



Bundesamt für
Auswärtige Angelegenheiten
Zentralstelle für das Auslandsschulwesen



Deutsches Sprachdiplom

der Kultusministerkonferenz

Handreichungen für die
Schriftliche Kommunikation



DSD I-Länderprojekt Frankreich

Handreichungen für die Schriftliche Kommunikation im DSD I

DSD I-Länderprojekt Frankreich

Inhaltsverzeichnis

1. Einführung in den Prüfungsteil Schriftliche Kommunikation im DSD I	S. 4
1.1 Allgemeine Hinweise zum Prüfungsteil Schriftliche Kommunikation (SK)	S. 4
1.2 Zielgruppe und Prüfungsziele	S. 4
1.3 Aufgabenformat	S. 6
1.4 Beispielaufgabe	S. 6
1.5 Bewertungsverfahren	S. 7
1.6 Operationalisierung der Bewertungskriterien im DSD I	S. 8
1.7 Zusätzliche Hinweise und häufig auftretende Bewertungsprobleme	S. 13
1.8 Verzerrung von Bewertungen	S. 14
2. Prüfungs- und Bewertungsunterlagen	S. 16
2.1 Bewertungsbogen für den Prüfungsteil Schriftliche Kommunikation - DSD I	S. 16
2.2 Bewertungskriterien für den Prüfungsteil Schriftliche Kommunikation - DSD I	S. 17
Anhang: Literaturhinweise	S. 18

1. Einführung in den Prüfungsteil Schriftliche Kommunikation im DSD I

1.1 Allgemeine Hinweise zum Prüfungsteil Schriftliche Kommunikation (SK)

Die Bewertung schriftlicher Prüfungsleistungen stellt eine besondere Herausforderung dar. Denn die Bewertung von Leistungen, die aufgrund offener Itemformate erbracht werden, wird durch Faktoren beeinflusst, die das Urteil verzerren können, so dass die Testgütekriterien, also die Validität, aber auch die Reliabilität und damit schlussendlich die Fairness beeinträchtigt werden. Zu diesen Faktoren zählen individuelle Maßstäbe, Vorlieben sowie andere Persönlichkeitsfaktoren.

Um dennoch Qualitätsstandards einzuhalten, sind bestimmte Maßnahmen erforderlich. Hierzu zählen

- die Standardisierung des Testformats,
- das kriterienorientierte Bewerten sowie
- entsprechende Schulungen.

Diese qualitätssichernden Maßnahmen sind ganz besonders für so genannte High-Stakes-Tests wichtig, also – wie im Falle des DSD – bei Prüfungen, von deren Ergebnis Entscheidungen abhängen, die für den weiteren Lebensweg der Prüflinge von Bedeutung sind.

In den vorliegenden Handreichungen werden die

die Zielgruppe sowie die Prüfungsziele skizziert und das Aufgabenformat eingehend erläutert.

Zudem werden das Bewertungsverfahren und die Bewertungskriterien definiert, um sicherzustellen, dass die Bewerterinnen und Bewerter zuverlässige und valide Einstufungen vornehmen.

1.2 Zielgruppe und Prüfungsziele

Beim DSD handelt es sich um eine standardisierte Prüfung, die weltweit durchgeführt wird und als High-Stakes-Test weitreichende Auswirkungen auf die Prüflinge, aber auch auf Curricula, Lehrmaterialien und Unterrichtsmethoden hat (Rückwirkungseffekt).

Aus diesen Gründen ist die Einhaltung zentraler Testgütekriterien von großer Bedeutung. So muss nicht nur sichergestellt werden, dass die Validität und die Zuverlässigkeit des Tests gewährleistet werden. Vielmehr gilt es auch, die (ggf. negativen) Auswirkungen der Prüfung zu berücksichtigen – z. B. Prüfungsdrill, bei dem es nicht um das Lernen an sich, sondern ausschließlich um die Vorbereitung auf die Prüfung geht.

Die DSD I-Prüfung richtet sich an Schüler und Schülerinnen, deren Schulen das DSD anbieten. Die Zielgruppe besteht dabei aus Jugendlichen im Alter von ca. 14 bis 16 Jahren. Die Lernenden haben zum Zeitpunkt der Prüfung ca. 600 - 800 Stunden Deutschunterricht à 45 Min. durchlaufen.

Die Prüfung besteht aus vier Prüfungsteilen: Leseverstehen, Hörverstehen, Schriftliche Kommunikation und Mündliche Kommunikation. Jeder dieser Subtests ist mit 25% gleich gewichtet.

Im Prüfungsteil *Schriftliche Kommunikation* sollen die Prüflinge zeigen, dass sie in der Lage sind, anhand von Vorgaben und Teilaufgaben einen zusammenhängenden Text zu schreiben, der die Merkmale der geforderten mitteilungsbezogenen Textsorte aufweist. Die Prüflinge sollen dabei zu einem Thema bestimmte Schreibhandlungen bzw. Sprachfunktionen realisieren (wiedergeben / paraphrasieren, beschreiben / berichten / darstellen, die eigene Meinung äußern und einfach begründen).

Die Anforderungen der DSD I-Prüfung entsprechen den Niveaustufen A2 und B1 des GeR¹, so dass die folgenden Skalen zugrunde gelegt werden können, wobei die Prüfung nur dann als bestanden gilt, wenn in allen vier Kompetenzen das Niveau B1 erreicht wurde:

Schriftliche Produktion allgemein (GeR S. 67):

B1	Kann unkomplizierte, zusammenhängende Texte zu mehreren vertrauten Themen aus ihrem/seinem Interessengebiet verfassen, wobei einzelne kürzere Teile in linearer Abfolge verbunden werden.
A2	Kann eine Reihe einfacher Wendungen und Sätze schreiben und mit Konnektoren wie 'und', 'aber' oder 'weil' verbinden.

Kreatives Schreiben (GeR S. 67f.):

B1	Kann unkomplizierte detaillierte Beschreibungen zu einer Reihe verschiedener Themen aus ihrem/seinem Interessengebiet verfassen. Kann Erfahrungsberichte schreiben, in denen Gefühle und Reaktion in einem einfachen zusammenhängenden Text beschrieben werden. Kann eine Beschreibung eines realen oder fiktiven Ereignisses oder einer kürzlich unternommenen Reise verfassen. Kann eine Geschichte erzählen.
A2	Kann in Form verbundener Sätze etwas über alltägliche Aspekte des eigenen Umfelds schreiben, wie z. B. über Menschen, Orte, einen Job oder Studiererfahrungen. Kann eine sehr kurze, elementare Beschreibung von Ereignissen, vergangenen Handlungen und persönlichen Erfahrungen verfassen. Kann in einer Reihe einfacher Sätze über die eigene Familie, die Lebensumstände, den Bildungshintergrund oder die momentane oder vorige berufliche Tätigkeit schreiben. Kann kurze, einfache fiktive Biographien und einfache Gedichte über Menschen schreiben.

Schriftliche Interaktion allgemein bzw. Korrespondenz (GeR S. 86):

B1	Kann Informationen und Gedanken zu abstrakten wie konkreten Themen mitteilen, Informationen prüfen und einigermaßen präzise ein Problem erklären oder Fragen dazu stellen.
	Kann in persönlichen Briefen und Mitteilungen einfache Informationen von unmittelbarer Bedeutung geben oder erfragen und dabei deutlich machen, was er/sie für wichtig hält.
A2	Kann kurze, einfache, formelhafte Notizen machen, wenn es um unmittelbare notwendige Dinge geht.

Korrespondenz (GeR S. 86):

B1	Kann in einem persönlichen Brief Neuigkeiten mitteilen und Gedanken zu abstrakten oder kulturellen Themen (z. B. Musik, Film) ausdrücken.
	Kann einen persönlichen Brief schreiben und darin detailliert über Erfahrungen, Gefühle, Ereignisse berichten.
A2	Kann einen ganz einfachen persönlichen Brief schreiben und sich darin für etwas bedanken oder entschuldigen.

¹ Europarat, Rat für kulturelle Zusammenarbeit (Hrsg.): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin u.a.: Langenscheidt 2001, zugänglich unter: www.goethe.de/referenzrahmen

1.3 Aufgabenformat

Der Prüfungsteil Schriftliche Kommunikation dauert 60 Minuten. Die Prüflinge sollen in dieser Zeit die Aufgabenstellung erfassen und die Aufgabe umsetzen. Für die Erstellung des Textes steht Konzeptpapier zur Verfügung, das ggf. verwendet werden kann (aber nicht verwendet werden muss). Der auf dem Konzeptpapier stehende Text darf jedoch nicht bewertet werden. Bei der Bewertung ist nur der auf offiziellem Schreibpapier verfasste Text zu berücksichtigen.

Es handelt sich um eine weitgehend gesteuerte Aufgabe, bei der verschiedene Vorgaben zu verarbeiten und Teilaufgaben zu bearbeiten sind.

Die Aufgabenstellung verlangt eine angemessene Bearbeitung aller drei Teilaufgaben. Dabei muss sich der Text des Prüflings nicht an der vorgegebenen Abfolge der Teilaufgaben orientieren, sondern der Prüfling kann diese in einer von ihm selbst bestimmten Reihenfolge bearbeiten.

1.4 Beispielaufgabe: Brauchen wir Hausaufgaben?

In einem Internetforum gibt es eine Diskussion zum Thema „Brauchen wir Hausaufgaben?“. Dazu findest du hier folgende Aussagen:

Tim: Ich finde Hausaufgaben sehr wichtig. Da kann ich zuhause noch einmal in Ruhe üben und sehen, ob ich alles verstanden habe.

Laura: Hausaufgaben sind sinnlos. Wenn ich in der Schule etwas nicht verstanden habe, lerne ich es auch nicht, wenn ich es zu Hause nur wiederhole.

Klara: Wenn ich von der Schule nach Hause komme, habe ich überhaupt keine Lust mehr auf Hausaufgaben und Schule. Man braucht doch auch seine Freizeit.

Peter: An unserer Schule machen wir unsere Hausaufgaben in der Schule. Es gibt dafür spezielle Übungsstunden am Nachmittag. Da kann auch ein Lehrer helfen. Das finde ich gut.

Schreibe einen **Beitrag** für die Schülerzeitung deiner Schule.

Bearbeite in deinem Beitrag die folgenden drei Punkte:

- Gib alle vier Aussagen aus dem Internetforum **mit eigenen Worten** wieder.
- Wie sieht es an deiner Schule mit Hausaufgaben aus? Berichte **ausführlich**.
- Sind Hausaufgaben sinnvoll? Begründe deine Meinung **ausführlich**.

Du hast insgesamt **60 Minuten** Zeit.
Du brauchst die Wörter **nicht** zu zählen.

1.5 Bewertungsverfahren

Die handschriftlich fixierten Leistungen der Prüflinge werden kriterienorientiert, d. h. anhand eines festgelegten Kriterienrasters, bewertet. Das Kriterienraster besteht aus skalierten Deskriptoren. Die individuelle Leistung wird in Bezug zu den skalierten Deskriptoren und den zugrunde liegenden Kompetenzbeschreibungen des GeR gesetzt.

Die Schreibleistungen werden von speziell geschulten Bewerterinnen und Bewertern hinsichtlich der vorgegebenen Kriterien und nach einem standardisierten Bewertungsverfahren eingestuft. Beim Bewertungsverfahren gilt daher ein besonderes Augenmerk dem Training der Bewerterinnen und Bewerter. Schulungen und Kalibrierungen sollen dafür sorgen, dass die Bewertungskriterien weitgehend homogen interpretiert und die Maßstäbe möglichst in gleicher Weise bei der Leistungsbeurteilung angewendet werden.

Das Ergebnis wird mit Hilfe statistischer Verfahren ermittelt. Vor jeder Bewertungsphase finden so genannte Kalibrierungen mit dem Ziel statt, die Bewerterinnen und Bewerter zu „eichen“. Dabei erhalten alle Bewerterinnen und Bewerter zusätzlich zu ihrem „individuellen“ Bewertungsauftrag über 30 – 60 SK-Arbeiten eine Anzahl identischer SK-Arbeiten zur aktuellen Prüfungsaufgabe, so genannte Vergleichsarbeiten. Anhand der vorgenommenen Bewertung dieser Vergleichsarbeiten wird der Grad der Strenge bzw. Milde der jeweiligen Bewerterinnen und Bewerter ermittelt. Bei der Berechnung der Ergebnisse der SK-Arbeiten gehen dann zum einen die einzelnen Punktwerte ein, die die Bewerterin bzw. der Bewerter festgesetzt hat – die „Rohwerte“ –, zum anderen wird das Ergebnis dann entsprechend des Strengegrades der Bewerterin bzw. des Bewerter korrigiert, so dass alle Schreibleistungen auf Grundlage des ermittelten Bewertungsdurchschnitts bewertet werden.

Mit diesen Maßnahmen soll erreicht werden, dass bei der konkreten Bewertung möglichst einheitliche Maßstäbe angelegt werden, um ein hohes Maß an Reliabilität sicherzustellen.

Für die Zuordnung der Schreibleistung zu einem Niveau gilt die folgende Regel:

Niveaustufe	Punktzahl
B1 (DSD I)	12 - 24
A2	8 - 11
unter A2	0 - 7

1.6 Operationalisierung der Bewertungskriterien im DSD I

Die Bewertung der im Bereich *Schriftliche Kommunikation* angefertigten Texte erfolgt anhand der folgenden Aspekte:

Gesamteindruck

Beim Kriterium *Gesamteindruck* handelt es sich um ein holistisches Kriterium, bei dem der Eindruck des Textganzen auf die Leserin bzw. den Leser bewertet wird, ohne dass Einzelaspekte wie z. B. grammatische Korrektheit im Fokus stehen. Denn gerade der holistische Aspekt gewährleistet eine valide Bewertung des in realen Sprachverwendungssituationen relevanten Ersteindrucks (Stichwort: kommunikative Funktion). Eine Bewertung erfolgt deshalb nach dem ersten Lesen.

Die Deskriptoren dieses Kriteriums beschreiben die kommunikative Qualität der Leistung, indem die beiden Merkmale Textzusammenhang und Flüssigkeit der Lektüre betrachtet werden. Bewertet wird also, in welchem Ausmaß der Text die gewünschte kommunikative Funktion erfüllt.

Die zu bewertenden Merkmale sind:

- Flüssigkeit
- Textzusammenhang

Dabei werden sowohl sprachliche als auch inhaltliche Aspekte berücksichtigt. Es ist z. B. möglich, dass ein gut strukturierter Text aufgrund seiner sprachlichen Mängel nur mit Stocken rezipiert werden kann; im umgekehrten Fall kann ein Text sprachlich auf hohem Niveau verfasst sein, aber strukturelle/logische Schwächen aufweisen. In beiden Fällen ist keine Höchstbewertung abzugeben.

Bewertet wird hier, ob der Prüfling in der Lage ist, zu einem vorgegebenen Thema einen zusammenhängenden Text zu verfassen, der in der Rezeption durch die Leserin bzw. den Leser gut funktioniert.

Merkmal *Flüssigkeit*

Im Merkmal *Flüssigkeit* wird der Grad der Ungestörtheit der Lektüre, also der Lesefluss bewertet. Entscheidend ist hier, wie der Text in der ersten Lektüre funktioniert, ob er plausibel und eingängig wirkt bzw. in welchem Maße er die Kooperation der Leserin bzw. des Lesers benötigt. Bei der Bewertung ist Folgendes zu berücksichtigen:

- SK-Leistungen auf B1-Niveau erfordern eine gewisse Kooperation seitens der Leserin bzw. des Lesers.
- Eine Abwertung muss dann vorgenommen werden, wenn sich der Lesefluss über diese erforderliche Kooperationsleistung hinaus an mehreren Stellen spürbar verlangsamt (2 Punkte).
- Gerät der Lesefluss ins Stocken, d.h. müssen bestimmte Textpassagen **mehrfach** gelesen werden, dann kann nur 1 Punkt vergeben werden.
- Ist ein Text größtenteils unverständlich, so ist er mit 0 Punkten zu bewerten.

Da eine Beeinträchtigung des Leseflusses ebenso wie die leichte Lesbarkeit des Textes sich aus schlechten bzw. guten Leistungen in den übrigen Kriterien (wie z. B. Grammatische Korrektheit) ergibt, findet hier auf jeden Fall eine **Doppelsanktionierung** statt, die in den anderen Kriterien zu vermeiden ist.

Merkmal *Textzusammenhang*

Im Merkmal *Textzusammenhang* wird zum einen der Grad der Strukturiertheit auf der Ebene der Makrostruktur des Textes (Einleitung, die die situative Einbettung des Textes verdeutlicht, gegliederter Hauptteil und ein z.B. durch die Meinungsäußerung realisierter Schlussteil), zum anderen der Grad der Schlüssigkeit der Gedankenführung bewertet. Bei letzterem geht es insbesondere um den kohärenten Aufbau des Textes. Eine optimale Leistung ist in diesem Merkmal dann gegeben, wenn der Text sinnvoll eingeleitet und abgeschlossen wird, die verschiedenen Textpassagen nicht isoliert nebeneinander stehen, die Binnenstruktur in den einzelnen Aufgabenteilen funktioniert und die Leseführung so angelegt ist, dass die Leserin bzw. der Leser die Funktion einer Textpassage klar erfassen kann.

Störungen der Nachvollziehbarkeit innerhalb einer Teilaufgabe werden unter dem jeweiligen inhaltlichen Kriterium sanktioniert. Beispiel: In Teilaufgabe 2 – *Wie sieht es an deiner Schule mit Hausaufgaben aus?* – ist ein einzelner Aspekt nicht nachvollziehbar dargestellt, ansonsten ist aber die entsprechende Passage gut in den Gesamttext eingebettet, die „übergeordnete“ Intention (Schilderung der Situation an der eigenen Schule) wird deutlich. In diesem Fall wird der Mangel im Kriterium *eigene Erfahrungen* sanktioniert.

ACHTUNG: Bezug zur Schreibsituation (Thema, Referenz mit Quellenangabe) und *Textsorte*

Die Art, wie die in der Aufgabenstellung eingeforderte Textsorte umgesetzt wird, wird ebenfalls im Kriterium *Gesamteindruck* berücksichtigt:

Durch die situative Einbettung ist ein Rahmen für die erwartete Textsorte (Leserbrief oder Artikel) geschaffen. Es wird erwartet, dass in der SK-Arbeit einleitend auf die Situation bzw. auf das Thema Bezug genommen wird, z. B. in Form einer Einleitung und/oder einer Anrede/Überschrift. Die Textsorteneinbettung (Anrede/Überschrift, Themennennung, Quellenbezug, Referenz) muss angemessen gegeben sein. Die Art des Bezugs auf die Schreibsituation wird hier bewertet. Fehlt die Textsorteneinbettung gänzlich, muss im Kriterium *Gesamteindruck* bei der Endbewertung insgesamt ein Punkt abgezogen werden.

Inhalt

Das Kriterium *Inhalt* wird durch die drei Subkriterien *Wiedergabe*, *eigene Erfahrungen* und *eigene Meinung* ausdifferenziert. Jedes Subkriterium wird gesondert bewertet. Dabei gibt die Aufgabenstellung ausdrücklich vor, dass alle drei Teilaufgaben angemessen behandelt werden sollen.

Dabei können alle drei Teilaufgaben in beliebiger Reihenfolge und in beliebiger Kombination bearbeitet werden. Folglich wird weder eine bestimmte Abfolge in der Bearbeitung der Teilaufgaben verlangt noch deren separate Behandlung.

Unter dem Kriterium *Inhalt* werden sowohl quantitative als auch qualitative Aspekte der inhaltlichen Umsetzung erfasst: Zum einen wird bewertet, inwiefern die angegebenen Teilaufgaben in ausreichendem Maße bearbeitet wurden, zum anderen, inwiefern die geforderten Schreibhandlungen angemessen umgesetzt sind. Sprachliche Aspekte treten bei diesem Kriterium in den Hintergrund.

Die drei Teilaufgaben sind in der Aufgabenstellung nicht nummeriert, weil die Reihenfolge der Bearbeitung nicht vorgegeben ist. Es darf daher nicht sanktioniert werden, wenn die Schülerinnen und Schüler ihren individuellen Zugang zum Thema finden, also beispielsweise mit der letzten Teilaufgabe beginnen statt mit der ersten. Ausschlaggebend ist allein, dass alles in ausreichendem Maße und in angemessener Weise behandelt wird.

Bei der inhaltlichen Umsetzung ist unwichtig, ob die Bewerterin bzw. der Bewerter inhaltlich mit den Positionen des Prüflings übereinstimmt; es wird nur bewertet, ob diese schlüssig/nachvollziehbar entwickelt sind. Auch wenn es manchmal schwer fällt: Wir bewerten nicht die Meinungen der Prüf-

linge. Auch sachliche Fehler sollten nur dann ein Problem sein und entsprechend sanktioniert werden, wenn sie auf eine unzureichende Verarbeitung der Vorgaben (Texte) gründen.

Bei der Bewertung ist nach Möglichkeit darauf zu achten, nicht über Gebühr Weltwissen bei den Lernenden vorauszusetzen. Prinzipiell, und dies ist auch Aufgabe der Testerstellung, gilt es, durch die Aufgabenstellung ausreichend Material zur Verfügung zu stellen, um die Aufgabe auch ohne tiefer gehendes Vorwissen in ausreichendem Maße bewältigen zu können. Ist das bei einer Aufgabe nur eingeschränkt der Fall, so darf dies nicht zu Lasten der Prüflinge bewertet werden. Inwiefern jedoch Wissen zum Allgemeinwissen gehört, das in verschiedenen Ländern der Erde vorausgesetzt werden darf, ist ein letztlich nicht festzulegender Aspekt, der stets bei der Entwicklung neuer Aufgaben zu berücksichtigen ist.

1. Subkriterium *Wiedergabe*

Bei dem Subkriterium *Wiedergabe* wird der Grad der *Angemessenheit* und der Grad der *Eigenständigkeit* der Reproduktion der vier in der Aufgabenstellung formulierten Textaussagen bewertet. Dabei müssen Textaussagen als solche sprachlich markiert werden, da sie andernfalls nicht als Fremdaussagen erkannt werden können und mit Äußerungen des Prüflings vermengt werden.

Merkmal Angemessenheit

Eine angemessene Wiedergabe liegt dann vor, wenn alle vier Textaussagen inhaltlich richtig wiedergegeben wurden. Dabei wird erwartet, dass die wichtigen Äußerungen der vier Jugendlichen zum Thema, wenn auch nicht mit sämtlichen Details, wiedergegeben werden. Der einzelne Jugendliche muss aufgrund der Wiedergabe identifizierbar sein. Die Namen der Jugendlichen müssen dabei nicht genannt werden.

Eine genaue Beschreibung der zu erwartenden Wiedergabeleistung wird in den Spezifischen Hinweisen zur jeweiligen SK-Aufgabe einer Prüfung geliefert.

Merkmal Eigenständigkeit

Das Merkmal *Eigenständigkeit* umfasst die Bereiche Lexik und Syntax. Die Anforderung, Textaussagen sprachlich eigenständig wiederzugeben, verlangt also zum einen, dass eigene Formulierungen gefunden werden. Das bedeutet nicht, dass für jedes Wort ein Synonym verwendet werden muss, aber ein Abschreiben ganzer Textstellen ist nicht zulässig. Werden die Aussagen in wörtlicher Rede angeführt, so ist das nicht als eine eigenständige Leistung des Prüflings zu bewerten.

Der geringste Grad syntaktischer Eigenständigkeit besteht darin, die in der ersten Person formulierten Aussagen in die dritte Person zu transformieren, indem Verben der Redeeinleitung (meinen, sagen, berichten usw.) und dass-Sätze verwendet werden. In diesem Fall liegt nur eine begrenzt eigenständige Leistung vor. Liegt eine Transformation im obigen Sinne vor, kann auch eine inhaltlich vollständige und richtige Wiedergabe nur mit einem Punkt bewertet werden.

Formen strukturierter Textwiedergabe, zusammenfassende und abstrahierende Formen der Textdarstellung sind eigenständige Strukturierungsleistungen und werden positiv bewertet, sofern sie inhaltlich richtig sind. Beschränkt sich jedoch die Wiedergabe auf eine solche knappe Zusammenfassung wie „Laura und Klara sind gegen Hausaufgaben“, dann sind die Anforderungen nicht erfüllt, denn die Identifizierbarkeit der Aussagen ist nicht gegeben.

Der Konjunktiv I bei der Redewiedergabe kann auf dem Niveau B1 **nicht** erwartet werden.

Allgemein gilt: Werden mindestens 50% der in den spezifischen Hinweisen als erforderlich angegebenen Teilaussagen richtig wiedergegeben, wird für das Merkmal Angemessenheit 1 Punkt vergeben.

Wenn im Merkmal Angemessenheit 0 Punkte vergeben werden, kann die Gesamtpunktzahl für das Kriterium Wiedergabe durch gute sprachliche Leistungen (Merkmal Eigenständigkeit) nicht angehoben werden.

Die Bewertung der Wiedergabe erfolgt ansonsten immer unter Mittelung der beiden Merkmale Angemessenheit und Eigenständigkeit.

2. Subkriterium *eigene Erfahrungen*

Bei dem Subkriterium *eigene Erfahrungen* wird der Grad der Detailliertheit und Nachvollziehbarkeit der im Aufsatz dargestellten eigenen Erfahrungen bewertet. Es geht also zum einen um ein eher quantitatives Bewertungskriterium, das die Ausführlichkeit der Darstellung fokussiert, zum anderen um ein qualitatives Bewertungskriterium, das die logische Kohärenz der vermittelten Erfahrungen ins Auge fasst. Die Nachvollziehbarkeit ist in diesem Kriterium nicht in erster Linie oder nicht allein als sprachliche Verständlichkeit zu fassen, sondern vor allem als eine Verständlichkeit, die aus einer sinnvollen inhaltlichen Kontextualisierung der gelieferten Informationen erwächst.

Wenn die eigenen Erfahrungen in nur einem Hauptsatz abgehandelt werden, wird weder das Merkmal der Ausführlichkeit noch der Nachvollziehbarkeit erfüllt. Beispiel: „An meiner Schule gibt es manchmal viele, manchmal wenige Hausaufgaben.“ Die Frage, wann und warum es Unterschiede hinsichtlich des Hausaufgabenpensums gibt, bleibt ungeklärt. Somit ist die Nachvollziehbarkeit dieser äußerst knappen Aussage deutlich beeinträchtigt. In diesem Fall können nur 0 Punkte vergeben werden.

3. Subkriterium *eigene Meinung*

Bei dem Subkriterium *eigene Meinung* geht es zum einen um die Darstellung der eigenen Meinung, zum anderen um den Grad der Begründetheit des eigenen Standpunkts. Auf A2- und B1-Niveau wird vorausgesetzt, dass die eigene Meinung nachvollziehbar dargestellt wird. Die eigene Meinung ist dabei mit entsprechenden Aussagen/Beispielen zu begründen bzw. zu veranschaulichen.

Für eine das A2-Niveau überschreitende Leistung ist der Grad der Begründetheit der Meinungsäußerung bewertungsrelevant. Die Verdeutlichung der eigenen Meinung ohne argumentative Stützung kann maximal mit einem Punkt (A2) bewertet werden. Eine differenziert entwickelte Argumentation mit These, Begründung, Belegen und Schlussfolgerungen kann auf dem Niveau B1 nicht erwartet werden.

Auch die Aussage, keine Meinung zum Thema zu haben bzw. sich nicht klar für oder gegen etwas positionieren zu können, ist als Meinungsäußerung zu werten!

Verfügbarkeit sprachlicher Mittel

Das Kriterium *Verfügbarkeit sprachlicher Mittel* wird durch die Subkriterien *Wortschatz* und *Strukturen* konkretisiert. Dabei wird die Korrektheit der verwendeten sprachlichen Mittel im Hinblick auf die lexikalischen Mittel im Subkriterium *Wortschatz* selbst, im Hinblick auf die Morphosyntax im Kriterium *Korrektheit* unter dem Subkriterium *Grammatische Korrektheit* bewertet.

Jedes Subkriterium wird gesondert bewertet. Wenn in einer Prüflingsarbeit kaum eigenes Sprachmaterial vorhanden ist, kann die Leistung in Bezug auf die beiden Subkriterien ggf. nicht beurteilt werden und muss dann entsprechend mit 0 Punkten bewertet werden.

Sprachliche Mittel werden im Hinblick auf ihre Funktionalität bei der Bearbeitung der Aufgabenstellung bewertet.

1. Subkriterium *Wortschatz*

Bei dem Subkriterium *Wortschatz* geht es um den Grad der im Aufsatz realisierten lexikalischen Kompetenz. Während auf B1-Niveau davon ausgegangen werden kann, dass mit dem hier vorhandenen Wortschatz auch abstraktere Sachverhalte ausgedrückt werden können, reicht der Wortschatz auf dem darunter liegenden Niveau nur dazu aus, konkrete Alltagssituationen zu versprachlichen.

Fehler in der Wortverwendung und beim Gebrauch der Präpositionen werden in diesem Subkriterium mitbewertet (s. o.).

Bei der Bewertung des Umfangs und der Angemessenheit des Wortschatzes hinsichtlich der Aufgabenbearbeitung sollte die dem Prüfling zur Verfügung stehende Textvorlage zur Abgrenzung des prüflingseigenen Sprachmaterials mit einbezogen werden. Auf diese Weise lässt sich die Bandbreite des verwendeten Wortschatzes erfassen.

2. Subkriterium *Strukturen*

Bei dem Subkriterium *Strukturen* wird der Grad der Komplexität sowie das Spektrum der verwendeten sprachlichen Strukturen bewertet. Auf B1-Niveau bedeutet dies, dass der Prüfling überwiegend noch mit einfachen/elementaren Strukturen arbeitet. Komplexe Strukturen kommen gelegentlich vor. Auch fehlerhaft umgesetzte Strukturen werden unter diesem Kriterium positiv bewertet, da sie zeigen, dass der Prüfling diese Strukturen kennt und praktisch anwendet. Die Korrektheit der verwendeten morphosyntaktischen Strukturen wird unter dem Kriterium *Korrektheit* bewertet (s. o.).

Zu komplexen Strukturen auf B1-Niveau zählen u.a.:

- mehrgliedrige Hypotaxen
- mehrgliedrige Konnektoren
- indirekte Fragen
- Passiv und Passiversatz
- Futur I, Präteritum, Plusquamperfekt
- Konjunktiv II (auch „würde“-Form)
- (erweiterter) Infinitiv mit „zu“
- Partizip I
- erweiterte Nominalgruppen (z.B. mit Genitiv)
- Indefinitpronomen und Pronominaladverbien
- Relativpronomen mit Präposition

Korrektheit

Das Kriterium *Korrektheit* untergliedert sich in die Subkriterien *Grammatische Korrektheit* und *Orthografische Korrektheit*.

1. Subkriterium *grammatische Korrektheit*

Bei dem Subkriterium *Grammatische Korrektheit* geht es um den Grad der Beherrschung der Grammatik sowie um den Grad der Verständlichkeit eines Textes mit morphosyntaktischen Mängeln. Dabei wird davon ausgegangen, dass auf der Stufe B1 bei einem sicheren Umgang mit der Grammatik die noch vorhandenen grammatischen Fehler nicht mehr die Verständlichkeit des Textes beeinträchtigen. Auf der Niveaustufe A2 ist hingegen der Erwerb elementarer grammatischer Strukturen noch nicht abgeschlossen bzw. deren Verwendung noch fehlerhaft.

Auch wenn die in einem Text verwendeten komplexen Strukturen fehlerfrei sind, die einfachen Strukturen jedoch elementare Mängel aufweisen, kann keine Höchstbewertung abgegeben werden. Finden sich z.B. in einem Text mehrere fehlerfreie Passagen, so ist dies ein Hinweis auf eine gute Beherrschung von grammatischer Korrektheit.

Hinweis: Artikelfehler werden in diesem Subkriterium bewertet!

2. Subkriterium *orthografische Korrektheit*

Bei dem Subkriterium *Orthografische Korrektheit* geht es um den Grad der Beherrschung der Orthografie **und** der Interpunktion. Es wird davon ausgegangen, dass im Rahmen eines schulischen Fremdspracherwerbs die Aneignung der Rechtschreib- und Interpunktionsregeln auf der Niveaustufe B1 weitgehend abgeschlossen ist. Dies bedeutet nicht, dass wir in diesem Bereich einem gänzlich fehlerfreien Text begegnen. Finden sich z.B. in einem längeren Text mehrere längere fehlerfreie Passagen, so ist dies ein Hinweis auf eine gute Beherrschung von Orthographie und Interpunktion. In die Bewertung der Beherrschung orthografischer Korrektheit fließt zudem das Merkmal *Verständlichkeit* ein.

1.7 Zusätzliche Hinweise und häufig auftretende Bewertungsprobleme

Gewichtung der Kriterien

Alle acht Kriterien (Gesamteindruck, Inhalt usw.) sind gleich gewichtet. Dies gilt prinzipiell auch für die einem Kriterium zugeordneten Einzelmerkmale. Allerdings sind diese Einzelmerkmale in den Schreibleistungen nicht immer auf dem gleichen Niveau realisiert. Beispielweise können in einem Text die eigenen Erfahrungen umfangreich, aber teilweise nicht ganz nachvollziehbar dargestellt werden. In einem solchen Falle werden die Ergebnisse im Subkriterium *eigene Erfahrungen* **gemittelt**. Im Einzelfall liegt es im Ermessen und in der Verantwortung der Bewerterin bzw. des Bewerbers, eine angemessene Entscheidung zu treffen.

Doppelsanktionierung

Bei einer kriterienorientierten Bewertung besteht stets die Gefahr, Merkmale einer Leistung unter unterschiedlichen Kriterien zu bewerten.

Prinzipiell gilt: Die Risikobereitschaft der Lernenden darf nicht „bestraft“ werden. Ein kreativer und mutiger Umgang mit Sprache muss belohnt werden, denn Prüflinge sollen zeigen, was sie können und keine „minimalistischen“ Texte produzieren (Stichwort: Fehlervermeidungsstrategien). Andernfalls könnte die Bewertung negative Rückwirkungseffekte in der Form haben, dass sich die Lernenden, aber auch die Lehrkräfte auf Vermeidungsstrategien konzentrieren. Aus der Risikobereitschaft resultierende Fehler müssen jedoch in dem Kriterium „Korrektheit“ angemessen sanktioniert werden.

Umfang der Arbeit

Eine bestimmte Wortanzahl wird nicht gefordert. Der Umfang der Arbeit wird durch die Ausführlichkeit, in der auf die Teilaufgaben eingegangen wird, bestimmt.

Kein Leistungsvergleich

Die schriftlichen Leistungen der einzelnen Prüflinge dürfen bei der Bewertung nicht verglichen werden.

Zentrale Grundlage der Bewertung sind allein die Kriterien.

Bewertungsbogen vs. Markierungen im Text

Im Text **dürfen keine** Markierungen vorgenommen bzw. Korrekturzeichen verwendet werden. Auch muss Abstand davon genommen werden, Fehler zu zählen. Der Ansatz des Fehlerzählens ist negativ und steht in Widerspruch zum GeR und seinem kommunikativen Ansatz. Im Bewertungsbogen darf daher nicht die Fehleranzahl, sondern es sollen Zeilennummern bzw. Beispiele festgehalten werden, auf die sich die Bewertung stützt. Dies ist wichtig, erstens zur eigenen Disziplinierung bei der Bewertung, und zweitens, um ggf. in Zweifelsfällen die Bewertung für andere nachvollziehbar zu machen.

DaF-Hintergrund der Bewerterinnen und Bewerter

„Wir verstehen immer, was der Dichter uns sagen will“ – diese *déformation professionnelle* sollte bei der Bewertung von Prüfungsleistungen nach Möglichkeit bedacht werden. Das bedeutet: Bewerterinnen und Bewerter sollten aufgrund ihres DaF-Hintergrundes nicht zu kooperativ rezipieren und über Mängel in einer sprachlichen Äußerung hinwegsehen, die eine Person ohne entsprechende Erfahrung mit Lernenden und Interimssprache nicht verstehen würde, denn die Aussage über die Leistung ist hinsichtlich des Testkonstrukts dann nur noch eingeschränkt valide.

1.8. Verzerrung von Bewertungen

Trotz des standardisierten und kriterienorientierten Bewertungsverfahrens treten bei der Bewertungsarbeit unerwünschte Effekte auf, die zu Verzerrungen führen. Sie sollen im Folgenden kurz erläutert werden.

Besonders kritisch sind Persönlichkeitsfaktoren (affektive Faktoren) sowie akute Dispositionen wie Tagesform, Konzentrationsfähigkeit, aber auch externe Konditionen, unter denen die Bewertungsarbeit geleistet wird, wie akustische Bedingungen etc. So kann die individuelle Strenge bzw. die individuelle Wahrnehmung einer Leistung zu unterschiedlichen Tageszeiten schwanken; Pausen und (unfreiwillige) Unterbrechungen der Bewertungsarbeit können zur Verschiebung von Maßstäben führen; äußere ebenso wie psychische Faktoren beeinflussen Konzentration und damit Bewertungsmaßstäbe. Einen nicht zu unterschätzenden Einfluss haben zudem subjektive Theorien, persönliche Gewichtungen und Vorlieben, die nicht zuletzt durch die Berufserfahrung geprägt sind.

Der so genannte Halo-Effekt stellt sich dann ein, wenn von demselben Prüfling bereits andere Leistungen (Teilaspekte, Teilfertigkeiten) bewertet wurden und diese Urteile die weiteren Urteile mitbestimmen, indem sich beispielsweise eine negative oder positive Erwartung aufgrund nicht relevanter Faktoren konstituiert. Gerade bei analytischem Vorgehen besteht die Gefahr des Halo-Effekts. Die Bewertung eines isolierten Aspekts (z. B. der grammatischen Korrektheit) beeinflusst die Wahrnehmung anderer Aspekte (z. B. die Angemessenheit der Wortwahl). So kann beispielsweise ein als gut bewerteter Gesamteindruck dazu führen, dass auch die folgenden analytischen Aspekte tendenziell höher eingestuft werden.

Als Kontaminationseffekt wird ein Bewertungsfehler bezeichnet, der auf Idiosynkrasien oder Präferenzen zurückzuführen ist. Bei schriftlichen Leistungen beispielsweise kann die allgemeine Präsentation des Textes sowie die Handschrift das Urteil beeinflussen. Besonders stark machen sich diese Effekte bei mündlichen Prüfungen (face-to-face-Prüfungssituation) bemerkbar, da Sympathie, Auftreten etc. das Urteil der Prüfenden beeinflussen – oft ohne dass dieser Effekt bewusst wird. Gerade im schulischen Kontext, wo sich Prüfende und Geprüfte oftmals sehr gut kennen, sind solche Effekte besonders virulent.

Beim Positionseffekt (auch Reihenfolgen- oder serialer Effekt) handelt es sich um einen Messfehler, der sich aus der Position der zu bewertenden Arbeit im Kontext anderer zu bewertender Arbeiten ergibt. Beispielsweise können die ersten Arbeiten aus einem Stapel zu bewertender Texte eher milde, die späteren hingegen eher streng bewertet werden (oder umgekehrt). Das bedeutet: Die bereits bewerteten Leistungen beeinflussen die Wahrnehmung der nachfolgenden Leistungen. Dies ist darauf zurückzuführen, dass offensichtlich die Leistungen untereinander verglichen werden, also

bewusst oder unbewusst auch Rangreihen gebildet werden (Text A ist besser als Text B). Dies ist bei einem kriterienorientierten Bewertungsverfahren jedoch unzulässig (s. o.), denn hierbei muss die individuelle Leistung unabhängig von den anderen Leistungen der (zufälligen!) Stichprobe bewertet werden und zwar im Hinblick auf **vor** dem Test festgesetzte Bewertungsmaßstäbe.

Oft tendieren Bewerberinnen und Bewerber dazu, Leistungen im mittleren Kompetenzbereich einzustufen, auch wenn es sich um Leistungen handelt, die im oberen oder unteren Bereich anzusiedeln sind. Dieses –Zentraltendenz genannte – Verhalten hat eine unerwünschte Verschiebung der Kompetenzstufen zur Folge: Im Falle der Zentraltendenz nehmen die mittleren Niveaustufen eine breitere, die obere sowie die untere Stufe hingegen eine schmalere Bandbreite ein, was folglich zu einer Inkonsistenz zwischen der tatsächlichen Leistung und der jeweils zugrunde liegenden Kann-Beschreibung führt. Wenn wir uns die Kompetenzstufen als Treppe vorstellen, dann führt die Zentraltendenz dazu, dass die Stufen unterschiedliche Höhen haben.

Die genannten Effekte können niemals vollständig kontrolliert werden. Ziel muss es jedoch sein, die Bewerberinnen und Bewerber zu sensibilisieren für das eigene Handeln und Verhalten. Diese Sensibilisierung ist Aufgabe von Schulungen und Kalibrierungen.

2. Prüfungs- und Bewertungsunterlagen

2.1 Bewertungsbogen für den Prüfungsteil Schriftliche Kommunikation – DSD I

Prüfling: _____

Bewerter/-in: _____

	Begründungen und Belege	Punkte
Gesamt- eindruck	Textzusammenhang	
	Flüssigkeit	
Inhalt	Wiedergabe Angemessenheit	
	Eigenständigkeit	
	eigene Erfahrungen Detailliertheit	
	Nachvollziehbarkeit	
Verfügbarkeit sprachlicher Mittel	eigene Meinung	
	Wortschatz	
	Strukturen	
Korrektheit	Grammatik	
	Orthografie Rechtschreibung	
	Interpunktion	
	Gesamtpunktzahl:	

2.2. Bewertungskriterien für den Prüfungsteil Schriftliche Kommunikation – DSD I

Kriterium		B1 3 Punkte	B1/A2 2 Punkte	A2 1 Punkt	unter A2 0 Punkte
Gesamteindruck		Der Text ist zusammenhängend und insgesamt flüssig zu lesen.	Der Text ist weitgehend zusammenhängend. Mehrere Textstellen lesen sich nicht flüssig.	Es liegen einfache Sätze vor, die inhaltlich verbunden sind. An mehreren Textstellen wird der Lesefluss deutlich unterbrochen.	Es liegen überwiegend Wendungen und Sätze vor, die weder inhaltlich noch sprachlich miteinander verbunden sind.
Inhalt	<i>Wiedergabe</i>	Die Aussagen werden angemessen und eigenständig wiedergegeben.	Die Aussagen werden weitgehend angemessen und weitgehend eigenständig wiedergegeben.	Die Aussagen werden nur zum Teil angemessen und nur zum Teil eigenständig wiedergegeben.	Die Aussagen werden nicht angemessen oder gar nicht wiedergegeben bzw. abgeschrieben.
	<i>eigene Erfahrungen</i>	Der Bericht enthält mehrere Aspekte, die detailliert und nachvollziehbar dargestellt werden.	Der Bericht enthält einige Aspekte, die knapp, aber insgesamt nachvollziehbar dargestellt werden.	Der Bericht ist insgesamt knapp und teilweise nicht ganz nachvollziehbar.	Der Bericht ist sehr kurz und kaum noch nachvollziehbar.
	<i>eigene Meinung</i>	Die eigene Meinung wird angemessen begründet.	Die eigene Meinung wird nur knapp begründet, ist aber noch nachvollziehbar.	Die eigene Meinung wird deutlich, aber nicht begründet.	Die eigene Meinung wird nicht geäußert bzw. ist unverständlich.
Verfügbarkeit sprachlicher Mittel	<i>Wortschatz</i>	Der Wortschatz ermöglicht eine angemessene Bearbeitung der Aufgabe. Wortfehler treten überwiegend im Zusammenhang mit einer komplexer angelegten Lexik auf.	Der Wortschatz lässt eine Bearbeitung der Aufgabe weitgehend zu. Wortschatzlücken und Wortfehler treten auch bei einer weniger komplex angelegten Lexik gelegentlich auf.	Der Wortschatz lässt eine Bearbeitung der Aufgabe nur begrenzt zu. Wortschatzlücken und Wortfehler treten häufiger auf.	Der Wortschatz ist so begrenzt, dass er nicht ausreicht, um die Aufgabe zu bearbeiten.
	<i>Strukturen</i>	Die Strukturen ermöglichen eine angemessene Bearbeitung der Aufgabe. Gelegentlich werden komplexe Strukturen verwendet.	Die Strukturen lassen eine Bearbeitung der Aufgabe weitgehend zu. Komplexe Strukturen kommen kaum vor.	Die Strukturen lassen eine Bearbeitung der Aufgabe nur begrenzt zu. Komplexe Strukturen kommen nicht vor.	Die Strukturen sind so begrenzt, dass sie nicht ausreichen, um die Aufgabe zu bearbeiten.
Korrektheit	<i>grammatische Korrektheit</i>	Die grammatischen Strukturen werden mit wenigen Ausnahmen korrekt verwendet. Diese Ausnahmen beeinträchtigen die Verständlichkeit nicht.	Einfache Strukturen werden überwiegend korrekt verwendet. Bei der Verwendung komplexer Strukturen kommen Fehler vor, die die Verständlichkeit beeinträchtigen können.	Einige einfache Strukturen werden korrekt verwendet, allerdings zeigen sich viele elementare Fehler. Es wird jedoch überwiegend klar, was ausgedrückt werden soll.	Auch wenn nur einfache Strukturen verwendet werden, ist der Text sehr fehlerhaft. Dadurch wird die Verständlichkeit an einigen Stellen beeinträchtigt.
	<i>orthografische Korrektheit</i>	Orthografie und Interpunktion sind weitgehend korrekt.	Orthografie- und Interpunktionsfehler kommen vor, ohne die Verständlichkeit zu beeinträchtigen.	Orthografie- und Interpunktionsfehler treten häufig auf, jedoch wird die Verständlichkeit kaum beeinträchtigt.	Orthografie und Interpunktion sind so fehlerhaft, dass die Verständlichkeit beeinträchtigt wird.

Anhang: Literaturhinweise

Literaturhinweise rund ums Testen und Bewerten

Referenzrahmen

Europarat (2001): Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen. Berlin et al.: Langenscheidt.

Glaboniat, M. et al. (2005): Profile deutsch. Berlin et al.: Langenscheidt

Lenz, P./ Studer, T. (2005): Neue Instrumente für die Beurteilung der Französisch- und Englischkompetenzen von Deutschschweizer Schülerinnen und Schülern. In: Gohard-Radenkovic, A. (ed.): Plurilinguisme et interculturalité dans la didactique des langues étrangères / Réflexions à partir d'un contexte bilingue // Mehrsprachigkeit und Interkulturalität in der Fremdsprachendidaktik / Überlegungen aus einem zweisprachigen Kontext. Lang

Leistungsmessung allgemein

Bachman, L. F. / Palmer, A. (1996): Language Testing in Practice. Oxford: Oxford University Press

Eckes, T. (2005). Analyse und Evaluation sprachproduktiver Prüfungen beim TestDaF. In I. Kühn, M. Lehker & W. Timmermann (ed.), Sprachtests in der Diskussion. Frankfurt: Lang, 60–93

Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. Language Assessment Quarterly, 2, 197–221

Grotjahn, R. (2006): Prüfen - Testen - Bewerten In: Jung, O. H. (ed.): Praktische Handreichung für Fremdsprachenlehrer (4., völlig neu bearb. Aufl.). Frankfurt am Main: Lang, 221-230

Hughes, A. (2003): Testing for Language Teachers. Second edition. Cambridge: Cambridge University Press

Lewkowicz, J. A. (2000): Authenticity in language testing: Some outstanding questions. In: Language Testing 17/19, 43-64

Milanovic, Michael (ed.) (1998): Multilingual glossary of language testing terms, Studies in Language Testing 6, Cambridge: Cambridge University Press

Kniffka, G. (2003): Prüfen und Bewerten. In: Bausch, K.-R./Christ, H./Krumm, H.-J. (eds.): Handbuch Fremdsprachenunterricht. 4. Aufl. Tübingen, Basel: Francke, 373-377

Vollmer, H. J. (2003): Leistungsmessung, Lernerfolgskontrolle, Selbstbeurteilung: Überblick. In: Bausch, K.-R./Christ, H./Krumm, H.-J. (eds.): Handbuch Fremdsprachenunterricht. 4. Aufl. Tübingen, Basel: Francke, 365-370

Testen der Leseverstehenskompetenz

Alderson, C. (2000). Assessing Reading. Cambridge: Cambridge University Press

Arras, U. (2006): Testen und Beurteilen des Leseverstehens in der Fremdsprache. In: *Babylonia* 3/2006, 81-86.

Grotjahn, R. (2000): Determinanten der Schwierigkeit von Leseverstehensaufgaben. In: Bolton, S. (ed.) *TestDaF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Experten-seminar.* München: Goethe-Institut: 7-56

Testen der Hörverstehenskompetenz

Buck, G. (2001): *Assessing Listening.* Cambridge: Cambridge University Press

Grotjahn, R. (2005): Testen und Bewerten des Hörverstehens. In: M. Ó Dúill/R. Zahn/K. Höppner (eds.), *Zusammenarbeiten: Eine Festschrift für Bernd Voss.* Bochum: AKS-Verlag 2005, 115-144

Paschke, P. (2001): Zum Problem der Authentizität in L2-Hörverstehenstests. In: Henrici, G. / Königs, F. / Zöfgen, E. (eds.): *FLuL, Leistungsmessung und Leistungsevaluation*, 30. Jahrgang

Solmecke, G. (2000): Faktoren der Schwierigkeit von Hörtests. In: *TestDaF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar,* Gilde Verlag, München

Testen schriftlicher Kompetenz

Arras, U. (2009): Wie es zu einer Beurteilung kommt: Ein Forschungsbericht zu Strategien bei der Beurteilung schriftlicher Leistungen im Kontext der Prüfung TestDaF. In: A. Hunstiger & U. Koreik (Hrsg.), *Chance Deutsch: Schule – Studium – Arbeitswelt (Materialien Deutsch als Fremdsprache, Bd. 78, S. 179–196).* Göttingen: Universitätsverlag

Weigle, S. C. (2000). *Assessing Writing.* Cambridge: Cambridge University Press

Testen mündlicher Kompetenz

Arras, U. (2006): Was macht eine Aufgabe eigentlich schwierig? Schwierigkeitsdeterminanten in Tests zur Überprüfung mündlicher Kompetenzen am Beispiel von TestDaF-Aufgaben des Subtests Mündlicher Ausdruck. In: M. Oesterreicher & R. Zahn (Hrsg.), *Lingua Franca – Lingua Academica: Mehrsprachigkeit im europäischen Hochschulraum (Dokumentation der 24. Arbeitstagung 2006 des AKS, S. 211–225).* Bochum: AKS-Verlag

Glenn F. (2003): *Testing Second Language Speaking.* London et al.: Pearson/Longman

Luoma, S. (2004): *Assessing Speaking.* Cambridge: Cambridge University Press

Wiesmann, B. (1999): *Mündliche Kommunikation im Studium : Diskursanalysen von Lehrveranstaltungen und Konzeptualisierung der Sprachqualifizierung ausländischer Studienbewerber.* München : Iudicium

Testen anderer Aspekte von Sprachkompetenz

Purpura, J. (2004): *Assessing Grammar;* Cambridge: Cambridge University Press

Read, J. (2000): *Assessing Vocabulary,* Cambridge: Cambridge University Press