



STAATSINSTITUT FÜR SCHULQUALITÄT  
UND BILDUNGSFORSCHUNG  
MÜNCHEN

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

---

## **Ergänzende Hinweise zu den klassenbezogenen Rückmeldungen von Testergebnissen**

**aus den Vergleichsarbeiten 2008  
Mathematik Jahrgangsstufe 8**

Die Vergleichsarbeiten Mathematik Jahrgangsstufe 8 stellen ein länderübergreifendes Projekt im Sinne der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring dar. Die Lernstandserhebung ist im Frühjahr 2008 zeitgleich in den Ländern Bayern, Berlin, Brandenburg, Bremen, Hamburg, Hessen, Niedersachsen, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt, Schleswig-Holstein und Thüringen durchgeführt worden. Das länderfinanzierte Institut zur Qualitätssicherung im Bildungswesen (IQB) an der Humboldt-Universität in Berlin erhielt im Frühsommer 2007 den Auftrag, für die Testdurchführung einen kompetenzorientierten Aufgabenpool bereitzustellen. Dabei sollten die Aufgaben einerseits auf ihre Tauglichkeit gründlich vorgetestet werden und andererseits anschlussfähig sein an die „Metrik“ der Bildungsstandards für den mittleren Abschluss. Tatsächlich ist es aufgrund der Vorarbeit des IQB möglich, die Testleistung der Schülerinnen und Schüler auf der Fähigkeitsskala (vgl. Begriffserklärung in Abschnitt 1 dieser Erläuterung) der Bildungsstandards zu beschreiben.<sup>1</sup>

Mit der Bereitstellung der Aufgabenhefte und den dazugehörigen Korrekturanweisungen endete die Verantwortlichkeit des IQB. Dementsprechend finden ausschließlich länderinterne Auswertungen statt und keine zentralen Analysen oder gar Ländervergleiche. In Bayern ist ein wichtiger Anspruch an diese Auswertungen, den Lehrerinnen und Lehrern Rückmeldung über die Ergebnisse ihrer Klassen zu geben. Dieser Anspruch liegt den klassenbezogenen Ergebnisberichten zugrunde. In den per OWA verschickten Ergebnisberichten erfolgte eine knappe Darstellung der wichtigsten klassenbezogenen Ergebnisse. Im Folgenden gibt es einige Hintergrundinformationen zu den Analysen des Gesamttests bzw. zu den Leitideen (1.) und der Einzelaufgaben (2.).

### **1. Ergebnisse im Gesamttest und bei den Leitideen**

Die Ermittlung der Testleistungen der teilnehmenden Schülerinnen und Schüler erfolgte nicht aufgrund eines Schemas, das 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  $\theta$ , 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 Vergleichsarbeiten mit denen des Schuljahres 2008/09 zu vergleichen.

Ähnlich wie bei PISA liegen die Fähigkeitsausprägungen in der aktuellen Auswertung um den Wert 500. Der Wert 500 wurde so gewählt, dass er der durchschnittlichen Schülerleistung bei der Normierung der Bildungsstandards für den mittleren Bildungsabschluss im Jahre 2006 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.  $0^{\circ}\text{C}$ ,  $100^{\circ}\text{C}$  bzw.  $\theta = 500$ ); bei anderen Werten ist hingegen etwas mehr Mühe notwendig, um eine konkrete Vorstellung zu vermitteln. So kann man beispielsweise eine Temperatur von  $13^{\circ}\text{C}$  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  $15^{\circ}\text{C}$  ähnlich zutreffend.

Analog hierzu lässt sich eine Fähigkeitsskala am einfachsten anhand von Aufgaben(-merkmalen) veranschaulichen: Es lässt sich für jede Fähigkeitsausprägung die

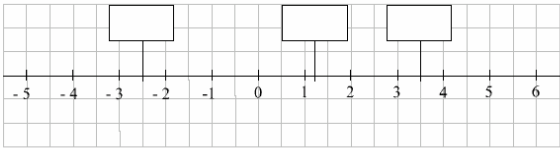
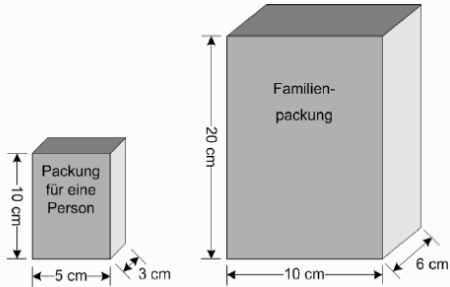
<sup>1</sup> In Bayern wurde bei der diesjährigen Auswertung auf eine Zuordnung zu definierten Kompetenzstufen verzichtet. Der gegenwärtige Vorschlag des IQB dafür bedarf noch weiterer Klärung.

Wahrscheinlichkeit angeben, mit der eine Aufgabe gelöst wird. Tabelle 1 verdeutlicht diesen Zusammenhang. Dort ist beispielhaft für drei Aufgaben angegeben, wie deren Lösungswahrscheinlichkeit bei Fähigkeitsausprägungen im Bereich von  $\theta = 200$  bis  $\theta = 800$  variiert. Aus der Tabelle ergibt sich unter anderem, dass ein Schüler mit der Fähigkeitsausprägung  $\theta = 700$  (drittletzte Spalte) die Aufgabe 15 mit einer Wahrscheinlichkeit von 97,5% löst, die Aufgabe 23 mit 88,5% und die Aufgabe 21 mit 65,5%. Dieser Schüler hätte somit eine gute Chance, alle drei Aufgaben korrekt zu lösen. Dies ist anders bei einem Schüler mit der Fähigkeit  $\theta = 450$ . Bezogen auf Kriterien der Bildungsstandards für den mittleren Abschluss würde man annehmen, dass nur Aufgabe 15 hinreichend sicher von dem Schüler beherrscht wird ( $p = 72,6\%$ ).

Es ist allerdings zu betonen, dass es sich bei den Angaben in Tabelle 1 um Wahrscheinlichkeitsaussagen handelt. Somit ist beispielsweise nicht sicher, dass ein Schüler mit der Fähigkeit  $\theta = 700$  tatsächlich alle drei der aufgeführten Aufgaben löst. Würde man eine hinreichend große Schülergruppe betrachten, in der alle exakt die Fähigkeit  $\theta = 700$  aufweisen, fände man, dass 56,9% alle drei Aufgaben lösen, 38,3% zwei Aufgaben, 4,8% nur eine Aufgabe und 0,1% sogar gar keine.

Tabelle 1 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 1: Veranschaulichung des Leistungsniveaus unterschiedlicher Fähigkeitsausprägungen**

		Aufgabenbezogene Lösungswahrscheinlichkeit für Schülerinnen und Schüler mit der Fähigkeitsausprägung $\theta =$												
		200	250	300	350	400	450	500	550	600	650	700	750	800
<p><b>Aufgabe 15: Zahlenstrahl</b></p>  <p>Trage in die leeren Kästchen die zugehörigen Zahlen ein.</p>		15,5%	23,8%	34,8%	47,7%	60,8%	72,6%	81,9%	88,5%	92,9%	95,7%	97,5%	98,5%	99,1%
<p><b>Aufgabe 23: Cornflakes</b></p>  <p>Die beiden abgebildeten Packungen für Cornflakes haben die gleiche Form und sind beide vollständig mit Cornflakes gefüllt. Die kleine Packung enthält die Menge Cornflakes, die normalerweise für eine Person reicht. Wie viele solcher Portionen Cornflakes enthält dann die Familienpackung? Kreuze an.</p> <p><input type="checkbox"/> 2    <input type="checkbox"/> 4    <input type="checkbox"/> 6    <input type="checkbox"/> 8    <input type="checkbox"/> 12</p>		3,5%	5,9%	9,6%	15,4%	23,7%	34,6%	47,5%	60,7%	72,5%	81,8%	88,5%	92,9%	95,7%
<p><b>Aufgabe 21: Runden</b></p> <p>Zwei verschiedene natürliche Zahlen werden auf Zehner gerundet. In beiden Fällen erhält man 20. Um wie viele Einer können sich die beiden Zahlen höchstens unterscheiden? Kreuze an.</p> <p><input type="checkbox"/> Um 3 Einer            <input type="checkbox"/> Um 4 Einer  <input type="checkbox"/> Um 5 Einer            <input type="checkbox"/> Um 9 Einer  <input type="checkbox"/> Um 10 Einer</p>		0,9%	1,5%	2,6%	4,4%	7,2%	11,8%	18,5%	28,0%	39,9%	53,1%	65,9%	76,7%	84,9%

## **2. Tatsächliche und erwartete Lösungsquoten**

Bei der Auswertung des Gesamttests bzw. der Leitideen im Abschnitt 1 des klassenbezogenen Ergebnisberichts standen so genannte soziale Vergleiche im Vordergrund: Die Ergebnisse der jeweiligen Klasse wurden zu einer Referenzgruppe in Bezug gesetzt. Bei der Analyse der Einzelaufgaben im Abschnitt 2 des klassenbezogenen Ergebnisberichts wurde hingegen kein Bezug zu einer externen Gruppe hergestellt. Vielmehr kann das Leistungsniveau jeder 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 großen, bundesweiten Voruntersuchungen (Normierung der Bildungsstandards im Jahre 2006 und Feldtest zur Erprobung der aktuell verwendeten Aufgaben). Auf der Grundlage dieser Schwierigkeitsparameter lassen sich dann die Fähigkeitsausprägungen für alle Testteilnehmer ermitteln (vgl. Abschnitt 1 dieser Erläuterung). Gleichzeitig können aber auch aus der Verteilung der Fähigkeitsausprägungen in einer Schülergruppe die für diese Gruppe erwarteten Lösungswahrscheinlichkeiten der Einzelaufgaben berechnet werden (analog zur Darstellung in Tabelle 1). Solche Berechnungen lassen sich dann wiederum für eine Art „Gegenprobe“ nutzen, bei der eine Gegenüberstellung von tatsächlichen und erwarteten Lösungshäufigkeiten vorgenommen wird. Während bundesweit über alle Testteilnehmer bei dieser „Gegenprobe“ eine nahezu 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 2 des klassenbezogenen Ergebnisberichts 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 Tabelle 4 des klassenbezogenen Ergebnisberichts nicht weiter hervorgehoben.