

Manuela Glaboniat & Guenther Sigott

# Das Konzept der kriterienorientierten Bewertung Dimensionen und Niveaus

ACHTUNG: Dieser Text bezieht sich auf eine frühere Beurteilungsvariante, die aus juristischen Gründen nicht übernommen werden konnte. Für die SRDP Deutsch gelten ab 2013 (Schulversuch) die Erläuterungen vom 11. Juni 2012, die auf der Website des BIFIE unter <https://www.bifie.at/node/1490> publiziert sind.

Je nach Ziel und Funktion einer Prüfung oder eines Tests<sup>1</sup> wird in der Testtheorie zwischen verschiedenen *Testarten* unterschieden. Im Bereich der schulischen Leistungsmessung und -evaluation werden v. a. Verfahren zur Messung des Lernfortschritts bzw. der schulischen Leistung (*Lernfortschritts- und Leistungstests*) eingesetzt. Sie verfolgen v. a. das Ziel, Lehrenden und Lernenden aufzuzeigen, ob und in welchem Ausmaß Fortschritte erzielt und Lernziele einer bestimmten Themeneinheit oder eines Schuljahres erreicht werden konnten. Daher beziehen sich diese Testverfahren und die daraus resultierenden Beurteilungen (Noten) in erster Linie auf die im Unterricht jeweils bearbeiteten Inhalte. Hinsichtlich der Beurteilung kommt hinzu, dass im schulischen Kontext meist *norm-* oder *individualorientiert* beurteilt wird.

*Normorientiert* bedeutet, dass – selbst bei Verwendung von (individuellen, schulinternen) Beurteilungskriterien – die Leistung einer Schülerin/eines Schülers (z. B. ein Aufsatz) *relativ* zur Gruppe (Mitschüler/innen in der Klasse) bzw. zum Klassendurchschnitt bewertet wird. Die jeweilige Note hängt also davon ab, wie gut die Leistungen der anderen sind.

Eine *individualorientierte* Beurteilung berücksichtigt (zusätzlich noch) die individuellen Faktoren einer/eines Lernenden (z. B. persönliche Schwächen und Stärken, der individuelle Leistungs- und Lernzuwachs u. Ä.).

Beide Bezugsnormen sind aus pädagogischen und lernpsychologischen Gründen legitim und haben ihren festen Platz in schulischer Leistungsmessung. Allerdings sind die daraus resultierenden Noten (wegen ihrer Abhängigkeit von den oben genannten Faktoren) immer relativ und sagen daher nur wenig über die konkreten Fähigkeiten einer Person aus – was im Rahmen der alltäglichen schulischen Leistungsmessung auch nicht oberstes Ziel ist.

Anders ist es bei der Reife- und Diplomprüfung. Diese Prüfung soll nicht nur möglichst valide und zuverlässig über die Fähigkeiten einer Person Auskunft geben können (Berichtsfunktion), sondern darüber hinaus auch noch eine Prognose über die zukünftige „Bewährung“ der Prüfungskandidatinnen und -kandidaten erlauben (z. B.: Ist jemand „reif“ für das Studium/eine bestimmte Arbeit etc.?). Die Reife- und Diplomprüfung ist somit auch ein Selektionsinstrument (Berechtigungs- und Selektionsfunktion), d. h. wichtige Entscheidungen (wie z. B. spezifische Studienplätze bzw. der allgemeine Zugang zu Universitäten

---

<sup>1</sup> Die Begriffe *Test* und *Prüfung* werden hier synonym verwendet.

oder anderen Bildungseinrichtungen, Stipendien, Auswahlverfahren am Arbeitsmarkt u. Ä.) hängen vom Bestehen bzw. von den Noten bei diesen Prüfungen ab.

Es zeigt sich, dass die Reife- und Diplomprüfung ganz andere Ziele als die oben genannten (nicht standardisierten) Test- und Beurteilungsverfahren verfolgt und daher anderen – und als ein sogenannter *high-stakes test* auch höheren – Qualitätsstandards verpflichtet ist.

Aus testtheoretischer Sicht entspricht die Reife- und Diplomprüfung ihren Anforderungen zufolge den Prinzipien einer standardisierten *Feststellungsprüfung* (auch: *Qualifikationstest*, *proficiency test*). Ein Hauptmerkmal von Feststellungsprüfungen ist, dass sie den Ausprägungsgrad von vorab definierten Fähigkeiten und Anforderungen in Bezug auf abschätzbare bzw. mögliche zukünftige Verwendungssituationen erfassen, d. h. sie beziehen sich auf vorher festgelegte Kompetenzmodelle bzw. Kompetenzbeschreibungen.

Das Kompetenzmodell bzw. die Kompetenzbeschreibungen bilden in weiterer Folge die Grundlage für die konkreten Prüfungsaufgaben sowie die Beurteilungskriterien, die für die Bewältigung der jeweiligen Aufgabe relevant sind (z. B. Erfüllung der Aufgabe, Textaufbau, formale Korrektheit u. Ä.).

Im Unterschied zu den im schulischen Alltag meist stärker norm- und individualorientierten Beurteilungsverfahren setzen Feststellungsprüfungen in jedem Fall eine sach- und kriterienorientierte Beurteilung voraus (Sachnorm). Das bedeutet, die Leistungen einer Schülerin/eines Schülers müssen hauptsächlich in Relation zu den erwarteten Kompetenzen und Lernzielen gemessen (festgestellt) werden. Alle anderen Faktoren (wie die Leistungen der anderen Lernenden, individuelle Vorlieben und Erwartungen von Lehrenden) sollten dabei eine möglichst geringe Rolle spielen. Zusätzlich erfordern standardisierte Feststellungsprüfungen mit hoher Berichts- und Selektionsfunktion (*high-stakes tests*) wie erwähnt die Einhaltung höchster Qualitätsstandards, die sich vor allem in den klassischen Gütekriterien der Reliabilität und Validität zeigen.

Eine besondere Herausforderung bei der schriftlichen Reife- und Diplomprüfung liegt dabei in der zuverlässigen Beurteilung von Aufsätzen, also in der Auswertungsreliabilität bzw. Beurteilungsübereinstimmung. Mehr oder weniger unproblematisch ist in dieser Hinsicht das Testen und Beurteilen von Sachwissen (z. B. Geografie, Geschichte, Mathematik) oder rezeptiver Sprachkompetenzen (Hörverstehen, Leseverstehen). In diesen Bereichen können geschlossene oder halboffene Testformate (Multiple-Choice-Aufgaben, Lückentexte usw.) eingesetzt werden, die nicht nur einfach und schnell auswertbar sind, sondern a priori auch eine sehr hohe Übereinstimmung bei der Beurteilung ergeben, da meist lediglich die richtigen Antworten gezählt werden müssen. Dies ist bei der Beurteilung von Aufsätzen selbstverständlich nicht möglich, denn hier entsteht das Urteil aus dem Zusammenspiel bestimmter Eigenschaften der Schülertexte, die bei der/dem Beurteilenden (Rater/in) einen Eindruck hinterlassen. Der Ansatz zur Kompetenzmessung bei der schriftlichen Reife- und Diplomprüfung ist somit nicht *itemorientiert*, sondern *performanzorientiert*. Beim performanzorientierten Testen kommt der Rolle der Beurteilenden (Rater/innen) große Bedeutung zu, da ihre Einschätzung der Schülerperformanzen bzw. ihre Interpretation der Beurteilungskriterien und -skalen wesentlich die Reliabilität und Validität des Tests bestimmen. Und genau darin liegt das große Problem: Verschiedene Beurteiler/innen kommen nicht selten zu höchst unterschiedlichen Beurteilungen bzw. Noten.

Um diesem Problem Rechnung zu tragen, sind bei der Aufsatzbeurteilung einige Schritte unerlässlich. Zunächst muss ganz grundsätzlich festgelegt werden, worauf bei der Beurteilung überhaupt geachtet werden soll. Dies erfordert im Vorfeld oft nochmals einen kritischen Blick auf Testaufgaben, Kompetenzmodell und Testkonstrukt mit der Frage, ob und inwieweit auch wirklich (nur) das überprüft wird, was überprüft werden soll. In einem weiteren Schritt muss geklärt werden, welche Aspekte des sprachlichen Handelns für die Erfüllung der Testaufgaben schwerpunktmäßig eine Rolle spielen und bei der Beurteilung erfasst werden sollen. In der schulischen Beurteilungspraxis finden sich hier zum Teil große Unterschiede.

Viele Lehrende konzentrieren sich in diesem Zusammenhang gerne auf das Zählen von formalsprachlichen Fehlern – oft im guten Glauben, damit den Anforderungen der Reliabilität und Validität Genüge zu leisten. Dabei ist ihnen oft nicht bewusst, dass auf diese Art und Weise nur ein kleiner Teil (Orthografie) der schriftlichen Kompetenzen beurteilt wird und die daraus resultierenden Noten nur begrenzt aussagekräftig sind. Andere Lehrende achten bei der Beurteilung der Aufsätze vielleicht stärker auf den Inhalt oder Textaufbau, wieder andere definieren für sich Kriterien wie *gedankliche Komplexität*, *Kreativität* usw.

Wie auch immer, um eine möglichst hohe Beurteilungsreliabilität zu erzielen, ist es notwendig, sich auf der Basis des Testkonstrukts (Kompetenzmodells) auf *gemeinsame Beurteilungskriterien* (Dimensionen) zu einigen und diese möglichst genau zu definieren. In einem nächsten Schritt gilt es, für jede einzelne dieser Dimensionen verschiedene Qualitäts- oder Niveaustufen der erwarteten Leistung festzulegen und so eine analytische Skala zu entwickeln (siehe Beurteilungsraster unten). Ein letzter Schritt liegt schließlich in der Festlegung, in welchem Verhältnis die einzelnen Dimensionen zueinander stehen und wie aus der Beurteilung einer Leistung nach verschiedenen Dimensionen eine Gesamtnote ermittelt wird.

Mit der Ausarbeitung eines solchen Beurteilungssystems ist eine hohe Auswertungsreliabilität jedoch noch nicht gewährleistet, denn hohe Übereinstimmung lässt sich erst durch intensives Beurteilungstraining erzielen. Dabei werden die Beurteilenden (Rater/innen) vor allem anhand von ausgewählten Schülertexten (*Benchmarks*) geschult.

Bevor im Folgenden die analytische Skala der schriftlichen Reife- und Diplomprüfung Deutsch (SRPD) vorgestellt wird, sei noch darauf hingewiesen, dass zur Gewährleistung von Reliabilität und Validität noch eine Reihe weiterer Maßnahmen gehören: Viele davon befinden sich auf der Ebene der Testentwicklung und Aufgabenerstellung. Es ist zum Beispiel ein Grundprinzip, dass die Testaufgaben und Anweisungen klar und präzise formuliert sind und alle wichtigen Informationen – auch in Bezug auf die Beurteilung – enthalten (z. B. zur erwünschten Textlänge, Textsorte, Abhandlung der Leitpunkte etc.). In diesem Zusammenhang spielt gerade bei der Zielgruppe der Jugendlichen auch die *face validity* („Augenscheinvalidität“) eine große Rolle. Sie bezeichnet den „auf den ersten Blick“ gewonnenen Globaleindruck einer Prüfung. Eine hohe *face validity* ist dann gegeben, wenn die Testaufgabe als angemessen, sinnvoll, ansprechend, realitätsnah, transparent und nachvollziehbar eingeschätzt wird und somit nicht nur inhaltlich, sondern auch emotional hohe Akzeptanz findet. Testaufgaben wirken oft allein aufgrund ihrer äußerlich attraktiven und authentischen Aufbereitung (Layout, grafische und akustische Gestaltung) motivierender.

## Das Beurteilungssystem in der SRDP

Wie oben skizziert, werden in der SRDP Deutsch die zum Bestehen der Prüfung relevanten Kompetenzen durch Beurteilen von Schreibperformanzen gemessen. Dies bedeutet, dass es hier nicht um das *Zählen* von gelösten Aufgaben geht, sondern um die *Beurteilung* von jeweils zwei längeren Texten anhand klar definierter Dimensionen einer analytischen Skala. Diese Beurteilungen kommen nicht durch Zählen von diskreten, d. h. klar abgegrenzten Aufgaben, die entweder gelöst wurden oder nicht, zustande.

Der performanzorientierte Ansatz legt eine kriterienorientierte Interpretation der Prüfungsergebnisse nahe. Dabei wird die Leistung einer Kandidatin/eines Kandidaten nicht unter Bezugnahme auf die Leistungen anderer Kandidatinnen/Kandidaten oder unter Bezugnahme auf den Mittelwert aller Kandidatinnen/Kandidaten interpretiert. Vielmehr wird die Leistung unter Bezugnahme auf Beschreibungen von Kompetenzen interpretiert, die für bestimmte Niveaus bzw. Noten typisch sind. Der Vorteil der kriterienorientierten Interpretation von Prüfungsergebnissen besteht darin, dass sich die Anforderungen nicht von Schule zu Schule und auch nicht im Laufe der Zeit, also bei verschiedenen Maturaterminen, ändern. Außerdem erfordert es

dieser Ansatz, die Bedeutung von Prüfungsergebnissen inhaltlich zu definieren, und er stellt damit das Erfordernis der Validität in den Vordergrund.

Da das Resultat der Prüfung in hohem Maß vom Urteil der Rater/innen abhängt, ist es notwendig, den Prozess der Urteilsbildung zu standardisieren. Dies erfolgt durch die Verwendung einer analytischen Skala, die einerseits Dimensionen und andererseits Niveaustufen auf jeder dieser Dimensionen vorgibt. Die Urteilsbildung der einzelnen Raterin/des einzelnen Raters erfolgt innerhalb dieses konzeptuellen Rahmens.

Die Dimensionen und die Niveaubeschreibungen wurden von der Arbeitsgruppe SRDP Deutsch in ausführlicher Diskussion und unter Bezugnahme auf authentische Schülertexte entwickelt. Aus diesem Prozess ging eine analytische Skala mit den vier Dimensionen *Aufgabenerfüllung aus inhaltlicher Sicht*, *Aufgabenerfüllung aus textstruktureller Sicht*, *Aufgabenerfüllung in Bezug auf Stil und Ausdruck* und *Aufgabenerfüllung hinsichtlich formaler Richtigkeit* hervor. Auf jeder dieser Dimensionen werden vier Niveaustufen beschrieben, die den Schulnoten *Genügend*, *Befriedigend*, *Gut* und *Sehr gut* entsprechen und diesen somit konkrete Bedeutung aus kriterienorientierter Sicht verleihen. Die Niveaustufen stehen zueinander in implikatorischer Beziehung, das heißt, dass jede Niveaustufe die Kompetenzen der niedrigeren Niveaustufe(n) einschließt.

## Standardisierte kompetenzorientierte schriftliche Reifeprüfung / Reife- und Diplomprüfung Deutsch – Bewertungsraster

Niveaustufe	1	2	3	4
Note	Genügend	Befriedigend	Gut	Sehr gut
Aufgabenerfüllung aus <u>inhaltlicher</u> Sicht	Schreibhandlung(en) im Sinne der Textsorte überwiegend erkennbar	Schreibhandlung(en) im Sinne der Textsorte weitgehend realisiert	Schreibhandlung(en) im Sinne der Textsorte durchgehend realisiert	Schreibhandlung(en) im Sinne der Textsorte durchgehend realisiert
	Alle Arbeitsaufträge angesprochen und mindestens zwei bearbeitet	Alle Arbeitsaufträge angesprochen und mindestens zwei erfüllt	Alle Arbeitsaufträge erfüllt	Alle Arbeitsaufträge erfüllt
	Wichtige Einzelaussagen/-aspekte des Inputtexts erfasst	Kernaussage des Inputtexts erfasst	Inputtext vollständig erfasst	Inputtext vollständig erfasst
	In elementaren Punkten überwiegend sachlich richtig	In elementaren Punkten weitgehend sachlich richtig	In zentralen Passagen durchgehend sachlich richtig	Sachlich richtig
		Ansätze zur Eigenständigkeit	Über den Inputtext hinaus eigenständig	Nachvollziehbare Entwicklung eines eigenen Standpunktes
			Komplexität und Ideenreichtum	
Aufgabenerfüllung aus <u>textstruktureller</u> Sicht	Gedankliche Grobstruktur des Textes erkennbar	Text gedanklich und grafisch überwiegend der Textsorte angemessen strukturiert	Text gedanklich und grafisch der Textsorte angemessen klar strukturiert	Text gedanklich und grafisch der Textsorte angemessen klar strukturiert
	Erkennbare Bezugnahme auf den Inputtext	Eindeutige Bezugnahme auf den Inputtext	Weitgehend gelungene Verknüpfung mit dem Inputtext	Gelungene Verknüpfung mit dem Inputtext
	Überwiegend kohärenter Aufbau innerhalb der Absätze	Gut erkennbare Kohärenz innerhalb der Absätze, nachvollziehbarer Einsatz von Kohäsionsmitteln	Leicht nachvollziehbare Binnengliederung, zielgerichteter, sicherer Einsatz von Kohäsionsmitteln; kohärent und frei von Gedankensprüngen	Leicht nachvollziehbare Binnengliederung, zielgerichteter, sicherer Einsatz von passenden Textorganismen; kohärent und frei von Gedankensprüngen, zielgerechter Einsatz von metakommunikativen Mitteln
Aufgabenerfüllung in <u>Bezug auf Stil und Ausdruck</u>	Ansätze zur schreibhandlungs- und situationsadäquaten Sprachverwendung	Überwiegend schreibhandlungs- und situationsadäquate Sprachverwendung	Weitgehend schreibhandlungs- und situationsadäquate Sprachverwendung	Durchwegs schreibhandlungs- und situationsadäquate Sprachverwendung mit gegebenenfalls entsprechenden Stilmitteln
	In den Schlüsselbegriffen treffend, überwiegend angemessene Ausdrucksweise, geringe Varianz in Wortwahl	Weitgehend präzise Wortwahl und angemessene Ausdrucksweise, erkennbare Varianz in Wortwahl	Präzise und variantenreiche Wortwahl, idiomatisch, dem Inhalt und der Textsorte entsprechend	Durchgehend differenzierte und variantenreiche Wortwahl, dem Inhalt und der Textsorte entsprechend; Verwendung einer angemessenen Fachsprache; idiomatisch gewandt; feinere Bedeutungsnuancen auch bei komplexeren Sachverhalten deutlich
	In Ansätzen erkennbare Varianz in Satzstruktur	Erkennbare Varianz in Satzstruktur	Weitgehend variantenreiche und komplexe bzw. der Textsorte angemessene Satzstrukturen	Variantenreiche und komplexe bzw. der Textsorte angemessene Satzstrukturen
	An den Inputtext angelehnte Formulierungen, vieles wortwörtlich übernommen	Ansätze zu eigenständigen Formulierungen in Bezug auf den Inputtext	Weitgehend eigenständige Formulierungen in Bezug auf den Inputtext	Eigenständige Formulierungen in Bezug auf den Inputtext
Aufgabenerfüllung hinsichtlich <u>formaler Richtigkeit</u>	Überwiegend richtige Anwendung der Prinzipien der deutschen Schreibung und Zeichensetzung	Weitgehend richtige Anwendung der Prinzipien der deutschen Schreibung und Zeichensetzung	In Orthografie und Zeichensetzung weitgehend fehlerfrei	In Orthografie und Zeichensetzung nahezu fehlerfrei
	Überwiegend korrekte Anwendung der grundlegenden Grammatikregeln	Weitgehend korrekte Anwendung der Grammatikregeln	Grammatikalisch weitgehend fehlerfrei	Grammatikalisch nahezu fehlerfrei

Im Folgenden werden die vier Dimensionen im Detail beschrieben und diskutiert.

### Aufgabenerfüllung aus inhaltlicher Sicht

Diese Dimension beschreibt zunächst, inwieweit die Intention oder der Zweck des Outputtextes erkennbar ist. Dafür wird die Bezeichnung *Schreibhandlung* verwendet. In der Regel erfordert die Aufgabenstellung die Produktion von Texten, in denen zwei oder mehrere Schreibhandlungen (z. B. *Zusammenfassen* und *Argumentieren*) realisiert werden müssen. Das Spektrum reicht von der bloßen Erkennbarkeit der Schreibhandlungen bis zu deren makelloser Realisierung, die sie für die Leserin/den Leser mühelos identifizierbar macht. Dies wird bereits ab Niveau 2 gefordert. Performanzen auf Niveau 2 unterscheiden sich in dieser Hinsicht nicht von jenen auf Niveau 1. Weiters beinhaltet diese Dimension die Erfüllung der drei oder vier in der Aufgabenstellung explizit aufgelisteten Arbeitsaufträge. Zur Erreichung von Niveau 4 müssen alle Arbeitsaufträge angesprochen und zwei davon eingehend behandelt worden sein. Bei der Beurteilung ist zwischen angesprochenen und bearbeiteten Arbeitsaufträgen zu unterscheiden. Ein Arbeitsauftrag gilt als bearbeitet, wenn eine deutliche vertiefte Auseinandersetzung (z. B. ausführliche Begründung, detaillierte Beschreibung) erfolgt ist. Ab Niveau 2 müssen alle Arbeitsaufträge bearbeitet worden sein. Ein weiterer Aspekt dieser Dimension ist das Verstehen des Inputtextes. Zur Erreichung von Niveau 4 muss aus dem Outputtext hervorgehen, dass die wichtigen Einzelaussagen des Inputtextes verstanden wurden. Für Niveau 3 ist es überdies erforderlich, die Kernaussage des Inputtextes zu verstehen, während ab Niveau 2 vollständiges Verstehen einschließlich relevanter Textdetails nachzuweisen ist. Außerdem beinhaltet diese Dimension den Aspekt der sachlichen Richtigkeit. Damit wird allgemeines Weltwissen angesprochen. Aber bereits zur Erreichung des Niveaus 4 müssen die für den Outputtext wesentlichen Punkte frei von objektiv überprüfbar falschen Behauptungen sein. Dasselbe gilt auch für Niveau 3. Niveau 2 erfordert überdies, dass die zentralen Passagen, somit ausgedehnte Textabschnitte, frei von solchen Behauptungen sind. Texte auf Niveau 1 müssen durchgehend sachlich richtig sein. Ein weiterer Aspekt dieser Dimension ist die gedankliche Loslösung vom Inputtext, ohne jedoch diesen zu ignorieren. Dies ist in Ansätzen für Niveau 3 erforderlich, während für Niveau 1 die Entwicklung eines klar nachvollziehbaren eigenen Standpunktes typisch ist, sofern die Aufgabenstellung dies zulässt bzw. erfordert. Performanzen des Niveaus 1 zeichnen sich überdies durch Komplexität und Vielfalt der dem Outputtext zugrunde liegenden Ideen aus.

### Aufgabenerfüllung aus textstruktureller Sicht

Auf dieser Dimension geht es um Kohärenz und Kohäsion sowohl innerhalb des Outputtextes als auch zwischen dem Inputtext und dem Outputtext. Wichtige Aspekte sind die Anordnung der Ideen im Sinne von Absätzen und die Anordnung der Absätze zueinander sowie die grafische Signalisierung von Absätzen durch Einrücken oder Freizeilen im Outputtext. Das Spektrum reicht von der bloßen Erkennbarkeit der inhaltlichen Grobstruktur für Niveau 4 bis zu einem inhaltlich und grafisch klar strukturierten Text, in dem die Anordnung der Ideen mit der grafischen Markierung im Einklang steht, für die Niveaus 2 und 1. Darüber hinaus beschreibt diese Dimension den Grad der Markierung der (gedanklichen) Kohärenzbeziehungen durch kohäsive (sprachliche) Mittel und damit die Leichtigkeit, mit der die Kohärenzbeziehungen durch die Leserin/den Leser nachvollzogen werden können. Dies reicht von durch Kohäsionsmittel gut erkennbar gemachter Kohärenz innerhalb der Absätze auf Niveau 3 über die durch Kohäsionsmittel explizit markierte und dadurch leicht nachvollziehbare Kohärenz zwischen den Absätzen, die Binnengliederung, auf Niveau 2, bis zu Niveau 1, wo die Kohärenz zwischen den Absätzen zusätzlich durch den zielgerech-

ten Einsatz von metakognitiven Mitteln verdeutlicht wird.<sup>2</sup> Weiters umfasst diese Dimension die Beziehung zwischen Inputtext und Outputtext, wobei das Spektrum von einer erkennbaren Bezugnahme auf Niveau 4 über eine eindeutige Bezugnahme auf Niveau 3 bis zu einer gelungenen Bezugnahme auf Niveau 1 reicht.

### Aufgabenerfüllung in Bezug auf Stil und Ausdruck

Diese Dimension beschreibt den Grad der Angemessenheit der sprachlichen Mittel, insbesondere der Wortwahl und des Satzbaus, für die jeweilige Schreibhandlung und die kommunikative Situation. Dabei werden für Niveau 4 Ansätze zur Verwendung für die jeweilige Schreibhandlung und Situation typischer sprachlicher Mittel gefordert, während für Niveau 1 zusätzlich zur durchgehenden registeradäquaten Sprachverwendung, sofern die Textsorte dies zulässt, auch die Verwendung entsprechender Stilmittel gefordert wird. In puncto Wortwahl zeichnen sich Performanzen des Niveaus 4 durch oftmals unpräzise Wortwahl wegen der Verwendung von verallgemeinernden Wörtern und durch eine schmale Bandbreite des Wortschatzes aus. Auf Niveau 3 ist die Wortwahl präzise, wenngleich die Variationsbreite des Wortschatzes begrenzt bleibt. Niveau 2 zeigt treffende und variantenreiche Wortwahl, wobei auch idiomatische – d. h. für das Deutsche natürliche und typische – Wendungen auftreten. Niveau 1 ist gekennzeichnet durch große Präzision in der Wortwahl, die bis zur korrekten Verwendung von Fachvokabular reicht, sowie durch eine große Varianz des Wortschatzes, wodurch auch bei komplexen Sachverhalten feine Bedeutungsnuancen ausgedrückt werden. Darüber hinaus besteht ein bemerkenswertes Maß an idiomatischer Gewandtheit. In Bezug auf die geforderte Textsorte ist das Repertoire an Satzstrukturen auf Niveau 4 gering, wesentlich breiter auf Niveau 3 und paart sich auf den Niveaus 2 und 1 mit bemerkenswerter Komplexität im Satzbau. Schließlich kennzeichnet diese Dimension auch den Grad der Eigenständigkeit der Formulierung in Bezug auf den Inputtext. Das Spektrum reicht hier von einer Tendenz zur wörtlichen Übernahme von Formulierungen aus dem Inputtext auf Niveau 4 bis zur eigenständigen Formulierung, ohne dass dadurch der Konnex zum Inputtext leidet, auf den Niveaus 2 und 1.

### Aufgabenerfüllung hinsichtlich formaler Richtigkeit

Obwohl formale Richtigkeit durch die Präsenz oder Absenz von Verstößen gegen Konventionen und Regeln auf Wort- und Satzebene definiert ist, zielt diese Dimension nicht auf das Zählen solcher Phänomene ab. Vielmehr ist der Eindruck zu beschreiben, der durch das Auftreten von Regelverstößen auf lexikalischer oder syntaktischer Ebene entsteht. So sind Performanzen des Niveaus 4 und 3 frei von Verstößen gegen Orthografie und Zeichensetzung, die kommunikationsstörend wirken. Der Grad der Fehlerfreiheit steigert sich bis zu Niveau 1, wo Verstöße gegen Orthografie und Zeichensetzung sehr selten sind und eher den Charakter von Flüchtigkeitsfehlern haben. In puncto Grammatik (Wortgrammatik, Satzgrammatik) weisen Performanzen des Niveaus 4 Verstöße gegen mehrere Grammatikregeln (z. B. Wortbildung, Tempusverwendung, Satzbau, Modus in indirekter Rede) auf. Wiederholte Verstöße gegen mehr als eine Grammatikregel kommen jedoch nicht vor. Performanzen der Niveaus 3 und 2 sind grammatikalisch überwiegend fehlerfrei, während auf Niveau 1 Grammatikfehler kaum vorkommen. Es ist bei alledem zu betonen, dass nicht die numerische Häufigkeit bestimmter Fehler(typen) entscheidend ist, sondern der Gesamteindruck.

---

<sup>2</sup> Metakognition: Wissen über das eigene Wissen. Gemeint sind hier Formulierungen, die den Textzusammenhang erläutern. Um diese verfassen zu können, bedarf es eben eines Wissens um die Schwierigkeiten der Textrezeption und die Fähigkeit, einen Text bewusst zu strukturieren.

## Die Gesamtnote

Da die Prüfung aus zwei getrennten Aufgaben besteht, werden diese getrennt voneinander beurteilt. Jede der beiden Aufgaben wird zunächst auf allen vier Dimensionen beurteilt. Aus den Beurteilungen auf den vier Dimensionen wird der ungewichtete Mittelwert gebildet. Dabei kommt die sogenannte Vetoregel zum Tragen: Ist die Beurteilung auf der Dimension *Aufgabenerfüllung aus inhaltlicher Sicht* negativ, dann ist die Beurteilung für die Aufgabe negativ. Die Beurteilung für die Aufgabe ist auch negativ, wenn die Beurteilung auf mindestens zwei Dimensionen negativ ist. Schließlich ist die Beurteilung einer Aufgabe auch negativ, wenn die Aufgabe nicht bearbeitet wurde. Aus den Teilbeurteilungen der beiden Aufgaben wird sodann die Gesamtnote wiederum durch Bildung des ungewichteten Mittelwertes ermittelt, wobei Dezimalwerte von 0,5 und darüber auf die nächstgrößere ganze Zahl aufgerundet werden.

Durch die Anwendung dieser Vorgangsweise wird den Erfordernissen der Reliabilität und Validität Rechnung getragen und die Beurteilung für alle Beteiligten nachvollziehbar gemacht.



## Literatur

Bachmann, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Glaboniat, M. (2006). *Das Papier nicht wert ... Zum Problem schulischer Leistungsmessung und Benotung und neue Chancen durch Qualitäts- und Leistungsstandards*. In *ide – informationen zur deutschdidaktik* 4/2006. S. 32–51. Verfügbar unter [http://www.uni-klu.ac.at/deutschdidaktik/downloads/Glaboniat\\_2\\_06.pdf](http://www.uni-klu.ac.at/deutschdidaktik/downloads/Glaboniat_2_06.pdf) [22.02.2012].

Ingenkamp, K. (Hrsg.) (1995). *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte*. 9. Auflage. Weinheim: Beltz.

Vierlinger, R. (1999). *Leistung spricht für sich selbst. „Direkte Leistungsvorlage“ (Portfolios) statt Ziffernzensuren und Notenfetischismus*. Heinsberg: Dieck.