Blick in die Statistik. In: ZfA (Hrsg.): *40 Jahre Deutsches Sprachdiplom der Kultusministerkonferenz*, 32–33.
- Mislevy, Robert J.; Almond, Russell G. & Lukas, Janice F. (2003): A brief introduction to evidence-centered design. *ETS Research Report Series* 1, i–29.
- Montanari, Elke (2016): Das Deutsche Sprachdiplom für neu zugewanderte Jugendliche im Handlungsmuster des Prüfungsgesprächs. In: Benholz, Claudia; Frank, Magnus & Niederhaus, Constanze (Hrsg.): *Neu zugewanderte Schülerinnen und Schüler – eine Gruppe mit besonderen Potentialen. Beiträge aus Forschung und Schulpraxis*. Münster: Waxmann, 285–298.
- Moss, Pamela A. (2013): Validity in Action: Lessons From Studies of Data Use. *Journal of Educational Measurement* 50: 1, 91–98.
- Neugebauer, Uwe & Becker-Mrotzek, Michael (2013): Die Qualität von Sprachstandsverfahren im Elementarbereich. Eine Analyse und Bewertung. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. http://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Mercator-Institut_Qualitaet_Sprachstandsverfahren_Web_03.pdf (27.01.2017).
- North, Brian (2014): *The CEFR in Practice*. Cambridge: Cambridge University Press.
- Pižorn, Karmen & Huhta, Ari (2016): Assessment in educational settings. In: Tsagari, Dina & Banerjee, Jayanti (Eds.): *Handbook of Second Language Assessment*. Boston: De Gruyter, 239–254.
- Pochon-Berger, Evelyne & Lenz, Peter (2014): *Sprachliche Anforderungen und Sprachtests im Migrations- und Integrationsbereich*. Université de Fribourg.
- Redder, Angelika; Naumann, Johannes & Tracy, Rosemarie (Hrsg.) (2015): *Forschungsinitiative Sprachdiagnostik und Sprachförderung – Ergebnisse*. Münster, New York: Waxmann.
- Reich, Hans H.; Roth, Hans-Joachim & Gantefort, Christoph (2008): Der Sturz ins Tulpenbeet. Deutsche Sprachversion. Auswertungsbogen und Auswertungshinweise. In: Klinger, Torsten; Schwippert, Knut & Leiblein, Birgit (Hrsg.): *Evaluation im Modellprogramm FörMig. Planung und Realisierung eines Evaluationskonzepts*. Münster u.a.: Waxmann, 209–237.
- Reich, Hans H.; Roth, Hans-Joachim & Döll, Marion (2009): Fast Catch Bumerang – Auswertungshinweise, Schreibimpuls und Auswertungsbogen. In: Lengyel, Dorit; Reich, Hans H.; Roth, Hans-Joachim & Döll, Marion (Hrsg.): *Von der Sprachdiagnose zur Sprachförderung*. Münster, München, Berlin u.a.: Waxmann, 207–241.

- Ricart Brede, Julia; Maak, Diana; Pliska, Enisa (Hrsg.) (2018): *Deutsch als Zweitsprache und Mehrsprachigkeit. Beiträge aus dem „Workshop Deutsch als Zweitsprache, Migration und Mehrsprachigkeit 2016“*. Stuttgart: Fillibach bei Klett.
- Ricart Brede, Julia & Schrage, Hannah (2018): Das Deutsche Sprachdiplom (DSD) als Motivationsturbo – sogar mit Nutzen für Regelunterricht und Berufsvorbereitung? Sichtweisen von schulischen Akteurinnen und Akteuren zum Einsatz des DSD I im deutschen Inlandsschulwesen. In: Ricart Brede, Julia; Maak, Diana; Pliska, Enisa (Hrsg.): *Deutsch als Zweitsprache und Mehrsprachigkeit. Beiträge aus dem „Workshop Deutsch als Zweitsprache, Migration und Mehrsprachigkeit 2016“*. Stuttgart: Fillibach bei Klett, 211–232.
- Riehl, Claudia Maria & Blanco López, Julia (2019): Mehrsprachiges Potenzial und die Transferierbarkeit sprachlicher Kompetenzen. In: Ender, Andrea; Greiner, Ulrike & Strasser, Margareta (Hrsg.): *Deutsch in mehrsprachigem Umfeld. Sprachkompetenzen auf Sekundarstufe*. Seelze, Zug: Klett und Kallmeyer, 306–319.
- Saleh, Sien-Lie (2019): *Potenzial/Perspektive. Ein Analyseverfahren für neu Zugewanderte*. <http://www.2p-bw.de/Lde/Startseite/2P-Bausteine/Aufgabenbeispiele> (15.10.2019).
- Saville, Nick (2009): *Developing a Model for Investigating the Impact of Language Assessments within Educational Contexts by a Public Examination Provider*. Bedfordshire: University of Bedfordshire. Dissertation.
- Saville, Nick & Khalifa, Hanan (2016): The impact of language assessment. In: Tsagari, Dina & Banerjee, Jayanti (Eds.): *Handbook of Second Language Assessment*. Boston: De Gruyter Mouton, 77–94.
- Schneider, Günther; Lenz, Peter & Forster Vosicki, Brigitte (2019): Gemeinsamer europäischer Referenzrahmen für Sprachen (GER) und Europäisches Sprachenportfolio (ESP). In: Forster Vosicki, Brigitte; Gick, Cornelia & Studer, Thomas (Hrsg.): *IDT 2017, Bd. 3. Sprachenpolitik: Expertenberichte und Freiburger Resolution*. Berlin: Erich Schmidt, 184–209.
- Schuth, Elisabeth; Heppt, Birgit; Köhne, Judith; Weinert, Sabine & Stanat, Petra (2015): Die Erfassung schulisch relevanter Sprachkompetenzen bei Grundschulkindern. Entwicklung eines Testinstruments. In: Redder, Angelika; Naumann, Johannes & R. Tracy, Rosemarie (Hrsg.): *Forschungsinitiative Sprachdiagnostik und Sprachförderung – Ergebnisse*, Münster, New York, Waxmann, 93–112.
- Shohamy, Elana (2001): *The power of tests: a critical perspective on the uses of language tests*. Harlow, New York: Longman.
- Shohamy, Elana (2007): Language tests as language policy tools. *Assessment in Education: Principles, Policy & Practice*. 14: 1, 117–130.

- Sievers, Isabel (2019): Bildungszugänge und -barrieren in der Migrationsgesellschaft. Aktuelle Ansätze, Diskurse und Beobachtungen im Kontext von Neumigration. In: Natajara, Radhika (Hrsg.): *Sprache, Flucht, Migration*. Wiesbaden: Springer, 235–305.
- Studer, Thomas (2020): Jetzt skaliert! Plurikulturelle und mehrsprachige Kompetenzen im erweiterten Referenzrahmen. *Deutsch als Fremdsprache* 1, 5–26.
- Taylor, Lynda (2005): Washback and impact. *ELT Journal* 59: 2, 154–155.
- Tsagari, Dina & Banerjee, Jayanti (Eds.) (2016): *Handbook of Second Language Assessment*. Boston: De Gruyter Mouton.
- Toledo, Heike & Mayer, Thomas (2014): Die Philosophie des Deutschen Sprachdiploms. In: ZfA (Hrsg.): *40 Jahre Deutsches Sprachdiplom der Kultusministerkonferenz*, 14–15.
- Universität Bamberg (2018): *BiSpra-Aufgaben: Weiterentwicklung zu einem diagnostisch nutzbaren Testinstrument und Prüfung der Sensitivität für Fördereffekte*. <https://www.uni-bamberg.de/entwicklungspsychologie/forschung/bispra/bispra-aufgaben/> (20.4.2019).
- Wilhelm, Oliver & Kunina, Olga (2009): Pädagogisch-psychologische Diagnostik. In: Wild, Elke & Möller, Jens (Hrsg.): *Pädagogische Psychologie*. Berlin u.a.: Springer, 305–328.
- Wisniewski, Katrin (2014): *Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen*. Frankfurt a.M. u.a.: Peter Lang.
- Wisniewski, Katrin (2019): Zur Skalierung von Strategien im Gemeinsamen europäischen Referenzrahmen: Eine kritische Auseinandersetzung. *Deutsch als Fremdsprache* 3, 142–151.
- Wisniewski, Katrin, Bärenfänger, Olaf & Tschirner, Erwin (2019): *Der Begleitband zum Gemeinsamen europäischen Referenzrahmen für Sprachen. Einleitung in den Themenschwerpunkt*. *Deutsch als Fremdsprache* 2, 67–76.
- Woerfel, Seda Yılmaz & Riehl, Claudia Maria (2016): Mehrschrittlichkeit: Wechselseitige Einflüsse von Textkompetenz, Sprachbewusstheit und außersprachlichen Faktoren. In: Rosenberg, Peter & Schroeder, Christoph (Hrsg.): *Mehrsprachigkeit als Ressource in der Schriftlichkeit*. Berlin, Boston: De Gruyter, 305–336.
- Zepter, Alexandra L. (2016). Zweitspracherwerb: Verläufe und diagnostische Aspekte. In: Heinrichs, Petra (Hrsg.): *Unterricht für neu zugewanderte Kinder und Jugendliche in der Sekundarstufe I. Fachliche und fachdidaktische Perspektiven*. Köln: Bezirksregierung Köln, 96–113.
- Zentralstelle für das Auslandsschulwesen [ZfA] (Hrsg.) (2009): *Rahmenplan Deutsch als Fremdsprache für das Auslandsschulwesen*. https://www.auslandsschulwesen.de/SharedDocs/Downloads/Webs/ZfA/DE/DSD/DaF-Rahmenplan.pdf?__blob=publicationFile&v=2 (20.08.2020).

- Zentralstelle für das Auslandsschulwesen [ZfA] (2017): *Anhang „Berufsorientiertes Deutsch“ zum Rahmenplan Deutsch als Fremdsprache für das Auslandsschulwesen*. https://www.auslandsschulwesen.de/SharedDocs/Downloads/Webs/ZfA/DE/DSD/DSDIPRO/Anhang-BerufsorientiertesDeutsch.pdf?__blob=publicationFile&v=2 (15.07.2019).
- Zentralstelle für das Auslandsschulwesen; Kulturministerkonferenz & Auswärtiges Amt [ZfA] (2018|2019): *Ausführungsbestimmungen zur Prüfungsordnung für die Prüfungen zum Deutschen Sprachdiplom der Kultusministerkonferenz. Beschluss der Kultusministerkonferenz vom 13.09.2018 in der Fassung vom 16.07.2019*. https://www.auslandsschulwesen.de/SharedDocs/Downloads/Webs/ZfA/DE/DS/Ausf%C3%BChrungsbestimmungenzurPr%C3%BCfungsordnung.pdf?__blob=publicationFile&v=2 (22.11.2019).
- Zentralstelle für das Auslandsschulwesen & Kultusministerkonferenz [ZfA] (2018a): *Prüfungsordnung für die Prüfungen zum Deutschen Sprachdiplom der Kultusministerkonferenz. Beschluss der KMK vom 13.09.2018*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2018/2018_09_13-DSD-Pruefungsordnung.pdf (15.10.2019).
- Zentralstelle für das Auslandsschulwesen [ZfA] (2018b): *Deutsches Sprachdiplom der Kultusministerkonferenz. Handreichungen für die Mündliche Kommunikation*. https://www.auslandsschulwesen.de/SharedDocs/Downloads/Webs/ZfA/DE/DSD/DSDI/HandreichungenDSDI/download_A2-B1_MK.pdf?__blob=publicationFile&v=4 (22.11.2019).
- Zentralstelle für das Auslandsschulwesen [ZfA] (2019): *Deutsches Sprachdiplom der Kultusministerkonferenz. Handreichungen für die Schriftliche Kommunikation im DSD I*. https://www.auslandsschulwesen.de/SharedDocs/Downloads/Webs/ZfA/DE/DSD/DSDI/HandreichungenDSDI/download_A2-B1_B2-C1_SK.pdf?__blob=publicationFile&v=6 (22.11.2019).
- Zentralstelle für das Auslandsschulwesen [ZfA] (2019a): *Das Deutsche Sprachdiplom (DSD)*. https://www.auslandsschulwesen.de/Webs/ZfA/DE/Deutsch-lernen/DSD/dsd_node.html (7.11.2019).

Kurzbio:

Dr. Katrin Wisniewski ist akademische Assistentin am Herder-Institut der Universität Leipzig und vertritt derzeit die Professur für Deutsch als Fremdsprache an der Justus-Liebig-Universität Gießen. Sie promovierte zur Validität der Skalen des GeR und befasst sich mit sprachdiagnostischen Themen, dem Zusammenhang von Sprachkompetenzen und Studienerfolg sowie mit der Lernerkorpuslinguistik.

Robin Ide ist Wissenschaftlicher Mitarbeiter im Projekt UniSpracheDeutsch des Academic Lab | Wissenschaftslabor der Universität Leipzig und Projektmanager am Institut für Testforschung und Testentwicklung e.V. in Leipzig. In seiner Promotion beschäftigt er sich mit den Konsequenzen des Einsatzes standardisierter Sprachtests am Beispiel des Inlandseinsatzes des Deutschen Sprachdiploms I.

Matthias Schwendemann ist Wissenschaftlicher Mitarbeiter am Herder-Institut der Universität Leipzig und am Max-Planck-Institut für Kognitions- und Neurowissenschaften in Leipzig. Er beschäftigt sich im Rahmen seiner Promotion mit dem Erwerb und der Entwicklung syntaktischer Kompetenzen bei erwachsenen Lernenden mit Deutsch als Zweitsprache und Arabisch als Ausgangssprache.

Anschrift:

Katrin Wisniewski

Robin Ide

Matthias Schwendemann

Herder-Institut, Universität Leipzig

Beethovenstr.15, 04107 Leipzig

robin.ide@uni-leipzig.de

matthias.schwendemann@uni-leipzig.de

katrin.wisniewski@uni-leipzig.de