



STAATSIKITUT FÜR SCHULQUALITÄT
UND BILDUNGSFORSCHUNG
MÜNCHEN

Qualitätsagentur
Schellingstraße 155 • 80797 München • Tel.: 089 2170-2273 • Fax: -2816
E-Mail: qualitaetsagentur@isb.bayern.de

Ergänzende Hinweise zu den klassenbezogenen Rückmeldungen von Testergebnissen

**aus den Orientierungsarbeiten 2008
Rechtschreiben Jahrgangsstufe 2**

Die nachfolgenden Ausführungen geben – bei Interesse – Einblick in die wissenschaftlich begründeten Arbeitsgänge, die eine hohe Spezialkompetenz im empirischen Arbeiten und in Statistik erfordern, die in der Regel nicht zur Ausbildung von Lehrkräften gehören. Auch wenn wir uns dessen bewusst sind, dass demzufolge die Ausführungen zum größten Teil schwierig zu verstehen sein mögen, wollen wir sie doch als Zusatzangebot zur Verfügung stellen, um die Grundlagen der Rückmeldungen so weit wie möglich transparent zu machen. Es ist ein Zusatzangebot an Lehrkräfte, die an wissenschaftlichen Grundlagen für Leistungsvergleiche interessiert sind.

Die Ihnen zugesandten Rückmeldungen zu den Orientierungsarbeiten 2008 Rechtschreiben Jahrgangsstufe 2 enthalten Informationen, die einerseits für sich selbst sprechen, die andererseits aber auch Konstrukte enthalten, die vielleicht mancher Lehrerin und manchem Lehrer nicht geläufig sind: z. B. Fähigkeitsniveau, Lösungsquote/Lösungswahrscheinlichkeit. Gerade in diesen Konstrukten liegt eine wesentliche Leistung von Orientierungsarbeiten, nämlich die erbrachten Schülerleistungen im Hinblick auf das erzielte Leistungs- (= Fähigkeits-) niveau sowohl in der Zeitreihe, also über Schuljahre hinweg, als auch im Vergleich zu anderen Klassen oder Schulen zu interpretieren und auszuwerten. Dabei sind einige Bedingungen zu beachten:

- Die Aufgaben müssen, z. B. nach ihrer Schwierigkeit, auf einer Skala eindeutig verortet werden, damit die mit ihnen gegebenen Anforderungen in verschiedenen Tests miteinander in Beziehung gesetzt werden können
- Die von den Schülern zu erwartenden Fähigkeiten müssen zur leichteren Interpretation durch die Lehrer letztendlich nach Niveaustufen definiert und gruppiert werden
- Auffälligkeiten im Leistungsprofil einer Klasse sind in der Regel nicht schon bei einer Abweichung vom „Durchschnitt“ gegeben; hier ist das gesamte Leistungsspektrum einer Klasse differenziert in Augenschein zu nehmen

Die Qualitätsagentur freut sich über Ihr Interesse an diesen wissenschaftlichen Grundlagen.

Dr. Otmar Schießl, OStD
Leiter der Qualitätsagentur

1. Darstellung der durchschnittlichen Testleistung

Die Ermittlung der Testleistungen der teilnehmenden Schülerinnen und Schüler erfolgte nicht aufgrund einer Schematik, die für das Lösen einer bestimmten Aufgabe eine bestimmte Punktzahl vorsieht. Stattdessen wurde eine Fähigkeitsskala verwendet. Eine Fähigkeitsskala stellt eine Maßeinheit zur Verfügung, um Ausprägungen von Kompetenzen quantifizieren zu können. Die bekannteste Fähigkeitsskala ist die PISA-Skala, deren Ausprägungen um den Wert 500 justiert wurden. Derartige Werte auf einer Fähigkeitsskala werden Fähigkeitsausprägung oder -parameter (Abkürzung θ , griechisch Theta) genannt.

Der größte Vorteil einer Auswertung von Fähigkeitsausprägungen besteht darin, dass Leistungen von Testteilnehmern direkt miteinander verglichen werden können, selbst wenn sie völlig unterschiedliche Aufgaben (derselben Fähigkeitsskala) bearbeitet haben. Somit wird es möglich sein, die Ergebnisse der diesjährigen Orientierungsarbeiten mit denen des Schuljahres 2008/09 zu vergleichen.

Anders als bei PISA liegen die Fähigkeitsausprägungen in der aktuellen Auswertung um den Wert 0. Der Wert 0 wurde so gewählt, dass er der durchschnittlichen Schülerleistung in diesem Jahr entspricht. Die Zuordnung einer Schülerleistung zu einer Fähigkeitsausprägung ergibt sich nach einer so genannten Skalierung aus den individuell gelösten Aufgaben. Je besser eine Schülerleistung ist, desto höher ist der Wert auf der Fähigkeitsskala.


Will man die Ausprägungen einer Fähigkeitsskala veranschaulichen, ergibt sich eine ähnliche Situation wie bei der Beschreibung der Celsius-Temperaturskala: Einzelne Werte sind leicht zu definieren (z. B. 0°C , 100°C bzw. $\theta = 0$); bei anderen Werten ist hingegen etwas mehr Mühe notwendig, um eine konkrete Vorstellung zu vermitteln. So kann man beispielsweise eine Temperatur von 13°C damit beschreiben, dass das Tragen einer Jacke angemessen ist, man mit Winterkleidung wie gefüttertem Mantel und Handschuhen eher schwitzen würde, wohingegen man mit T-Shirt und kurzer Hose frieren würde. Eine solche Beschreibung ist einerseits durchaus anschaulich, andererseits bleibt sie auch etwas vage und uneindeutig, denn sie wäre für $11,12,14$ oder 15°C ähnlich zutreffend.

Analog hierzu lässt sich eine Fähigkeitsskala am einfachsten anhand von Aufgaben(-merkmalen) veranschaulichen: Es lässt sich für jede Fähigkeitsausprägung die Wahrscheinlichkeit angeben, mit der eine Aufgabe gelöst wird. Tabelle A verdeutlicht diesen Zusammenhang. Dort ist beispielhaft für drei Aufgaben angegeben, wie deren Lösungswahrscheinlichkeit bei Fähigkeitsausprägungen im Bereich von $\theta = -3,5$ bis $\theta = 3,5$ variiert. Aus der Tabelle ergibt sich unter anderem, dass ein Schüler mit der Fähigkeitsausprägung $\theta = 2,0$ (viertletzte Spalte) die Aufgabe 2a („Stuhl“) mit einer Wahrscheinlichkeit von 99,3% löst, die Aufgabe 8a („singt“) mit 92,2% und die Aufgabe 4c („Ohrring“) mit 81,8%. Dieser Schüler hätte somit eine gute Chance, alle drei Aufgaben korrekt zu lösen. Dies ist anders bei einem Schüler mit der Fähigkeit $\theta = -2,0$. Bezogen auf PISA-Kriterien würde man annehmen, dass nur Aufgabe 5b hinreichend sicher von dem Schüler beherrscht wird ($p = 71,1\%$).

Es ist allerdings zu betonen, dass es sich bei den Angaben in Tabelle A um Wahrscheinlichkeitsaussagen handelt. Somit ist beispielsweise nicht sicher, dass ein Schüler mit der Fähigkeit $\theta = 2,0$ tatsächlich alle drei der aufgeführten Aufgaben löst. Würde man eine hinreichend große Schülergruppe betrachten, in der alle exakt die Fähigkeit $\theta = 2,0$ aufweisen, fände man, dass 74,9% alle drei Aufgaben lösen, 23,5% zwei Aufgaben und 1,6% nur eine Aufgabe.

Tabelle A dient der Veranschaulichung des Leistungsniveaus individueller Fähigkeitsausprägungen. Sie soll einen Eindruck von der verwendeten Skala vermitteln. Für die Interpretation der durchschnittlichen Klassenleistung kann sie nicht herangezogen werden.

Tabelle A: Veranschaulichung des Leistungsniveaus unterschiedlicher Fähigkeitsausprägung

		Aufgabenbezogene Lösungswahrscheinlichkeit für Schülerinnen und Schüler mit der Fähigkeitsausprägung $\theta =$														
		-3,5	-3,0	-2,5	-2,0	-1,5	-1,0	-0,5	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5
Aufgabe 5b	Setze die fehlenden Buchstaben richtig ein.															
	 _____uhl	35,4%	47,5%	59,9%	71,1%	80,2%	87,0%	91,7%	94,8%	96,8%	98,0	98,8%	99,3%	99,6%	99,7%	99,8%
Aufgabe 8a	Überprüfe die fettgedruckten Wörter. Schreibe sie dann richtig auf die Zeilen.															
	Das Kind sinkt ein Lied. Das Kind _____ ein Lied.	4,6%	7,4%	11,6%	17,8%	26,3%	37,1%	49,3%	61,5%	72,5%	81,3	87,8%	92,2%	95,1%	97,0%	98,1%
Aufgabe 4c	Nicht alle zusammengesetzten Wörter sind richtig geschrieben. Schreibe <u>alle</u> Wörter richtig auf.															
	Ohring _____	1,8%	2,9%	4,7%	7,6%	11,9%	18,2%	26,9%	37,8%	50,0%	62,2%	73,1%	81,8%	88,1%	92,4%	95,3%
$\theta =$		-3,5	-3,0	-2,5	-2,0	-1,5	-1,0	-0,5	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5
Fähigkeitsniveau		unter 1		1			2		3							

2. Berücksichtigung von Fähigkeitsniveaus

2.1 Was sind Fähigkeitsniveaus?

Fähigkeitsniveaus werden gebildet, indem eine Fähigkeitsskala in verschiedene Abschnitte unterteilt wird. Die Unterteilung der hier verwendeten Fähigkeitsskala ist im unteren Bereich der Tabelle A angezeigt.

Es sind verschiedene Vorgehensweisen möglich, um auf einer Fähigkeitsskala Schwellen zur Unterteilung interpretierbarer und kommunizierbarer Niveaus zu bestimmen. Bevor die entsprechenden Niveaustufen als empirisch gesichert und didaktisch fundiert gelten können, ist eine Reihe von interdisziplinären Arbeitsschritten notwendig. Diese Arbeiten befinden sich für die Orientierungsarbeiten Rechtschreiben in der Jahrgangsstufe 2 noch in einem sehr frühen Stadium. Aus diesem Grund müssen die Einteilung der Fähigkeitsniveaus bei der diesjährigen Durchführung der Orientierungsarbeiten als vorläufig betrachtet werden.

Das **Fähigkeitsniveau 1** spiegelt **grundlegende Fähigkeiten** wider. Schülerinnen und Schüler auf diesem Niveau kennen die richtige Schreibweise vertrauter Wörter und können einfache Ableitungen richtig umsetzen. Sie erkennen einzelne Buchstabengruppen und nutzen diese für das Richtigschreiben. Die Laut-Buchstaben-Zuordnung gelingt meistens. Häufig vorkommende einfache Wörter werden mehrheitlich richtig geschrieben. Die Groß- bzw. Kleinschreibung von Substantiven und Verben ist größtenteils gesichert. Aufgabe 5b (vgl. Tabelle A) stellt eine Beispielaufgabe aus dem (unteren) Anforderungsbereich des Fähigkeitsniveaus 1 dar. Entsprechend den internationalen Gepflogenheiten bei der Einteilung von Fähigkeitsniveaus haben Schülerinnen und Schüler auf dem Niveau 1 bei dieser Aufgabe eine Lösungswahrscheinlichkeit von deutlich über 60%.

Das **Fähigkeitsniveau 2** bezieht sich auf **erweiterte Fähigkeiten**. Schülerinnen und Schüler auf diesem Niveau denken über Schreibweisen nach und können fehlerhafte Schreibweisen berichtigen. Die Laut-Buchstaben-Zuordnung gelingt fast immer. Elementares Regelwissen ist verfügbar und bei Bedarf werden einfache Rechtschreibstrategien angewandt. Die Auslautverhärtung stellt größtenteils kein Problem mehr dar. Beispielhaft verdeutlicht Aufgabe 8a (vgl. Tabelle A) die Anforderungen des Fähigkeitsniveaus 2.

Das **Fähigkeitsniveau 3** kennzeichnet **fortgeschrittene Fähigkeiten**. Schülerinnen und Schüler auf diesem Niveau haben erste Einsichten in die Prinzipien der Rechtschreibung gewonnen. Ein grundsätzliches Rechtschreibgespür ist erkennbar und erstes morphematisches Wissen kann festgestellt werden. Das alphabetische Ordnen der Wörter nach dem Erst- und teilweise auch nach dem Zweitbuchstaben gelingt überwiegend. Falsche Getrennschreibungen werden vermieden. Bei Aufgabe 4c (vgl. Tabelle A) handelt es sich um eine Beispielaufgabe für die Anforderungen des Fähigkeitsniveaus 3.

Die Kompetenzen von Schülerinnen und Schülern, deren Testleistung **unter Fähigkeitsniveau 1** liegen, lassen sich nur schwer beschreiben. Insgesamt sollte daher die eher zurückhaltende Interpretation vorgenommen werden, dass kein hinreichender Nachweis für das Erreichen von Fähigkeitsniveau 1 vorliegt.

2.2 Bewertung des Klassenergebnisses anhand der Fähigkeitsniveaus

Abschnitt 2 der Ergebnismeldung („Verteilung auf Fähigkeitsniveaus“) bietet zwei Beurteilungsebenen. Zum einen ist eine kriteriumsbezogene Beurteilung möglich. Hierfür ist ein Vergleich mit den landesweiten Ergebnissen eher unwichtig. Vielmehr geht es um den persönlichen oder schulischen Anspruch, wie groß (oder klein) der Anteil der Schülerschaft in der ausgewerteten Klasse sein soll, der sich jeweils auf einem bestimmten Fähigkeitsniveau befindet.

Die zweite Beurteilungsebene im Abschnitt 2 der Ergebnismeldung bezieht dann durch die Berücksichtigung der landesweiten Ergebnisse wieder vergleichende Gesichtspunkte ein. Eine einfache Betrachtung der landesweiten Durchschnittsverteilung im unteren Teil der Abbildung 2 reicht dabei noch nicht aus. Es stellt sich nämlich die Frage, welche Abweichungen von dieser Durchschnittsverteilung überhaupt „üblich“ sind. So macht der Anteil der Schülerinnen und Schüler auf Fähigkeitsniveau 2 in einigen Klassen weniger als 10% aus, in anderen hingegen mehr als 80%. Vor diesem Hintergrund wurde bei der aktuellen Auswertung aus den Landeswerten für jede Niveaustufe ein prozentualer

Kernbereich ermittelt, der per se als unauffällig („üblich“) gelten muss. Für die Interessierten ist diese Ermittlung anhand eines fiktiven Beispiels im Kasten 1 veranschaulicht.

Kasten 1: Fiktives Beispiel zur Ermittlung eines Kernbereichs

An einer Vergleichsarbeit haben zwölf Klassen teilgenommen. Insgesamt liegen 30% der Schülerinnen und Schüler auf Fähigkeitsniveau 2. Der Anteil der Schülerinnen und Schüler auf Fähigkeitsniveau 2 variiert in den zwölf Klassen folgendermaßen:

Anteil in Prozent	Anzahl der Klassen	Zugehörigkeit zum Kernbereich
22	1	nein
23		
24	1	nein
25		
26	1	nein
27	1	ja
28		
29	1	ja
30	2	ja
31		
32	2	ja
33		
34	1	nein
35		
36	1	nein
37		
38		
39	1	nein

Der Anteil der Schülerinnen und Schüler auf Fähigkeitsniveau 2 macht bei dem Beispiel im Minimalfall 22% aus, im Maximalfall 39%. Bei drei Klassen liegt der Anteil unter 27%, bei 3 Klassen über 32% und bei den „mittleren“ sechs Klassen zwischen 27 und 32%. Diese sechs Klassen stellen somit einerseits die Hälfte der Gesamtklassen dar und machen andererseits die „Mitte“ in Bezug auf die Anteile von Schülerinnen und Schülern auf Leistungsniveau 2 aus. Daher werden diese sechs Klassen zur Bestimmung des Kernbereichs für Fähigkeitsniveau 2 herangezogen; der entsprechende Kernbereich liegt also zwischen 27 und 32% in Bezug auf die Schüleranteile. Analog müssten die Kernbereiche für die anderen Fähigkeitsniveaus ermittelt werden.

Während im Beispiel von Kasten 1 nur zwölf Klassen zur Ermittlung des Kernbereichs herangezogen wurden, liegen der aktuellen Auswertung Daten von mehr als 5000 Klassen zu Grunde. Die hieraus für die einzelnen Fähigkeitsniveaus resultierenden Kernbereiche sind in Tabelle B dargestellt.

Tabelle B: Empirisch ermittelte Kernbereiche

Fähigkeitsniveau	Kernbereich
unter 1	0-5%
1	15-36%
2	32-48%
3	17-41%

Zwei Gründe lassen es sinnvoll erscheinen, nicht gleich von einem „auffälligen“ Ergebnis zu sprechen, wenn eine Klasse das eine oder andere Mal außerhalb der Kernbereiche der Fähigkeitsniveaus liegt. Zum einen ergibt sich aus der Darstellung in Kasten 1, dass bei jedem Fähigkeitsniveau jeweils ein Viertel der Klassen oberhalb und ein Viertel unterhalb

des Kernbereichs liegen. Es ist unangemessen, immer insgesamt die Hälfte der Klassenergebnisse als auffällig zu bewerten. Der zweite Grund, den Kernbereich nicht allein zur Beurteilung heranzuziehen, besteht darin, dass individuelle Testleistungen immer mit gewissen Zufälligkeiten verbunden sind. So kann durch eine kleine Unachtsamkeit eine Aufgabe falsch gelöst worden sein, obwohl der Schüler die Aufgabe eigentlich beherrscht. Oder durch willkürliches Raten kann eine Aufgabe richtig beantwortet worden sein. Aufgrund dieser Zufälligkeiten kann ein Schüler, der sich im Grenzbereich zwischen zwei Fähigkeitsniveaus befindet, auch unangemessen einem Niveau zugeordnet werden.

Aus diesen beiden Gründen wird in der vorliegenden Auswertung zusätzlich zum Kernbereich noch ein „Sicherheitsbereich“ berücksichtigt: Nur wenn die Abweichung vom Kernbereich nicht auf ein (Zufalls-)Ergebnis eines einzelnen Schülers zurückgeführt werden kann, ist ein Unterschied zwischen Klassen- und Landesergebnis in Abbildung 2 der Ergebnisrückmeldung hervorgehoben. Da der angelegte Sicherheitsbereich dem „prozentualen Gewicht“ eines einzelnen Schülers entspricht, ist er abhängig von der Klassengröße. Der Sicherheitsbereich kann den Ausführungen des Abschnitts 3 der Klassenrückmeldung entnommen werden. Dort ist angegeben, wie viel Prozent ein einzelner Schüler aufgrund der Klassengröße ausmacht.

Sind Unterschiede zwischen Klassen- und Landesergebnis in Abbildung 2 der Ergebnisrückmeldung hervorgehoben, lohnt es sich, über die möglichen Ursachen nachzudenken. Darüber hinaus sei an dieser Stelle noch einmal auf die Möglichkeit der weiter oben erwähnten kriteriumsbezogenen Beurteilung erinnert: Die persönlichen oder schulischen Ansprüche an die klassenspezifische Verteilung der Fähigkeitsniveaus können durchaus unabhängig sein vom Vergleich zwischen Klassen- und Landesergebnis.

2.3 Tatsächliche und erwartete Lösungsquoten

Bei der Auswertung des Gesamttests im Abschnitt 2 der Ergebnisrückmeldung standen so genannte soziale Vergleiche im Vordergrund: Das Ergebnis der Klasse wurde zur Gesamtheit der bayerischen Klassen in Bezug gesetzt. Bei der Analyse der Einzelaufgaben im Abschnitt 4 des Ergebnisberichts wurde hingegen kein Bezug zu einer externen Gruppe hergestellt. Vielmehr kann das Leistungsniveau Ihrer Klasse für die Ermittlung eigener Referenzwerte herangezogen werden. Möglich wird dies aufgrund der vorne erläuterten Eigenschaft von Fähigkeitsskalen der direkten Beziehung zwischen Fähigkeitsausprägungen und Lösungswahrscheinlichkeiten.

Der so genannte Schwierigkeitsparameter (der „Schwierigkeitsgrad“) jeder Aufgabe ist bekannt aus der repräsentativen Referenzgruppe für die Orientierungsarbeiten 2007/08. Auf der Grundlage dieser Schwierigkeitsparameter lassen sich dann die Fähigkeitsausprägungen für alle Testteilnehmer ermitteln (s. o.). Gleichzeitig können aber auch aus der Verteilung der Fähigkeitsausprägungen in einer Schülergruppe die für diese Gruppe zu erwartenden Lösungswahrscheinlichkeiten der Einzelaufgaben berechnet werden (analog zur Darstellung in Tabelle A). Solche Berechnungen lassen dann wiederum für eine Art „Gegenprobe“ nutzen, bei der eine Gegenüberstellung von tatsächlichen und erwartenden Lösungshäufigkeiten vorgenommen wird. Während landesweit über alle Testteilnehmer bei dieser „Gegenprobe“ eine perfekte Übereinstimmung zwischen erwarteten und tatsächlichen Lösungshäufigkeiten zu finden ist, können sich in homogenen Untergruppen wie Klassenverbänden durchaus Unterschiede bei einzelnen Aufgaben ergeben.

Eine solche „Gegenprobe“ liegt der Auswertung im Abschnitt 4 der Ergebnisrückmeldung zugrunde. Dabei ist sehr wohl berücksichtigt worden, dass einige falsche oder richtige Schülerlösungen in einer Klasse immer auch auf Zufälligkeiten zurückzuführen sein können. Dementsprechend sind Abweichungen zwischen tatsächlichen und erwarteten Lösungshäufigkeiten, die auf möglicherweise zufällige Antworten von zwei oder drei Schülern zurückzuführen sind, in der Darstellung des Ergebnisberichts nicht weiter hervorgehoben.