



Bundesamt für
Auswärtige Angelegenheiten
Zentralstelle für das Auslandsschulwesen



Deutsches Sprachdiplom

der Kultusministerkonferenz

Handreichungen für die
Schriftliche Kommunikation



DSD II

Handreichungen für die Schriftliche Kommunikation im DSD II

Inhaltsverzeichnis

1. Einführung in den Prüfungsteil Schriftliche Kommunikation im DSD II (Niveaustufe B2/C1)	
1.1 Allgemeine Hinweise zum Prüfungsteil Schriftliche Kommunikation (SK)	6
1.2 Zielgruppe und Prüfungsziele	6
1.3 Aufgabenformat	8
1.4 Beispielaufgabe	9
1.5 Bewertungsverfahren	11
1.6 Operationalisierung der Bewertungskriterien im DSD II	11
1.7 Zusätzliche Hinweise und häufig auftretende Bewertungsprobleme	18
1.8 Verzerrung von Bewertungen	20
2. Prüfungs- und Bewertungsunterlagen	
2.1 Bewertungsbogen für die Schriftliche Kommunikation – DSD II	22
2.2 Ergebnisblatt für die Schriftliche Kommunikation – DSD II	24
2.3 Bewertungskriterien für die Schriftliche Kommunikation – DSD II	25
Anhang: Literaturhinweise	26

1. Einführung in den Prüfungsteil Schriftliche Kommunikation im DSD II (Niveaustufe B2/C1)

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

Zudem werden das Bewertungsverfahren und die Bewertungskriterien definiert, um sicherzustellen, dass die Bewerterinnen und Bewerber 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 Reliabilität des Tests gewährleistet werden. Zudem 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 II-Prüfung richtet sich an Schülerinnen und Schüler anerkannter DSD-Schulen. Die Zielgruppe besteht somit aus Jugendlichen im Alter von ca. 16 bis 18 Jahren. Die Lernenden haben zum Zeitpunkt der Prüfung ca. 800 – 1200 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 Textsorte (textbasierter argumentativer Sachtext) aufweist. Die Prüflinge sollen dabei zu einem Thema bestimmte Schreibhandlungen realisieren (wiedergeben, beschreiben, auswerten, begründen, abwägen, Stellung nehmen etc.).

Die Anforderungen der Prüfung entsprechen den Niveaustufen B2 und C1 des GeR¹ und GeR-Begleitbands², so dass die folgenden Skalen zugrunde gelegt werden können:

Schriftliche Produktion allgemein (GeR-Begleitband S. 81)

C1	Kann klare, gut strukturierte Texte zu komplexen Themen verfassen und dabei die entscheidenden Punkte hervorheben, Standpunkte ausführlich darstellen und durch Unterpunkte oder geeignete Beispiele oder Begründungen stützen und den Text durch einen angemessenen Schluss abrunden. Kann die Struktur und die Konventionen verschiedener Genres verwenden und dabei Ton, Stil und Register adressatenbezogen, textsorten- und themengerecht variieren.
B2	Kann klare, detaillierte Texte zu verschiedenen Themen aus dem eigenen Interessensgebiet verfassen und dabei Informationen und Argumente aus verschiedenen Quellen zusammenführen und gegeneinander abwägen.

Schriftliche Interaktion allgemein (GeR-Begleitband S. 101)

C1	Kann sich klar und präzise ausdrücken und sich flexibel und effektiv auf die Adressaten beziehen.
B2	Kann Neuigkeiten und Standpunkte effektiv schriftlich ausdrücken und sich auf solche von anderen beziehen.

Berichte und Aufsätze schreiben (GeR-Begleitband S.83):

C1	Kann klare, gut strukturierte Ausführungen zu komplexen Themen verfassen und dabei zentrale Punkte hervorheben. Kann Standpunkte ausführlich darstellen und durch Unterpunkte, geeignete Beispiele oder Begründungen stützen. Kann eine angemessene Einleitung und einen angemessenen Schlusssatz zu einem längeren Bericht, Artikel oder zu einer wissenschaftlichen Arbeit über ein komplexes wissenschaftliches oder berufliches Thema verfassen, sofern dieses Thema im eigenen Interessensgebiet liegt und es möglich ist, den Text neu zu schreiben oder zu überarbeiten.
B2	Kann einen Aufsatz oder Bericht verfassen, in dem etwas systematisch erörtert wird, wobei entscheidende Punkte angemessen hervorgehoben und stützende Details angeführt werden. Kann eine detaillierte Beschreibung eines komplexen Vorgangs verfassen. Kann verschiedene Ideen oder Problemlösungen gegeneinander abwägen. Kann in einem Aufsatz oder Bericht etwas erörtern, dabei Gründe für oder gegen einen bestimmten Standpunkt angeben und Vor- und Nachteile verschiedener Optionen erläutern. Kann Informationen und Argumente aus verschiedenen Quellen zusammenführen.

¹ Europarat (Hrsg.): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin u. a.: Langenscheidt 2001.

² Europarat (Hrsg.): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Begleitband. Stuttgart: Klett 2020.

1.3 Aufgabenformat

Der Prüfungsteil Schriftliche Kommunikation dauert 120 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 Prüflinge dürfen ein einsprachiges und/oder zweisprachiges Wörterbuch benutzen, das ihnen von der Schule zur Verfügung gestellt wird.

1.4 Beispielaufgabe

Deutsches Sprachdiplom der KMK DSD II Schriftliche Kommunikation Aufgabe

Vorbilder

Schreiben Sie einen **zusammenhängenden Text** zum Thema „Vorbilder“. Bearbeiten Sie in Ihrem Text die folgenden drei Punkte:

- Arbeiten Sie wichtige Aussagen aus dem Text heraus.
- Werten Sie die Grafik anhand von wichtigen Daten aus.
- Nehmen Sie in Form einer ausgearbeiteten Argumentation ausführlich zum Thema „Vorbilder“ Stellung.

Sie haben insgesamt **120 Minuten** Zeit.

Die modernen Helden der Jugend

von Gerd Stein

Das waren noch Zeiten, als Deutschlands Jugendliche Nelson Mandela mehr verehrten als Leonardo Di Caprio und Umweltaktivisten von Greenpeace angesehener waren als Sportler wie Michael Schumacher. Das galt noch 1999. Heute sehen die Vorbilder anders aus: Einer aktuellen Studie zufolge zählen neben Familie und Freunden der Fußballer Michael Ballack, die Teenie-Band Tokio Hotel und Topmodel Heidi Klum zu den größten Idolen der Jugend in Deutschland.

Was haben diese Vorbilder gemeinsam? Sie stehen alle im Zentrum der Öffentlichkeit. Über sie werden Artikel geschrieben, sie treten in Talkshows auf oder moderieren selber Sendungen, die wie „Deutschland sucht den Superstar“ oder „Germany’s Next Topmodel“ bei den Jugendlichen äußerst beliebt sind. Die Studie bestätigt diese Beobachtung: Viele

Vorbilder der heutigen Jugendlichen sind ohne Medien kaum denkbar.

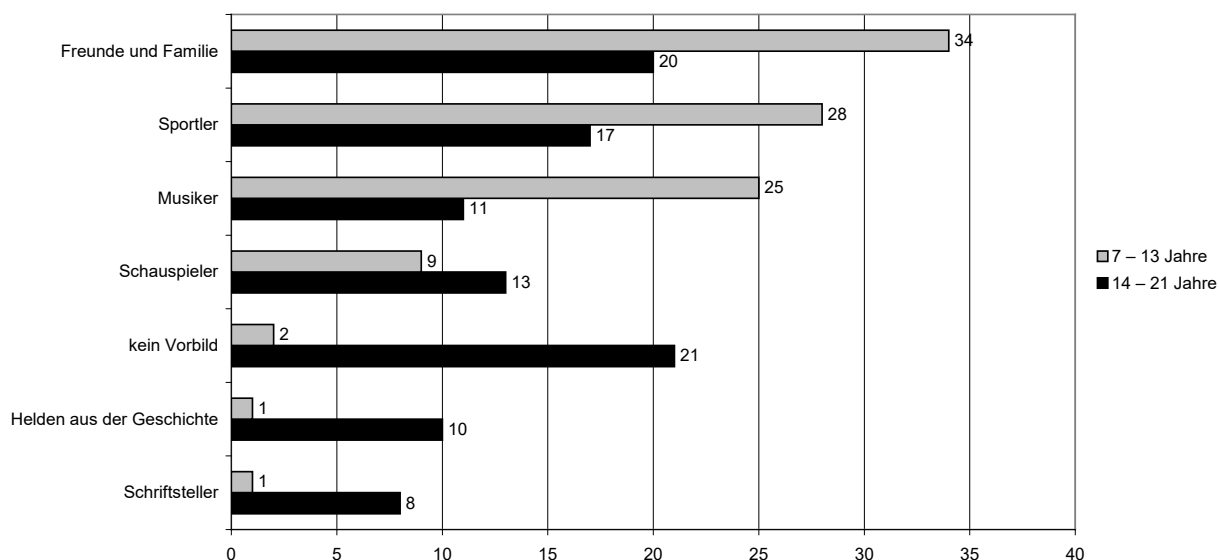
Bei der Wahl des Vorbilds spielt auch das Geschlecht eine bedeutende Rolle. So lässt sich beobachten, dass Mädchen eher Filmschauspielern und Sängern nacheifern, während Jungen eine klare Vorliebe für Sportler zeigen. Persönlichkeiten aus der Politik hingegen werden von Jugendlichen nur selten genannt.

Vorbilder sind nicht notwendigerweise positiv. Denn es stellt sich nicht nur die Frage, welche Werte sie verkörpern. Man muss außerdem bedenken, dass Wirtschaft, Medien und Politik Vorbilder häufig für ihre eigenen Zwecke benutzen.

Quelle: Städtische Rundschau vom 12. 02. 2010
(zu Prüfungszwecken bearbeitet)

Vorbilder von Kindern, Jugendlichen und jungen Erwachsenen

- Angaben in Prozent -



Quelle: Statistisches Landesbüro vom 12.02.2010

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 externen 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 Bewerber. Schulungen und Kalibrierungen sollen dafür sorgen, dass die Bewertungskriterien weitgehend homogen interpretiert und die Maßstäbe möglichst in gleicher Weise bei der Leistungsbewertung angewendet werden.

Das Ergebnis wird mit Hilfe statistischer Verfahren ermittelt. Vor jeder Bewertungsphase finden so genannte Kalibrierungen mit dem Ziel statt, die Bewerterinnen und Bewerber zu „eichen“. Dabei erhalten alle Bewerterinnen und Bewerber 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 Bewerber ermittelt. Bei der Berechnung der Ergebnisse der SK-Arbeiten gehen zum einen die einzelnen Punktwerte ein, die die Bewerberin bzw. der Bewerber festgesetzt hat – die „Rohwerte“ –, zum anderen wird das Ergebnis dann entsprechend des Strengegrades der Bewerberin bzw. des Bewerbers (sowie anderer statistisch ermittelter Werte) korrigiert, so dass alle Schreibleistungen auf Grundlage des ermittelten Bewertungsdurchschnitts fair 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
C1	12 – 24
B2	7 – 11
unter B2	0 – 6

1.6 Operationalisierung der Bewertungskriterien im DSD II

Die Bewertung erfolgt anhand der vier Kriterien *Gesamteindruck*, *Inhalt*, *sprachliche Mittel und Korrektheit*, die z. T. mithilfe verschiedener Subkriterien ausdifferenziert werden. Dabei weisen einzelne Subkriterien mehrere Merkmale auf. Die zu unterschiedlichen Merkmalen innerhalb eines Kriteriums erbrachten Leistungen sind selten identisch. So stößt man häufig auf SK-Arbeiten, bei denen etwa im Kriterium *Wortschatz* die Leistung hinsichtlich des *Wortschatzspektrums* anders ausfällt als die im Merkmal *Verwendung textsortenspezifischer Redemittel* oder im Merkmal *Wortschatzverwendung*. Die Bewerberin bzw. der Bewerber muss in einem solchen Fall eine Bewertung vornehmen, die alle drei Merkmale angemessen berücksichtigt.

Im Subkriterium *Gedankengang* und in allen inhaltlichen Subkriterien muss die Bewerberin bzw. der Bewerber bei sprachlich bedingten Verständnisschwierigkeiten kooperieren. Bei der *Textwiedergabe* und *Diagrammauswertung* ist jedoch darauf zu achten, dass man die Textwiedergabe ohne Textkenntnis und die Diagrammauswertung ohne Kenntnis der Grafik richtig verstehen können muss.

Generell muss gesagt werden, dass auch eine Schreibleistung, die sich sprachlich und inhaltlich auf C1-Niveau bewegt, nicht fehlerfrei ist. Demgemäß können über Flüchtigkeitsfehler hinaus in allen Kriterien Defizite vorhanden sein, die auf C1-Niveau noch nicht oder noch nicht konsequent vermieden werden können.

Gesamteindruck

Das Kriterium *Gesamteindruck* setzt sich aus einem holistischen (*Flüssigkeit*) und einem analytischen Subkriterium (*Gedankengang*) zusammen. Bei dem holistischen Subkriterium *Flüssigkeit* wird die Wirkung des Textes auf die Leserin bzw. den Leser bewertet, ohne dass Einzelaspekte wie z. B. Korrektheit direkt im Fokus stehen. Denn gerade der holistische Aspekt gewährleistet eine valide Bewertung des in realen Sprachverwendungssituationen relevanten Ersteindrucks (Stichwort: kommunikative Funktion). Eine Bewertung dieses holistischen Subkriteriums erfolgt deshalb nach dem ersten Lesen. Demgegenüber wird im analytischen Kriterium *Gedankengang* durch eine genaue Betrachtung der Makro- sowie der Binnenstruktur bewertet, inwieweit der Text einen Themenbezug hat, ob er in sich durchgehend strukturiert aufgebaut ist und ob die Textpassagen kohäsiv und kohärent entwickelt werden („roter Faden“).

Beide Subkriterien sind – soweit das Thema nicht verfehlt wird – **unabhängig** von der konkreten Aufgabenstellung zu bewerten. Ein Text etwa, der Text und Grafik als Material ohne jegliche Referenz auf die vorgegebenen Materialien in eine themenbezogene Erörterung integriert, kann in beiden Subkriterien mit 3 Punkten bewertet werden, sofern die Flüssigkeit der Lektüre und die Textkohärenz dies zulassen. Bewertet wird hier, ob der Prüfling in der Lage ist, zu einem vorgegebenen Thema einen integrierten Gesamttext zu verfassen, der in der Rezeption durch die Leserin bzw. den Leser gut funktioniert. Es wird nicht bewertet, ob eine exakte Aufgabenumsetzung vorliegt.

1. Subkriterium *Flüssigkeit*

Im Subkriterium *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:

- Auch SK-Leistungen auf C1-Niveau erfordern eine gewisse Kooperation seitens der Leserin bzw. des Lesers. Demzufolge ist eine Leistung auch dann mit 3 Punkten zu bewerten, wenn die Flüssigkeit nicht durchgängig gegeben ist.
- Eine Abwertung muss dann vorgenommen werden, wenn sich der Lesefluss über diese erforderliche Kooperationsleistung hinaus an einigen Stellen spürbar verlangsamt.
- Gerät der Lesefluss ins Stocken, d.h. müssen bestimmte Textpassagen **mehrfach** gelesen werden, können keine 2 Punkte mehr vergeben werden. Die Anzahl solcher Stellen entscheidet über die Vergabe von 1 Punkt oder von 0 Punkten.

Da eine Beeinträchtigung des Leseflusses ebenso wie die leichte Lesbarkeit des Textes sich aus schlechten bzw. guten Leistungen in den übrigen Kriterien ergibt, findet in diesem holistischen Subkriterium auf jeden Fall eine **Doppelsanktionierung** (siehe auch 1.7) statt, die in allen anderen Kriterien zu vermeiden ist.

2. Subkriterium *Gedankengang*

Im Subkriterium *Gedankengang* wird zum einen der Grad der Strukturiertheit auf der Ebene der Makrostruktur des Textes (Einleitung, gegliederter Hauptteil, Überleitungen und Schlussteil), zum anderen der Grad der formalen Schlüssigkeit der Gedankenführung innerhalb der einzelnen Aufgabenteile bzw. Textabschnitte bewertet. Der Gedankengang ist folglich nicht inhaltlich aufzufassen – dies wird in den inhaltlichen Subkriterien bewertet –, sondern formal im Sinne eines kohärent entfalteten Textes. Würde man den Gedankengang inhaltlich bewerten, käme es automatisch zu Doppelsanktionen, da dasselbe Phänomen auch im Kriterium *Inhalt* sanktioniert werden würde.

Eine optimale Leistung ist in diesem Subkriterium dann gegeben, wenn der Text sinnvoll eingeleitet und abgeschlossen wird, die verschiedenen Textpassagen gut miteinander verbunden werden, die Binnenstruktur in den einzelnen Textteilen formal funktioniert und die Leseführung so angelegt ist, dass die Leserin bzw. der Leser die kommunikative Funktion einer Textpassage klar erfassen kann.³ Dies bedeutet für die Makrostruktur im Einzelnen:

- a) Es muss eine zum Thema des Aufsatzes hinführende eigenständige Einleitung formuliert werden. Eine einfache Hinleitung zur Textwiedergabe etwa durch einen Basissatz ist unzureichend.
- b) Die Überleitungen sollen die Textteile sprachlich und inhaltlich miteinander verbinden. Mit Überleitungen, die nur darauf verweisen, was im nächsten Abschnitt behandelt werden soll (wie z.B.: Im Folgenden möchte ich ... / Jetzt komme ich ...), kann nicht die volle Punktzahl erzielt werden.
- c) Die SK-Arbeit muss einen eigenständigen Schluss (z. B. durch Zusammenfassung und/oder Schlussfolgerungen und/oder Ausblick) haben, durch den der Text abgerundet wird.

Inhalt

Das Kriterium *Inhalt* ist als analytisches Kriterium konzipiert. Hier werden sowohl quantitative als auch qualitative Aspekte der inhaltlichen Umsetzung der Aufgabenstellung 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 Behandlung nicht vorgegeben ist. Es darf daher nicht bestraft werden, wenn die Schülerinnen und Schüler ihren individuellen Zugang zum Thema auch im Bereich des Textaufbaus finden. Ausschlaggebend ist allein, dass die Aufgabenstellung in ausreichendem Maße und in angemessener Weise behandelt wird.

Bei der inhaltlichen Umsetzung ist es unwichtig, ob die Bewerterin bzw. der Bewerter mit den Positionen des Prüflings übereinstimmt; es wird nur bewertet, ob diese schlüssig/logisch entwickelt sind. Auch wenn es manchmal schwer fällt: Wir bewerten nicht die Meinungen der Prüflinge. Auch sachliche Fehler sollten nur dann ein Problem sein und entsprechend sanktioniert werden, wenn sie auf eine unzureichende oder inhaltlich falsche Verarbeitung der Vorgaben (statistische Daten, Texte) gründen.

Das Kriterium *Inhalt* wird durch die drei Subkriterien *Textwiedergabe*, *Diagrammauswertung* und *begründete Stellungnahme* spezifiziert, die jeweils einen Aspekt der dreigliedrigen Aufgabenstellung spiegeln.

1. Subkriterium *Textwiedergabe*:

Im Subkriterium *Textwiedergabe* wird bewertet, wie der Prüfling mit dem Vorlagentext umgeht. Für dieses Subkriterium gibt es folgende grundlegende Regelungen:

- a) Die Deskriptoren verlangen die vollständige Wiedergabe wichtiger Aussagen. Dies lässt dem Prüfling einen gewissen Spielraum, weil er nicht alle wichtigen Aussagen wiedergeben muss. Es wird somit keine Wiedergabe aller Textaussagen angestrebt. Vollständigkeit ist hier definiert **als inhaltliche Abdeckung aller Textabschnitte**. In diesem Sinne ist eine Wiedergabe defizitär, wenn ein ganzer Textabschnitt nicht wiedergegeben wird. Dies gilt auch dann, wenn andere Textabschnitte sehr differenziert präsentiert werden.

³ Probleme auf dieser Ebene treten häufig auf, wenn eine als Wiedergabe markierte Textwiedergabe nach der Diagrammauswertung weitergeführt wird, allerdings ohne als Wiedergabe signalisiert zu werden. In diesem Falle weiß die Leserin bzw. der Leser nicht, ob hier der Text wiedergegeben, eine eigene Meinung bzw. eigenes Wissen vorgeführt oder aber bereits mit der Erörterung begonnen wird.

- b) Werden die Basisinformationen zu Titel, Autor, Quelle, Thema, Erscheinungsdatum nicht oder teilweise nicht genannt, dann kann dies zu einer Abwertung um 1 Punkt im Bereich des Merkmals *Vollständigkeit* beitragen.
- c) Bei der Wiedergabe der Textaussagen müssen diese als Fremdaussagen sprachlich markiert werden, da sie andernfalls nicht als solche erkannt werden können und folglich als Äußerungen des Prüflings aufgefasst werden. Die Sprachhandlung „Wiedergabe fremder Aussagen“ wäre in diesem Falle nicht realisiert.

Wenn eine ansonsten alle Ansprüche erfüllende Textwiedergabe an **keiner** Stelle deutlich macht, dass hier fremde Aussagen wiedergegeben werden, dann kann die Textwiedergabe mit maximal 1 Punkt bewertet werden. Gibt es Mängel in einer solchen referenzlosen Darstellung des Textes, die normalerweise zu einem Abzug von 1 Punkt oder mehr führen würden, dann ist die Leistung in diesem Subkriterium mit 0 Punkten anzusetzen. Neben der vollständigen Wiedergabe wichtiger Textaussagen ist die eigenständige Wiedergabe das zweite bewertungsrelevante Merkmal in diesem Subkriterium. Eigenständigkeit verlangt auf der sprachlichen Ebene, dass der Prüfling die Textaussagen mit eigenen Strukturen und eigener Lexik darstellt. Dies bedeutet nicht, dass für jedes Wort ein Synonym verwendet werden muss, allerdings ist ein Abschreiben ganzer Textstellen nicht zulässig. **Kurze** wörtliche Übernahmen aus dem Text der Vorgabe sind erlaubt, wenn sie als Zitate kenntlich gemacht werden.

Eigenständigkeit kann zudem in der Strukturierung der Wiedergabe zum Ausdruck kommen, etwa wenn Aussagen aus verschiedenen Textpassagen sinnvoll zusammengezogen werden. Eine von der Textabfolge abweichende Strukturierungsleistung, die sinnvoll umgesetzt ist, ist an dieser Stelle positiv zu bewerten. Folgt die Wiedergabe nur dem Text, so ist dies nicht negativ zu bewerten.

- d) Die sprachliche und strukturelle Eigenständigkeit einer Textwiedergabe kann in der Bewertung **nicht** berücksichtigt werden, wenn im Merkmal *Vollständigkeit* die Textwiedergabe deshalb mit 0 Punkten bewertet wird, weil der Text **überwiegend** falsch wiedergegeben und/oder nicht wiedergegeben wurde. In diesem Falle kann die eigenständige sprachliche und/oder strukturelle Gestaltung nicht zu einer Anhebung der Bepunktung führen. Außerdem muss ein nahezu oder vollständig abgeschriebener Text mit 0 Punkten bewertet werden, auch wenn die Aussagen vollständig wiedergegeben sind.
- e) Kommentierende Passagen innerhalb der Textwiedergabe, die nicht als Kommentare kenntlich gemacht werden, sowie weiterführende Gedanken zum Text, die als solche nicht erkennbar sind und deshalb als Textaussagen aufgefasst werden können, werden im Rahmen der Wiedergabe negativ sanktioniert. Sie dürfen nicht im Rahmen des Subkriteriums *Gedankengang* sanktioniert werden, da im *Gedankengang* allein das Funktionieren des Textes insgesamt bewertet wird (siehe S. 9 f.).

Abschließend soll hervorgehoben werden, dass die Textwiedergabe im Unterschied zur klassischen Inhaltsangabe nicht nur sachlich distanziert, sondern auch kritisch engagiert abgefasst werden darf. Es ist folglich sowohl möglich, die Textaussagen in einer neutralen Form einfach nur zu referieren, als auch sie in einer kritisch oder zustimmend kommentierenden Form vorzustellen. Wichtig ist, dass alle über die Textwiedergabe hinausgehenden Aussagen als Aussagen des Prüflings kenntlich gemacht werden. Sollte die Textwiedergabe textanalytische Komponenten enthalten, also den Gedankengang des wiedergegebenen Textes beschreiben, so lässt sich dies nur positiv unter dem Merkmal *Eigenständigkeit* bewerten. Die Beschreibung des Gedankengangs und/oder des thematischen Textaufbaus ersetzen nicht die Wiedergabe der konkreten Textaussagen.

2. Subkriterium *Diagrammauswertung*

Bei der Bewertung der Grafikwiedergabe ist zu beachten: Im Zentrum dieses Subkriteriums steht nicht die Datenwiedergabe, sondern die Auswertungsleistung, die allerdings datengestützt durchgeführt werden muss. Es besteht deshalb keine Verpflichtung, alle Daten

als Text zu reproduzieren. Die Wiedergabe auswertungsrelevanter Basisdaten kann auch in den Auswertungsvorgang direkt integriert werden.

Für die Bewertung der Diagrammauswertung gelten folgende Orientierungspunkte:

- a) Werden die Basisinformationen zu Titel, Quelle, Thema oder Erscheinungsdatum nicht oder teilweise nicht genannt, dann kann dies zu einer Abwertung um 1 Punkt beitragen.
- b) Die Darstellung der Grafik in Form einer Wiedergabe von Basisdaten kann maximal mit 1 Punkt bewertet werden.
- c) Eine **sehr** differenzierte, fehlerfreie und präzise Wiedergabe der Daten einschließlich der Angabe von Quelle, Titel, Diagrammart kann mit 2 Punkten bewertet werden.
- d) Für 2 Punkte muss **zusätzlich** zu b) ein Vergleich zwischen den erhobenen Daten durchgeführt werden.
- e) Für 3 Punkte müssen zwei Vergleichsansätze entwickelt oder ein exemplarischer Vergleich mit Schlussfolgerung oder eine Interpretation bzw. ein Erklärungsansatz für die vorliegenden Daten ausformuliert werden. Die Entwicklung eines sinnvollen Erklärungsansatzes für die Grafik wird **nicht** gefordert, stellt allerdings die höchste interpretatorische Leistung dar und wäre folglich auf Basis einer auch knappen Diagrammbeschreibung mit 3 Punkten zu bewerten.
- f) Werden die Vergleiche nur formal durchgeführt (z. B. eine kontrastive Gegenüberstellung bloßer Daten: Während 73 % der befragten Männer für ... sind, sind nur 42 % der befragten Frauen dafür), können maximal nur 2 Punkte vergeben werden, selbst wenn die Daten durchgehend so präsentiert werden.
- g) Eine gute Auswertungsleistung erfordert keine vollständige Datenwiedergabe, um mit 3 Punkten bewertet zu werden. Eine gute Auswertungsleistung ohne jeden Datenbezug (allerdings sind konkrete Zahlen hierfür nicht unbedingt notwendig) kann jedoch maximal nur mit 2 Punkten bewertet werden.
- h) Schließlich muss bei der Leserin bzw. beim Leser eine angemessene Vorstellung von der Grafik hervorgerufen werden. Besteht diese zum Beispiel aus zwei Diagrammen, so kann eine Auswertung, die sich nur mit einem Diagramm in vorbildlicher Weise beschäftigt, nicht mit 3 Punkten bewertet werden.

3. Subkriterium *begründete Stellungnahme*

Für das Subkriterium *begründete Stellungnahme* muss die zentrale Bedeutung umfassender argumentativer Schreibhandlungen betont werden, die die entscheidende Differenz zur Niveaustufe B1 markieren. Gelungene beschreibende oder darstellende Textpassagen können demzufolge deutlichere argumentative Defizite nicht kompensieren.

Im Subkriterium *begründete Stellungnahme* wird allein der Grad der logischen Kohärenz sowie der argumentativen Entfaltung von Aussagen zum vorgegebenen Sachverhalt bewertet. Eine eigene Meinung muss auf jeden Fall deutlich werden.

Bei der Bewertung der SK-Arbeiten muss prinzipiell Folgendes beachtet werden:

- a) Die begründete Stellungnahme kann in Form einer dialektischen oder einer linearen Erörterung entfaltet werden. Beide Aufsatzformen stehen gleichberechtigt nebeneinander, weder die dialektische noch die lineare Erörterung ist an sich als höherwertig einzuschätzen. In beiden Aufsatzformen ist allein der Status der Argumentation relevant.
- b) Die dialektische Erörterung ist Ausdruck kontroversen Denkens, verlangt also vom Prüfling, sich in die von ihm nicht geteilte Haltung hineinzusetzen und diese argumentativ zu entfalten. In der linearen Erörterung entspricht dies der Widerlegung von Gegenargumenten bzw. der Begrenzung ihres Geltungsanspruchs. Fehlen solche

Argumentationsansätze in der linearen Erörterung, dann muss Argumentationsausbau und –umfang in besonderem Maße positiv ausfallen, um 3 Punkte vergeben zu können.

- c) Bewertungsrelevant ist in diesem Subkriterium allein die Entfaltung der Argumentation. Unter einer ausgebauten Argumentation wird verstanden, dass die Argumentation
- logisch widerspruchsfrei entwickelt wird und die Nachvollziehbarkeit nicht auf der Plausibilität des Gesagten basiert, sondern auf der Explizitheit des entwickelten Gedankengangs;
 - über These und Begründung – wobei die Ausdifferenziertheit der Begründung zu berücksichtigen ist – hinaus z. B. durch das Anführen von Belegen, die Relativierung von Geltungsansprüchen bzw. durch das Ziehen von begründeten Schlussfolgerungen ausgebaut und durch Beispiele oder andere Formen der Konkretion veranschaulicht wird.

Dabei muss nicht jedes Argument umfangreich entfaltet sein. Der Prüfling muss vielmehr nachweisen, dass er komplex argumentieren kann. Eine logisch kohärente und ausdifferenzierte Argumentation ist auch dann noch mit 3 Punkten zu bewerten, wenn der Grad der Anschaulichkeit gering ist.

d) Die eigene Meinung/Haltung zur Fragestellung muss auf jeden Fall deutlich werden. Sollte es zu keiner Meinungsäußerung kommen, muss 1 Punkt abgezogen werden. Dem Prüfling bleibt es in der dialektischen Erörterung freigestellt, ob er die eigene Meinung noch einmal begründet oder diese als Konsequenz aus seiner Erörterung herleitet. Eine dialektische Erörterung mit begründeter eigener Meinung ist nicht von vornherein besser als eine dialektische Erörterung, bei der die eigene Meinung nicht noch einmal gesondert begründet wird. Allerdings müssen die bei einer begründeten Meinung aufgeführten Argumente bei der Bewertung der Argumentation in der dialektischen Erörterung berücksichtigt werden, sofern diese über die in der Erörterung präsentierten Argumente hinausgehen.

Verfügbarkeit sprachlicher Mittel

Bei diesem Kriterium wird unter analytischer Herangehensweise überprüft, welche sprachlichen Mittel verfügbar sind. Dies bezieht sich auf die in der Aufgabenstellung geforderten Sprachhandlungen (z. B. wiedergeben, beschreiben, auswerten, begründen, abwägen, Stellung nehmen). Dieses Kriterium überprüft zudem, ob die für die geforderte Textsorte angemessenen Merkmale (z.B. der sachlich-argumentative Sprachgestus) vorhanden sind. Bei dem Subkriterium *Wortschatz* wird neben der Verfügbarkeit auch die Korrektheit der lexikalischen Mittel berücksichtigt. Demgegenüber beschränkt sich die Bewertung innerhalb des Subkriteriums *Strukturen* allein auf die Frage der Verfügbarkeit. Etwaige grammatische Fehler werden **allein** im Kriterium *Korrektheit* bewertet.

1. Subkriterium *Wortschatz*

In dem Subkriterium *Wortschatz* wird die Differenziertheit des Themenwortschatzes, die Verwendung textsortenspezifischer Redemittel sowie die Präzision bzw. die Korrektheit der eingesetzten lexikalischen Mittel bewertet. Alle drei Aspekte dieses Subkriteriums gehen zu je einem Drittel in die Bewertung ein.

Indikatoren für einen geringen Grad der Differenziertheit des verwendeten Wortschatzes sind z. B.

- Wortwiederholungen;
- die Verwendung von bedeutungsschwachen Verben (z. B. machen, tun, sein, haben) anstatt bedeutungsstarker Verben;

- Paraphrasen, sofern diese nicht bewusst als rhetorische Mittel eingesetzt werden, und ihre Genauigkeit;
- Wortschatzlücken.

Indikatoren für einen hohen Grad lexikalischer Differenziertheit sind z. B.

- idiomatische Wendungen;
- fachsprachliche und/oder fremdsprachliche Lexeme.

Unter textsortenspezifischen Redemitteln fassen wir alle im Zusammenhang mit der Markierung des Aktes der Textwiedergabe und Grafikauswertung (z. B. *der Autor behauptet, die Grafik verdeutlicht, dass...* usw.) verwendeten Operatoren, alle für die Vermittlung der Basisinformationen zu Text und Grafik erforderlichen Redemittel (*veröffentlicht in, Titel, Quelle* usw.), alle für die Grafikbeschreibung und -analyse relevanten Lexeme (*Balkendiagramm, Befragter, Prozentsatz* usw.) sowie Redemittel, die für die Textkohärenz und die Argumentation (*einerseits ... andererseits ...; daraus resultiert, dass...; außerdem* usw.) sowie der Formulierung der eigenen Meinung (*meiner Meinung nach, ich bin der Auffassung* usw.) eingesetzt werden.

Das Merkmal Präzision des sprachlichen Ausdrucks bezieht sich ausschließlich auf Fehler bei der Wortwahl. Ist ein Text beispielsweise aufgrund von vielen bedeutungsschwachen Verben lexikalisch unpräzise, so wird dies unter dem Merkmal Differenziertheit des Themenwortschatzes sanktioniert.

Im Merkmal Präzision des sprachlichen Ausdrucks bzw. Wortfehler wird auch die richtige bzw. fehlerhafte Verwendung aller Präpositionen bewertet. Die Artikelverwendung wird ausschließlich unter Korrektheit bewertet. Wird ein Wort fehlerhaft verwendet, so muss der Grad der Störung durch die falsche Wortwahl bei der Bewertung mit einbezogen werden. Wortfehler oder fehlerhafte Verwendung von Präpositionen, die das Verständnis nicht tangieren, können erst bei gehäuftem Auftreten zu einer deutlichen Abwertung in diesem Merkmalsbereich führen.

2. Subkriterium *Strukturen*

In dem Subkriterium *Strukturen* wird der Grad der Verfügbarkeit komplexer grammatischer Muster bewertet. Dabei wird nicht bewertet, ob die vorhandenen Mittel korrekt verwendet werden, da die sprachliche Richtigkeit der Mittel im Kriterium „Korrektheit“ behandelt wird.

Zu den komplexen Strukturen zählen

- hypotaktische Strukturen, die über die einfache Verbindung von Haupt- und Nebensatz hinausgehen und auch (erweiterte) Infinitivkonstruktionen umfassen können;
- Variation der Satzbaumuster durch Inversionen, durch voran- bzw. nachgestellte und eingeschobene Nebensätze und durch unterschiedliche Formen der Verb- rektion;
- Verwendung von Pronominaladverbien und Indefinitpronomen;
- komplexe Formen der Modalität (z. B. subjektiver Gebrauch von Modalverben);
- komplexe Nominalgruppen (z. B. durch Genitivattribute oder Appositionen und präpositionale Rechtsattribute sowie durch Linksattribute)
- erweiterte attributive und / oder prädikative Partizipialkonstruktionen;
- Passivkonstruktionen, insbesondere Passiv mit Modalverben, sowie Formen des Passiversatzes;

- Konjunktivkonstruktionen;
- mehrteilige Konnektoren

Ein Text kann auch dann als komplex bewertet werden, wenn nur ein Teil der aufgelisteten komplexen Strukturen in ihm zu finden sind. Wichtig ist, darauf hinzuweisen, dass nicht alle komplexen sprachlichen Formen in allen Textteilen präsent sein müssen, um einer Arbeit die Beibehaltung eines hohen Maßes an komplexen Strukturen bescheinigen zu können.

Generell ist hinzuzufügen, dass die Bewertung der verwendeten Sprachstrukturen als obersten Maßstab die kommunikativen Textfunktionen besitzt. Demgemäß ist es nicht sinnvoll, vorschreiben zu wollen, welche Sprachstrukturen unbedingt notwendig sind, um einen gut funktionierenden sachlich darstellenden und argumentativen Text zu verfassen.

Korrektheit

Hier wird die Richtigkeit grammatischer Strukturen – ebenfalls analytisch – bewertet. Dabei wird davon ausgegangen, dass der Text keine systematischen Grammatikfehler mehr aufweist. Wenn Unsicherheiten bestehen, ob eine SK-Arbeit in diesem Kriterium mit 2 oder mit 1 Punkt zu bewerten ist, kann die Fehlerverteilung als Kriterium hinzugezogen werden: Liegt ein passagenweiser fehlerfreier Text vor, dann ist dieser Text bei geringer Gesamtfehlerzahl mit 2 Punkten zu bewerten. Dies wäre ein Indikator dafür, dass in der SK-Arbeit trotz der noch vorhandenen Grammatikfehler „ein hohes Maß an grammatischer Korrektheit“ gegeben ist.

Der Deskriptor für 2 Punkte ist so aufzufassen, dass ein hohes Maß an Korrektheit die Gesamtarbeit kennzeichnet, auch wenn es in einzelnen Passagen zu gehäuftem Fehlern kommt. Es geht hier also nicht einfach um die Gesamtfehlerzahl, sondern um die Verteilung der Fehler über den Text.

Bei der Bewertung der Grammatikfehler ist zudem zu berücksichtigen, ob die Fehler aus einem grammatischen Bereich stammen oder aus unterschiedlichen. Texte etwa, die deutliche Defizite nur in der Deklination aufweisen, in anderen Feldern (z. B. Konjugation oder Satzstellung) allerdings relativ fehlerfrei sind, können noch mit 1 Punkt bewertet werden, auch wenn der Text (aufgrund der Deklinationsfehler) fast durchgehend fehlerhaft ist.

Auf den Niveaustufen B2 und C1 wird unterstellt, dass die Rechtschreibung und Interpunktion beherrscht wird. Sollte ein Text deutliche Mängel in diesen Bereichen aufweisen, ist das Ergebnis im Kriterium *Korrektheit* um 1 Punkt abzuwerten.

1.7 Zusätzliche Hinweise und häufig auftretende Bewertungsprobleme

Unvollständige Aufgabenbearbeitung

Das Fehlen von einem oder zwei Aufgabenteilen (Textwiedergabe, Diagrammauswertung, begründete Stellungnahme) hat grundsätzlich keinen Einfluss auf die Bewertung der Kriterien Flüssigkeit und Gedankengang sowie aller sprachlichen Kriterien (Wortschatz, Strukturen, Korrektheit). So bemisst sich beispielsweise die Bewertung des Gedankengangs allein danach, ob ein integrierter und gut strukturierter Gesamttext vorliegt, da der Gedankengang unabhängig von der Aufgabenstellung zu bewerten ist.

Bei der Bewertung der sprachlichen Mittel (Wortschatz und Strukturen) kann es möglicherweise dann zu einer geringeren Bepunktung kommen, wenn sich das Fehlen von Aufgabenteilen negativ auf den Umfang des gezeigten sprachlichen Materials auswirkt.

Doppelsanktionierung

Bei einer kriterienorientierten Bewertung besteht stets die Gefahr, Merkmale einer Leistung bei unterschiedlichen Kriterien zu bewerten. So besteht die Gefahr, etwa den Umfang der Strukturen auf die Frage nach der Fehlerhaftigkeit zu reduzieren, also nur richtig verwendete

Strukturen als wirklich vorhandene zu werten. Dies ist selbstverständlich nicht zulässig, da Umfang der in einem Text vorhandenen Strukturen und deren Richtigkeit zwei unterschiedlichen Kriterien zugeordnet sind.

Prinzipiell ist wichtig: 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. Deswegen gilt:

- Sprachliche Defizite dürfen weder im Gedankengang noch in den inhaltlichen Subkriterien sanktioniert werden, es sei denn Textpassagen sind komplett unverständlich.
- Logische Defizite innerhalb eines Arguments werden im Subkriterium *begründete Stellungnahme* sanktioniert, der Aufbau der Gesamtargumentation, d.h. die Art der Abfolge der Einzelargumente, im Gedankengang.
- Alle inhaltlichen Mängel im Bereich der Wiedergabe und Auswertung der Materialien werden in den Subkriterien *Textwiedergabe* und *Diagrammauswertung* behandelt und nicht im Subkriterium Gedankengang, zumal dieses Subkriterium das Funktionieren des Textes unabhängig von der Aufgabenstellung und den vorliegenden Materialien bewertet, sofern das Thema vom Aufsatz nicht verfehlt wird.

Kein Leistungsvergleich

Die schriftlichen Leistungen von Prüflingen dürfen bei der Bewertung nicht miteinander 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 können 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.

Bei extremen oder zweifelhaften Einstufungen (Stichwort *Widerspruch gegen das Prüfungsergebnis im Rahmen des so genannten Erweiterten Bewertungsverfahrens*) kann die Testinstitution mit Hilfe des Bewertungsbogens nachvollziehen, wie ein evtl. besonders strenger bzw. milder Bewerter zu seinen Entscheidungen gekommen ist.

Unlesbare Arbeiten

Schätzt die bewertende Person eine Arbeit als unlesbar ein, so muss sie diesbezüglich **vor** Abschluss ihres Bewertungsauftrags die ZfA zwecks Bestätigung kontaktieren.

Wird die Unlesbarkeit bestätigt, wird die Schreibleistung mit insgesamt 0 Punkten bewertet.

Thema verfehlt

„Thema verfehlt“ bedeutet, dass sich die gesamte Arbeit in **allen drei Teilaufgaben** nicht auf das Thema bezieht. Schätzt die bewertende Person eine Schreibleistung als thematisch verfehlt ein, so muss sie sich dies **vor** Abschluss ihres Bewertungsauftrags durch die ZfA bestätigen lassen.

Parallel zum Bestätigungsverfahren ist unbedingt eine vollständige Bewertung der Schreibleistung unter Anwendung aller (Sub-)Kriterien vorzunehmen. Dies ist v. a. wichtig für den Fall, dass die Einschätzung „Thema verfehlt“ nicht bestätigt würde. Sollte die Einschätzung bestätigt werden, wird die Schreibleistung mit insgesamt 0 Punkten bewertet.

DaF-Hintergrund der Bewerberinnen und Bewerber

„Wir verstehen immer, was der Dichter uns sagen will“ – diese *déformation professionnelle* sollte bei der Bewertung von Prüfungsleistungen nach Möglichkeit bedacht werden. Das bedeutet: Bewerberinnen und Bewerber sollten aufgrund ihres DaF-Hintergrundes nicht zu kooperativ rezipieren und deshalb 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 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 II

Prüfling: _____

Bewerter/in: _____

	Begründungen und Belege	Punkte
Gesamt-eindruck	Flüssigkeit	
	Gedankengang	
Inhalt	Textwiedergabe	
	Diagrammauswertung	
	begründete Stellungnahme: ☐ linear ☐ dialektisch (pro <u>und</u> contra)	
Verfügbarkeit sprachlicher Mittel	Wortschatz (1. themenspezifischer Wortschatz, 2. textsortenspez. Redemittel, 3. Wortschatzdefizite)	
	Strukturen (Morpho-Syntax)	
Korrektheit	Korrektheit (Grammatik)	
Gesamtpunktzahl:		

Bewertungsbogen für die Schriftliche Kommunikation – DSD II

Prüfling: _____

Bewerter/in: _____

Begründungen und Belege						Pkt.																		
Gesamteindruck	Flüssigkeit <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%;">verlangsamte Passagen</td> <td style="width: 33%;">wiederholtes Lesen</td> <td style="width: 33%;">unverständliche Passagen</td> </tr> <tr> <td style="height: 20px;"></td> <td></td> <td></td> </tr> </table>					verlangsamte Passagen	wiederholtes Lesen	unverständliche Passagen																
	verlangsamte Passagen	wiederholtes Lesen	unverständliche Passagen																					
	Gedankengang Einleitung: _____ Schlussteil: _____																							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="3" style="text-align: center;">Überleitungen:</th> <th colspan="3" style="text-align: center;">Binnenkohäsion:</th> </tr> <tr> <td style="width: 16.6%;">Text</td> <td style="width: 16.6%;">Grafik</td> <td style="width: 16.6%;">Stellungnahme</td> <td style="width: 16.6%;">Textwied.</td> <td style="width: 16.6%;">Grafik</td> <td style="width: 16.6%;">Stellungnahme</td> </tr> <tr> <td style="height: 40px;"></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>					Überleitungen:			Binnenkohäsion:			Text	Grafik	Stellungnahme	Textwied.	Grafik	Stellungnahme								
Überleitungen:			Binnenkohäsion:																					
Text	Grafik	Stellungnahme	Textwied.	Grafik	Stellungnahme																			
Inhalt	Textwiedergabe <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="6" style="text-align: center;">Textwiedergabe</th> </tr> <tr> <td style="width: 16.6%;">1. Abs.:</td> <td style="width: 16.6%;">2. Abs.:</td> <td style="width: 16.6%;">3. Abs.:</td> <td style="width: 16.6%;">4. Abs.:</td> <td style="width: 16.6%;">5. Abs.:</td> <td style="width: 16.6%;">Gesamtpunktzahl</td> </tr> <tr> <td style="height: 40px;"></td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: center; vertical-align: bottom;"> <input style="width: 30px; height: 20px;" type="checkbox"/> </td> </tr> </table>					Textwiedergabe						1. Abs.:	2. Abs.:	3. Abs.:	4. Abs.:	5. Abs.:	Gesamtpunktzahl						<input style="width: 30px; height: 20px;" type="checkbox"/>	
	Textwiedergabe																							
	1. Abs.:	2. Abs.:	3. Abs.:	4. Abs.:	5. Abs.:	Gesamtpunktzahl																		
						<input style="width: 30px; height: 20px;" type="checkbox"/>																		
Übernahmen aus dem Text: <input style="width: 30px; height: 20px;" type="checkbox"/>																								
Diagrammauswertung <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="3" style="text-align: center;">Diagramm</th> </tr> <tr> <td style="width: 33%;">Daten</td> <td style="width: 33%;">Vergleiche</td> <td style="width: 33%;">Sonstiges</td> </tr> <tr> <td style="height: 40px;"></td> <td></td> <td></td> </tr> </table>					Diagramm			Daten	Vergleiche	Sonstiges														
Diagramm																								
Daten	Vergleiche	Sonstiges																						
begründete Stellungnahme: ☐ linear ☐ dialektisch (pro <u>und</u> contra)																								
Verfügbarkeit sprachlicher Mittel	Wortschatz (1. themenspezifischer Wortschatz, 2. textsortenspezifische Redemittel, 3. Wortschatzdefizite)																							
	1.		2.		3.																			
	<input style="width: 30px; height: 20px;" type="checkbox"/>		<input style="width: 30px; height: 20px;" type="checkbox"/>		<input style="width: 30px; height: 20px;" type="checkbox"/>																			
Strukturen (Morpho-Syntax) <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 50%;">komplexe Hypotaxen</th> <th style="width: 50%;">andere komplexe Strukturen</th> </tr> <tr> <td style="height: 40px;"></td> <td></td> </tr> </table>					komplexe Hypotaxen	andere komplexe Strukturen																		
komplexe Hypotaxen	andere komplexe Strukturen																							
Korrektheit	Korrektheit (Grammatik) <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 33%;">Deklination</th> <th style="width: 33%;">Konjugation</th> <th style="width: 33%;">Satzbau</th> </tr> <tr> <td style="height: 40px;"></td> <td></td> <td></td> </tr> </table>					Deklination	Konjugation	Satzbau																
	Deklination	Konjugation	Satzbau																					
Gesamtpunktzahl:																								

2.2 Ergebnisblatt für die Schriftliche Kommunikation – DSD II

Bitte hier die Bewertungen eintragen						
	<i>Punkte</i>					Punkte pro Kriterium
	3	2	1	0		
GESAMTEINDRUCK						
Flüssigkeit	☐	☐	☐	☐	=	
Gedankengang	☐	☐	☐	☐	=	
INHALT						
Textwiedergabe	☐	☐	☐	☐	=	
Diagrammauswertung	☐	☐	☐	☐	=	
begründete Stellungnahme	☐	☐	☐	☐	=	
SPRACHLICHE MITTEL						
Wortschatz	☐	☐	☐	☐	=	
Strukturen	☐	☐	☐	☐	=	
KORREKTHEIT						
Grammatische Korrektheit	☐	☐	☐	☐	=	
Gesamtpunktzahl					=	

2.3 Bewertungskriterien für die Schriftliche Kommunikation – DSD II

Kriterium		C1 3 Punkte	C1/B2 2 Punkte	B2 1 Punkt	unter B2 0 Punkte
Gesamt- eindruck	Flüssigkeit	Der Text ist flüssig zu lesen.	Der Lesefluss ist an einigen Stellen verlangsamt.	Der Lesefluss stockt an einigen Stellen.	Der Lesefluss stockt an mehreren Stellen.
	Gedankengang	Der Text ist klar strukturiert.	Der Text ist insgesamt strukturiert.	Der Text ist nicht durchgehend strukturiert, aber als zusammenhängendes Ganzes verfasst.	Der Text zeigt nur vereinzelt Strukturelemente.
Inhalt	Wenn das Thema völlig verfehlt ist, wird der gesamte Prüfungsteil „Schriftliche Kommunikation“ mit 0 Punkten bewertet.				
	Textwiedergabe	Wichtige Textaussagen werden eigenständig und vollständig wiedergegeben.	Wichtige Textaussagen werden nahezu eigenständig und nahezu vollständig wiedergegeben.	Wichtige Textaussagen werden überwiegend eigenständig und überwiegend vollständig wiedergegeben.	Wichtige Textaussagen werden nur zum Teil und mehrfach nicht eigenständig wiedergegeben.
	Diagramm- auswertung	Die Grafik wird anhand wichtiger Daten angemessen ausgewertet.	Die Grafik wird anhand wichtiger Daten kurz ausgewertet.	Die Grafik wird anhand einzelner Daten kaum ausgewertet oder wichtige Daten werden weitgehend angemessen wiedergegeben, aber nicht ausgewertet.	Wichtige Daten werden nur zum Teil dargestellt. Eine Auswertung findet nicht statt.
	begründete Stellungnahme	Die eigene Meinung ist deutlich. Die Stellungnahme enthält eine differenziert ausgearbeitete Argumentation.	Die eigene Meinung ist deutlich. Die Stellungnahme enthält eine ausgearbeitete Argumentation. Die einzelnen Argumente gehen zumeist über These und Begründung hinaus.	Die eigene Meinung ist deutlich. Die Stellungnahme enthält eine knapp ausgearbeitete Argumentation. Die einzelnen Argumente bestehen zumeist nur aus These und Begründung.	Die eigene Meinung ist deutlich. Die Stellungnahme zerfällt in Einzelargumente, die kaum ausgearbeitet werden und/oder in sich widersprüchlich sind.
Verfügbarkeit sprachlicher Mittel	Wortschatz	Der Wortschatz ist differenziert. Textsortenspezifische Redemittel werden durchgehend verwendet. Sachverhalte werden präzise versprachlicht.	Der Wortschatz ist in Teilen differenziert. Textsortenspezifische Redemittel werden häufig verwendet. Sachverhalte werden weitgehend präzise versprachlicht. Falsche Wortwahl / Wortfehler sind selten.	Der Wortschatz geht über den Grundwortschatz hinaus, weist jedoch einige Lücken auf. Textsortenspezifische Redemittel werden begrenzt verwendet. Sachverhalte werden nicht immer präzise versprachlicht. Falsche Wortwahl / Wortfehler treten häufiger auf.	Der Wortschatz verbleibt auf der Ebene des Grundwortschatzes und weist deutliche Lücken auf. Textsortenspezifische Redemittel werden kaum verwendet. Eine unpräzise Lexik und falsche Wortwahl / Wortfehler prägen den Text an mehreren Stellen.
	Strukturen (Morpho-Syntax)	Ein hohes Maß an komplexen Strukturen wird durchgehend beibehalten.	Komplexe Strukturen dominieren, allerdings werden auch einfache Strukturen in einem relevanten Umfang verwendet.	Komplexe Strukturen werden in einem relevanten Umfang verwendet, allerdings dominieren einfache Strukturen.	Komplexe Strukturen kommen selten vor. Überwiegend werden einfache Strukturen verwendet.
Korrekt- heit	Korrektheit (Grammatik)	Der Text zeigt ein hohes Maß an grammatischer Korrektheit.	Der Text zeigt - von wenigen Textpassagen abgesehen - ein hohes Maß an grammatischer Korrektheit.	Der Text zeigt – abgesehen von gelegentlichen und nichtsystematischen Fehlern – ein zureichendes Maß an grammatischer Korrektheit.	Der Text zeigt deutliche Defizite in der Grammatik. Systematische Fehler kommen vor und/oder der Text ist fast durchgehend fehlerhaft.

Anhang: Literaturhinweise

Literaturhinweise rund ums Testen und Bewerten

Referenzrahmen

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

Europarat (2020): Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen. Begleitband. Stuttgart: Klett.

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 Expertenseminar. 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