



Institut für Bildungsevaluation

---

## **Check P4 Zug 2024**

Kantonsbericht zuhanden der Direktion für Bildung und Kultur, Kanton Zug

Zürich, 16. September 2024

## Anschrift

---

Hofwiesenstrasse 349  
8050 Zürich  
Telefon +41 43 268 39 60  
[www.ibe-edu.ch](http://www.ibe-edu.ch)  
[benjamin.wolf@ibe-edu.ch](mailto:benjamin.wolf@ibe-edu.ch)

---

## Zusammenfassung

Der Check P4 ermöglicht eine externe Standortbestimmung in den Fächern Deutsch, Englisch und Mathematik anhand standardisierter Online-Tests. Im Mai 2024 wurde der Check P4 in allen vierten Klassen der gemeindlichen Schulen im Kanton Zug durchgeführt. Die Ergebnisse dienen der individuellen Beurteilung sowie der Unterrichts- und Schulentwicklung. Darüber hinaus können die Ergebnisse für das Monitoring des kantonalen Bildungssystems verwendet werden, um beispielsweise Zusammenhänge mit individuellen Merkmalen (Geschlecht, Erstsprache, Alter, Lernziele) zu analysieren oder die Erreichung der Orientierungspunkte im zweiten Zyklus des Lehrplans 21 zu überprüfen. Die regelmässige Durchführung des Check P4 ermöglicht künftig den Vergleich der Leistungen in den getesteten Kompetenzbereichen über die Jahre hinweg.

Der vorliegende Bericht informiert über die Ergebnisse des Check P4 2024 auf kantonaler Ebene. Die Ergebnisse zeigen, dass eine besondere Stärke der Zuger Schülerinnen und Schüler im Fach Englisch liegt. In diesem Fach erreichen oder übertreffen mehr als 80% der Schülerinnen und Schüler die Kompetenzen, welche bis zum Orientierungspunkt im 2. Zyklus des Lehrplans 21 vorgesehen sind. In Deutsch und Mathematik beträgt dieser Anteil rund zwei Drittel. Der Lehrplan 21 sieht vor, dass die als Orientierungspunkt markierten Kompetenzstufen bis zum Ende der 4. Klasse im Unterricht behandelt werden. Das Erreichen der jeweiligen Kompetenzstufen ist bis zum Ende des 2. Zyklus vorgesehen.

Des Weiteren zeigen die Analysen, dass Schülerinnen und Schüler mit Deutsch als Erstsprache im Fach Englisch durchschnittlich weniger hohe Ergebnisse erzielen als Schülerinnen und Schüler mit Deutsch als Zweitsprache. Eine mögliche Erklärung für diesen Befund ist der im Kanton Zug vergleichsweise hohe Anteil an Kindern von internationalen Fach- und Führungskräften, deren Erstsprache häufig Englisch ist. In den Fächern Deutsch und Mathematik sind hingegen die Schülerinnen und Schüler mit Deutsch als Erstsprache im Vorteil. Ein weiterer interessanter Befund ist, dass Knaben gegenüber Mädchen im Fach Mathematik im Durchschnitt deutlich bessere Ergebnisse erzielen, wobei hier insbesondere das Ausmass überrascht.

Zusätzlich zu den Unterschieden zwischen Gruppen mit verschiedenen individuellen Merkmalen wurden auch die Unterschiede zwischen den Schulklassen untersucht. In allen drei geprüften Fächern zeigte sich eine grosse Spannbreite der Klassenergebnisse. Dies unterstreicht die Bedeutung einer vom Klassenkontext unabhängigen Standortbestimmung als Ergänzung zu den stärker klassenbezogenen Schulnoten.

Im Rahmen der Pilot-Durchführung wurden zwei Testdesigns – adaptive Tests und lineare Tests – eingesetzt und deren Ergebnisse und methodische Qualität verglichen. Der Vergleich zeigte, dass das adaptive Testdesign den beabsichtigten Zweck erfüllt und mehrere Vorteile gegenüber dem linearen Testdesign hat. Das adaptive Testdesign führte in allen fünf Tests zu einer besseren Passung zwischen Schwierigkeit der Aufgaben und Fähigkeit der Schülerinnen und Schüler. Zudem ermöglichte das adaptive Testdesign für die Mehrheit der Schülerinnen und Schüler in allen Tests eine genauere Messung der Fähigkeit bei gleicher Testlänge und -dauer.

Während des Schuljahres 2023/2024 hatten die Schülerinnen und Schüler der 4. Klasse Zugang zur Aufgabensammlung Mindsteps. Die Analyse der Nutzerzahlen zeigt, dass die grosse Mehrheit der Schülerinnen und Schüler mit Mindsteps arbeitete. Zudem geht in allen drei Fächern eine häufigere Nutzung von Mindsteps mit besseren Ergebnissen im Check P4 einher. Dies könnte darauf hindeuten, dass Schülerinnen und Schüler mit Mindsteps einen Lern- und Übungseffekt erzielen. Allerdings ist auch denkbar, dass weitere Einflussfaktoren – bspw. der Unterricht oder die Bedingungen an der Schule – sowohl die Fähigkeiten der Schülerinnen und Schüler als auch die Nutzung von Mindsteps positiv beeinflussen.

Zum Zeitpunkt der Ergebnisrückmeldung wurden die teilnehmenden Klassenlehrpersonen und Schulleitungen mit einem Online-Fragebogen zum Check P4 befragt. Im Fokus dieser Befragung standen Rückmeldungen zur Check-Durchführung. Als Stärken des Check P4 wurden die ausführlichen Informationen zur Vorbereitung, der Support durch

das IBE sowie die Form der Ergebnisrückmeldung genannt. Weniger positiv wurde die inhaltliche Abstimmung der Tests auf die Lehrmittel und auf die individuelle Unterrichtsplanung beurteilt. Die Analyse der Aufgabenschwierigkeiten und der Fähigkeiten der Schülerinnen und Schüler zeigt jedoch, dass insbesondere mit dem adaptiven Testdesign eine gute Passung erreicht wurde.

Kritische Rückmeldungen betrafen den Durchführungszeitpunkt des Check P4. Einerseits sind im Mai noch nicht alle Themen der 4. Klasse im Unterricht behandelt worden, andererseits können die individuellen Ergebnisse von den Lehrpersonen der 4. Klassen nicht mehr zur Förderung genutzt werden. Stattdessen werden die Ergebnisse an die Lehrpersonen der 5. Klassen weitergegeben. Unter diesen Voraussetzungen wurde der Aufwand für die Durchführung von einigen Lehrpersonen als zu hoch beurteilt.

Fazit: Der Check P4 2024 liefert wertvolle Erkenntnisse über den Lernstand der Schülerinnen und Schüler der 4. Klasse im Kanton Zug. Besonders hervorzuheben sind die Stärke der Schülerinnen und Schüler im Fach Englisch, die Vorteile des adaptiven Testdesigns und der positive Zusammenhang zwischen der Nutzung von Mindsteps und den Check-Ergebnissen. Weniger positiv wurde die inhaltliche Abstimmung der Tests auf die Lehrmittel und auf die individuelle Unterrichtsplanung beurteilt. Weitere Aussagekraft erhalten die Check-Ergebnisse bei wiederholter Durchführung, wodurch die Analyse von Trends möglich wird.

# Inhalt

<b>1</b>	<b>Einleitung</b>	<b>6</b>
1.1	Ausgangslage	6
1.2	Tests und Methoden	6
1.3	Teilnehmende Schülerinnen und Schüler	8
<b>2</b>	<b>Ergebnisse in Deutsch, Englisch und Mathematik</b>	<b>9</b>
2.1	Mittelwert und Streuung der Ergebnisse	9
2.2	Ergebnisse nach individuellen Merkmalen	10
2.3	Zusammenhang zwischen den Fächern	12
2.4	Unterschiede zwischen Klassen	13
2.5	Ergebnisse nach Kompetenzstufen	14
<b>3</b>	<b>Vergleich adaptives und lineares Testdesign</b>	<b>16</b>
3.1	Methodische Qualität des adaptiven und des linearen Testdesigns	17
3.2	Ergebnisse im adaptiven und linearen Testdesign	20
<b>4</b>	<b>Nutzung von Mindsteps und Ergebnisse im Check P4</b>	<b>21</b>
4.1	Nutzung von Mindsteps	22
4.2	Zusammenhang zwischen der Nutzung von Mindsteps und den Ergebnissen im Check P4	22
<b>5</b>	<b>Ergebnisse der Befragung zum Check P4</b>	<b>24</b>
5.1	Zusammenfassung der wichtigsten Rückmeldungen	24
5.2	Vollständige Ergebnisse der geschlossenen Fragen	26

# 1 Einleitung

## 1.1 Ausgangslage

Der Check P4 ermöglicht eine externe Leistungsmessung in den Fächern Deutsch, Englisch und Mathematik anhand standardisierter Online-Tests. Die Ergebnisse dienen der unabhängigen Standortbestimmung für Schülerinnen und Schüler, Lehrpersonen und Schulleitungen. Sie können für die individuelle Beurteilung, die Unterrichts- und Schulentwicklung sowie für das Monitoring des kantonalen Bildungssystems genutzt werden.

Der Check P4 fand im Mai 2024 in allen 4. Klassen der gemeindlichen Schulen im Kanton Zug statt. Dabei handelte es sich um eine Pilot-Durchführung. Ab Schuljahr 2024/25 wird die reguläre Durchführung des Check P4 jeweils im Frühling stattfinden.

Die Ergebnisse der Schülerinnen und Schüler sind seit dem 21. Juni 2024 auf der Onlineplattform *Check Dein Wissen* für die teilnehmenden Lehrpersonen verfügbar. Schulleitungen haben zudem Zugriff auf die aggregierten Ergebnisse der Klassen und der Schule.

Neben den Rückmeldungen für Schülerinnen und Schüler, Lehrpersonen, Schulleitungen und den Kanton verfolgte die Pilot-Durchführung drei zusätzliche Ziele. Erstens wurden zwei Testdesigns (adaptive Tests und lineare Tests) eingesetzt und die Ergebnisse sowie die methodische Qualität verglichen. Zweitens sollten mögliche Zusammenhänge zwischen den Ergebnissen im Check P4 und der Nutzung der Online-Aufgabensammlung Mindsteps untersucht werden. Drittens wurden die teilnehmenden Lehrpersonen mit einem Online-Fragebogen zur Akzeptanz der Tests und der Ergebnisse sowie zur Zufriedenheit mit dem Durchführungsprozess befragt.

Der vorliegende Bericht informiert über die Ergebnisse des Check P4 2024 auf kantonaler Ebene. Zunächst werden die standardisierten Online-Tests (Abs. 1.2) und die teilnehmenden Schülerinnen und Schüler (Abs. 1.3) beschrieben. Danach werden in Abschnitt 2 die Ergebnisse in den geprüften Fächern berichtet und mögliche Zusammenhänge mit individuellen Merkmalen sowie die Zusammenhänge zwischen den Fächern gezeigt. Ebenfalls berichtet werden in Abschnitt 2 die von den Schülerinnen und Schülern erreichten Kompetenzstufen, insbesondere im Hinblick auf den Orientierungspunkt im 2. Zyklus des Lehrplans 21. Abschnitt 3 vergleicht die zwei Testdesigns, die im Rahmen der Pilot-Durchführung des Check P4 zur Anwendung kamen, bezüglich methodischer Qualität und erzielter Ergebnisse. In Abschnitt 4 wird die Frage untersucht, inwiefern die Testergebnisse mit der Verwendung der digitalen Aufgabensammlung Mindsteps in Zusammenhang stehen. Abschliessend werden in Abschnitt 5 die Rückmeldungen der Lehrpersonen zum Check P4 zusammengefasst.

## 1.2 Tests und Methoden

Der Check P4 umfasst fünf Online-Tests zu insgesamt sieben Kompetenzbereichen des Lehrplans 21. Die drei Kompetenzbereiche des Fachs Mathematik wurden mit einem Test geprüft. Für die je zwei Kompetenzbereiche in Deutsch und Englisch wurde jeweils ein eigener Test durchgeführt. Jeder Test umfasste 29 bis 48 Aufgaben, die sich auf vier aufeinander folgende Testteile verteilten. Die Aufteilung der Kompetenzbereiche auf die Tests sowie die Anzahl der Aufgaben und die gesamte Dauer der Tests werden in Tabelle 1 aufgelistet. Die verfügbare Zeit für das Bearbeiten der Aufgaben war pro Testteil begrenzt.

Tabelle 1: Fächer, Kompetenzbereiche und Umfang der durchgeführten Tests

Fach	Kompetenzbereiche	Anz. Aufgaben	Testdauer
Deutsch (2 Tests)	Lesen	32	40 min
	Sprache im Fokus	29 – 32 *	20 min
Englisch (2 Tests)	Hören	34	40 min
	Lesen	32	40 min
Mathematik (1 Test)	Zahl und Variable	16	80 min
	Form und Raum	16	
	Grössen, Funktionen, Daten und Zufall	16	

\* Im adaptiven Testdesign umfasste der einfache Aufgabenblock jeweils eine Aufgabe weniger pro Testteil. Im linearen Testdesign umfasste der Test 32 Aufgaben.

Rund drei Viertel der Tests folgten einem adaptiven Testdesign, eine zufällige Auswahl von einem Viertel der Tests folgte hingegen einem linearen Testdesign. Ein adaptives Testdesign bedeutet, dass die Schwierigkeit der Aufgaben den Fähigkeiten der Schülerinnen und Schüler angepasst wird. Ziel dieses Designs ist es, jeden Schüler und jede Schülerin optimal herauszufordern, sodass er oder sie keinen deutlich über- oder unterfordernden Aufgaben begegnet. Die Anpassung der Aufgabenschwierigkeit bezweckt zudem eine präzisere und verlässlichere Messung der Fähigkeiten der Schülerinnen und Schüler.

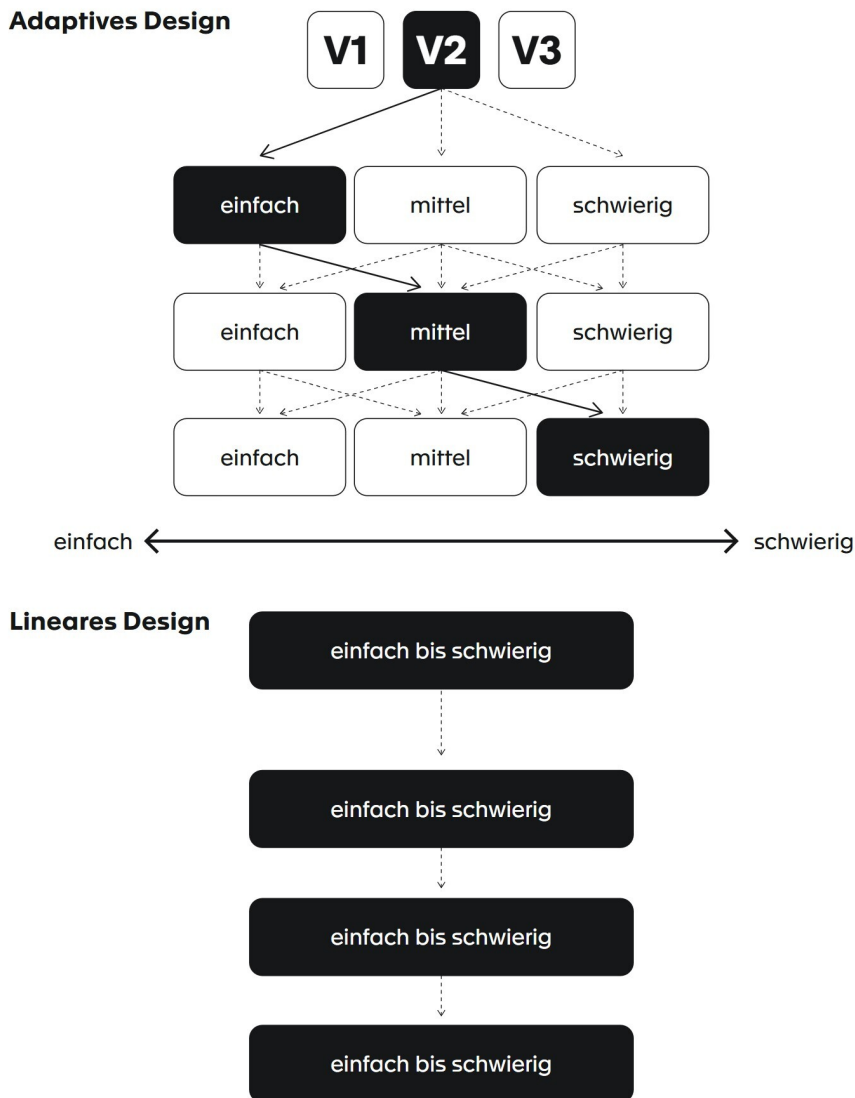
Das adaptive Testdesign ist in Abbildung 1 (oben) visualisiert. Für den ersten Testteil wird jeweils zufällig einer von drei gleich schwierigen Aufgabenblöcken ausgewählt (V1, V2, V3). Für die darauffolgenden Testteile wird entsprechend den Fähigkeiten bzw. dem Lösungsverhalten der Schülerinnen und Schüler ein Aufgabenblock durch einen Algorithmus ausgewählt. Zur Auswahl stehen pro Testteil drei unterschiedlich schwierige Aufgabenblöcke (einfach, mittel, schwierig). Dieses Design stellt sicher, dass alle Schülerinnen und Schüler dieselbe Anzahl Aufgaben in der gleichen Zeit bearbeiten. Lediglich im Test Deutsch Sprache im Fokus umfasst der einfachste Aufgabenblock jeweils eine Aufgabe weniger (7 anstatt 8). Darüber hinaus bietet das Design die Möglichkeit, Aufgaben innerhalb eines Aufgabenblocks zu überspringen oder zu überarbeiten.

Rund ein Viertel der Tests wurde zufällig einem linearen Testdesign zugeordnet. Dieses Design wird ebenfalls in Abbildung 1 (unten) dargestellt. In diesem Design erfolgt keine Anpassung der Aufgaben an die Fähigkeiten der Schülerinnen und Schüler. Stattdessen umfasst jeder Testteil ein breites Spektrum an Aufgabenschwierigkeiten. Die Anzahl Aufgaben und die verfügbare Zeit pro Testteil sind im linearen und adaptiven Testdesign jeweils gleich.

Bei beiden Testdesigns wird anhand der Lösungen mittels der Item Response Theorie die Fähigkeit der Schülerinnen und Schüler berechnet und auf der Check-Skala von 0 bis 1200 Punkten berichtet. Diese Skala ermöglicht es, den Lernfortschritt von der 3. Klasse der Primarschule bis zur 3. Klasse der Sekundarstufe I (Zyklus 2 und Zyklus 3 des Lehrplans 21) abzubilden. Die Ergebnisse des Check P4 sind daher prinzipiell mit Ergebnissen von Checks auf anderen Klassenstufen vergleichbar. Zudem sind sie mit weiteren Durchführungen des Check P4 in kommenden Jahren vergleichbar.<sup>1</sup>

<sup>1</sup> Weiterführende Informationen sind unter [www.ibe-edu.ch/methoden](http://www.ibe-edu.ch/methoden) zu finden.

Abbildung 1: Visualisierung des adaptiven (oben) und linearen (unten) Testdesigns



### 1.3 Teilnehmende Schülerinnen und Schüler

Tabelle 2 zeigt die Anzahl der teilnehmenden Schülerinnen und Schüler, Klassen und Schulen. Insgesamt haben 1'177 Schülerinnen und Schüler am Check P4 teilgenommen. Davon haben einige einen oder mehrere Tests nicht bearbeitet. Gründe dafür können unter anderem Abwesenheit wegen Krankheit oder die fehlende Möglichkeit eines Nachteilsausgleichs (längere Bearbeitungszeit) sein (vgl. Abs. 5.1).

Die genaue Anzahl der Schülerinnen und Schüler, die im Mai 2024 eine 4. Klasse in einer gemeindlichen Schule des Kantons Zug besuchten, ist nicht bekannt. Das Amt für gemeindliche Schulen schätzt die Zahl auf rund 1'200. Aufgrund dieser Annahme beträgt die Ausschöpfung je nach Test zwischen 94.8% und 95.7%, was einer hohen Ausschöpfung entspricht und somit eine gute Repräsentativität der Daten gewährleistet. In Tabelle 2 und den folgenden Auswertungen sind 12 Schülerinnen und Schüler nicht berücksichtigt, weil sie eine Privatschule besuchen. Die Anzahl der teilnehmenden Klassen weicht im Fach Englisch um 1 ab, weil eine kleine DaZ-Klasse einzig die Tests in diesem Fach durchführte.



Die im Rahmen des Check P4 durchgeführte Online-Befragung der Klassenlehrpersonen und Schulleitungen (s. Abs. 5) gibt Hinweise dazu, weshalb einige Schülerinnen und Schüler nicht am Check P4 teilnahmen oder einzelne Tests nicht durchführten. In vielen Fällen waren vorübergehende oder überdauernde Lernzielanpassungen die Begründung für vollständige oder teilweise Nichtteilnahme am Check P4. In mehreren Fällen handelt es sich bei den nicht teilnehmenden Schülerinnen oder Schülern um integrierte Sonderschüler. Weitere vereinzelte Begründungen betrafen Krankheit, Beurlaubung oder mangelnde Deutschkenntnisse der Schülerinnen und Schüler sowie technische Probleme bei der Durchführung eines Tests.

Tabelle 2: Teilnahme und Ausschöpfungsquote nach Test

Fach	Kompetenzbereich	SuS*	Klassen	Schulen	Ausschöpfung **
Deutsch	Lesen	1139	87	30	94.9%
	Sprache im Fokus	1137	87	30	94.8%
Englisch	Hören	1159	88	30	96.6%
	Lesen	1161	88	30	96.8%
Mathematik		1148	87	30	95.7%
Total		1177	88	30	98.1%

\* Schülerinnen und Schüler

\*\* Basierend auf einer Grundgesamtheit von 1'200 Schülerinnen und Schülern

Für die Anmeldung zum Check P4 wurden die Schülerinnen und Schüler auf der Online-Plattform *Check Dein Wissen* registriert. Dabei erfassten die zuständigen Personen der Schulen auch das Geburtsdatum, das Geschlecht, die Erstsprache (Deutsch als Erstsprache, andere Erstsprache) sowie Fächer, in denen die Schülerinnen oder Schüler angepasste Lernziele haben. Die teilnehmenden Schülerinnen und Schüler sind im Mittel 10.7 Jahre alt ( $SD = 0.41$ , Stichdatum 31. Mai 2024) und der Anteil an Mädchen liegt bei 51%. Der Anteil Schülerinnen und Schüler mit Erstsprache Deutsch beträgt 61.1%. In Deutsch haben 86 (7.2%), in Englisch 39 (3.3%) und in Mathematik 65 (5.5%) Schülerinnen und Schüler angepasste Lernziele.

## 2 Ergebnisse in Deutsch, Englisch und Mathematik

Die Ergebnisse werden im Folgenden mehrheitlich pro Fach dargestellt. Dazu wurde pro Schülerin, pro Schüler der Mittelwert aus den Ergebnissen der Kompetenzbereiche gebildet. Das Ergebnis pro Fach wurde nur berechnet, wenn für eine Schülerin oder einen Schüler in allen Kompetenzbereichen ein Ergebnis vorliegt.

### 2.1 Mittelwert und Streuung der Ergebnisse

Die durchschnittlichen Ergebnisse aller Schülerinnen und Schüler sind in Tabelle 3 pro Fach aufgelistet. Neben dem Mittelwert ist die Standardabweichung aufgeführt, welche die Streuung auf der Ergebnisskala beziffert. Sie kann vereinfacht als durchschnittliche Abweichung vom Mittelwert interpretiert werden.

Die Ergebnisse in den verschiedenen Fächern sind nicht direkt miteinander vergleichbar, da die Tests verschiedener Fächer keine gemeinsame Metrik haben. Allerdings fällt auf, dass der Mittelwert in Englisch im Vergleich zu den Mittelwerten in Deutsch und Mathematik sehr viel höher ist. Zieht man Check-Ergebnisse aus anderen Kantonen zur Beurteilung bei, dann zeigt sich, dass dieser Unterschied zwischen den Fachergebnissen in anderen Kantonen so nicht vorkommt. Die Leistungen der Schülerinnen und Schüler des Kantons Zug im Fach Englisch sind deutlich höher als erwartet. Dies hat vermutlich mit dem vergleichsweise hohen Anteil an Kindern internationaler Fach- und Führungskräfte (Expats) zu tun, die zuhause oft Englisch sprechen.

Tabelle 3: Mittelwert (M) und Standardabweichung (SD) der Ergebnisse in den geprüften Fächern

Fach	<i>M</i>	<i>SD</i>	<i>N</i>
Deutsch	550	122	1126
Englisch	622	117	1148
Mathematik	536	141	1147

## 2.2 Ergebnisse nach individuellen Merkmalen

In Kombination mit den erfassten Merkmalen der Schülerinnen und Schüler geben die Ergebnisse des Check P4 Aufschluss über Unterschiede oder Ähnlichkeiten in den Leistungen verschiedener Gruppen. Abbildung 2 zeigt den Mittelwert (dunkelblauer Punkt), die mittleren 50% (blauer Balken) und die mittleren 90% (hellblauer Balken) der Ergebnisse pro Fach aufgeteilt nach Erstsprache, Geschlecht und Lernzielen. Die Kennwerte zum Gruppenvergleich sind in Tabelle 4 zu finden.

Da die Ergebnisse verschiedener Fächer nicht direkt vergleichbar sind, ist das Ausmass der Gruppenunterschiede zwischen verschiedenen Fächern jeweils im Verhältnis zur Streuung der Ergebnisse zu betrachten. Zu diesem Zweck wird die Effektstärke  $d$  verwendet. Sie ergibt sich aus der Differenz der Mittelwerte im Verhältnis zur Streuung der gemessenen Werte innerhalb der beiden Gruppen. Die Effektstärke  $d$  drückt also in Standardabweichungen aus, wie gross der Unterschied zwischen zwei Gruppen ist. Gemäss einer Konvention in der experimentellen psychologischen Forschung gilt eine Effektstärke von  $d = 0.2$  als kleiner Effekt,  $d = 0.5$  als mittlerer Effekt und  $d = 0.8$  als grosser Effekt.<sup>2</sup>

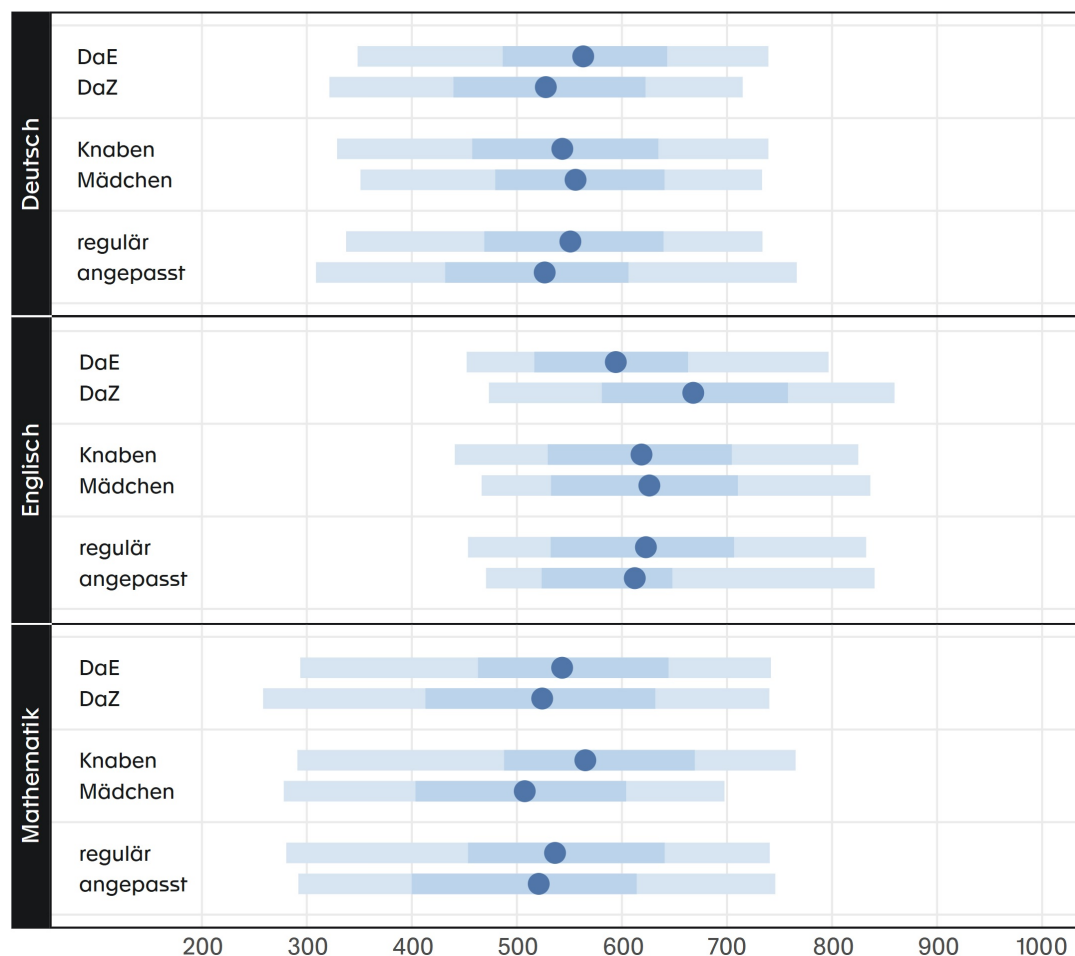
Zwischen Schülerinnen und Schülern mit Deutsch als Erstsprache (DaE) und jenen mit Deutsch als Zweitsprache (DaZ) finden sich bedeutsame Unterschiede in den Fächern Deutsch und Englisch. In Deutsch zeichnet sich ein Vorteil von 35 Punkten der Schülerinnen und Schüler mit Erstsprache Deutsch ab, was einem kleinen Effekt entspricht ( $d = 0.29$ ). In Englisch sind hingegen die Schülerinnen und Schüler mit Deutsch als Zweitsprache mit einem Vorsprung von 74 Punkten im Vorteil. Dies ist ein mittlerer Effekt ( $d = 0.66$ ). Dieses Ergebnis könnte darauf zurückzuführen sein, dass Englisch neben Deutsch eine häufige Erstsprache der Schülerinnen und Schüler im Kanton Zug ist.

Trotz dieser möglichen Erklärung ist das Ergebnis überraschend. Die Überprüfung der Grundkompetenzen (ÜGK) 2017 stellte für den Kanton Zug fest, dass Schülerinnen und Schüler im 8. Schuljahr, die zuhause nicht Deutsch sprechen, gleiche oder etwas weniger gute Ergebnisse in Englisch erzielen als jene mit Deutsch als Erstsprache.<sup>3</sup> Möglicherweise nimmt dieser Unterschied über die Schullaufbahn bis zum 8. Schuljahr bzw. bis zur 2. Klasse der Sekundarstufe I ab, sodass er dann nicht mehr oder nicht mehr in diesem Ausmass besteht. Zudem handelt es sich um einen Unterschied auf hohem Niveau. Die ÜGK unterscheidet lediglich zwischen Erreichen und Nichterreichen der Grundkompetenzen. Auch ein bedeutsamer Unterschied zwischen zwei Gruppen wird möglicherweise in der ÜGK nicht erkannt, wenn es sich um einen Unterschied deutlich oberhalb der Schwelle der Grundkompetenzen handelt.

<sup>2</sup> Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

<sup>3</sup> Konsortium ÜGK (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2017: Sprachen 8. Schuljahr*. EDK & SRED. <https://doi.org/10.18747/PHSG-coll3/id/385>

Abbildung 2: Ergebnisse im Check P4 pro Fach, separiert nach Geschlecht, Erstsprache Deutsch vs. andere Erstsprache sowie Lernzielen



Im Fach Mathematik ist der Unterschied zwischen Schülerinnen und Schülern mit Deutsch als Erstsprache und Deutsch als Zweitsprache mit 19 Punkten sehr gering ( $d = 0.13$ ). Die Richtung dieses Unterschieds ist konsistent mit Ergebnissen der ÜGK. Diese hat aber für das 11. Schuljahr nicht bloss einen sehr geringen, sondern einen markanten Vorsprung zugunsten der Schülerinnen und Schüler, die zuhause Deutsch sprechen, gefunden.<sup>4</sup> Es könnte sich also auch hier um einen Unterschied handeln, der sich über die Schullaufbahn hinweg verändert.

Zwischen den Ergebnissen von Mädchen und Knaben gibt es nur im Fach Mathematik bedeutsame Unterschiede. Die Mädchen erzielen etwas höhere Ergebnisse in Deutsch (13 Punkte,  $d = 0.10$ ) und in Englisch (7 Punkte,  $d = 0.06$ ). Das Ausmass dieser Differenzen ist aber sehr gering und daher nicht bedeutsam. In Mathematik ist der Mittelwert der Knaben hingegen 58 Punkte höher als jener der Mädchen, was einem kleinen bis mittleren Effekt entspricht ( $d = 0.42$ ). Der Leistungsvorsprung von Knaben gegenüber Mädchen in Mathematik wird in der Schweiz regelmässig festgestellt, beispielsweise in der PISA-Erhebung 2022<sup>5</sup> wie auch in der Überprüfung der Grundkompetenzen (ÜGK) 2016<sup>4</sup>. Diese Differenz ist im Kanton Zug auf der Primarstufe aber grösser als in anderen Kantonen (Bildungsraum Nordwestschweiz: ca. 20 Punkte Ende 5. Klasse<sup>6</sup>).

<sup>4</sup> Konsortium ÜGK (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2016: Mathematik 11. Schuljahr*. EDK & SRED. <https://doi.org/10.18747/PHSG-coll3/id/386>

<sup>5</sup> Erzinger, A. B., Pham, G., Prosperi, O., & Salvisberg, M. (2023). *PISA 2022: Die Schweiz im Fokus*. Universität Bern. <https://www.pisa-schweiz.ch/PISA2022-DieSchweizimFokus.pdf>

<sup>6</sup> IBE (2023). *Check P5 2023: Ergebnisbericht für den Bildungsraum Nordwestschweiz*. [https://www.bildungsraum-nw.ch/schwerpunkte/volksschule/ergebnisberichte\\_checks/copy3\\_of\\_ergebnisberichte2020/kantonsbericht-check-p5-2023.pdf](https://www.bildungsraum-nw.ch/schwerpunkte/volksschule/ergebnisberichte_checks/copy3_of_ergebnisberichte2020/kantonsbericht-check-p5-2023.pdf)

Tabelle 4: Ergebnisse pro Fach gruppiert nach individuellen Merkmalen

Fach	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
Erstsprache *	DaZ			DaE			
	Deutsch	528	127	421	563	117	705
	Englisch	668	121	440	594	104	708
	Mathematik	524	148	443	543	136	704
Geschlecht	Mädchen			Knaben			
	Deutsch	556	118	576	543	126	550
	Englisch	626	115	588	619	119	560
	Mathematik	507	134	586	565	142	561
Lernziele	regulär			angepasst			
	Deutsch	551	121	1075	526	140	51
	Englisch	623	116	1111	612	125	37
	Mathematik	536	141	1104	521	155	43

\* DaZ: Deutsch als Zweitsprache; DaE: Deutsch als Erstsprache

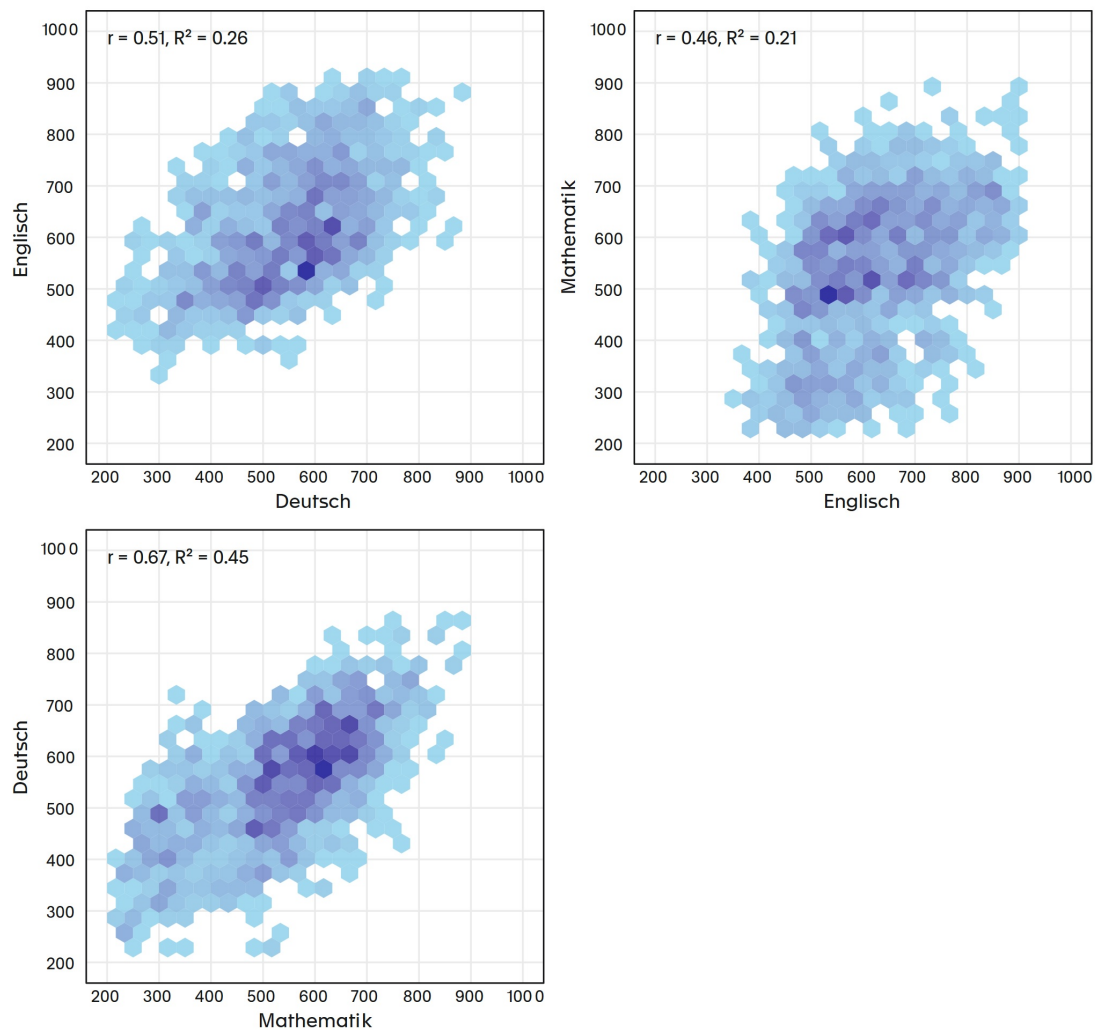
Für Schülerinnen und Schüler mit angepassten Lernzielen ist zu erwarten, dass sie im Mittel niedrigere Ergebnisse erzielen als Schülerinnen und Schüler mit regulären Lernzielen, da diese Massnahme bei Schülerinnen und Schülern mit besonderen Lernherausforderungen eingesetzt wird. Tatsächlich erzielen Schülerinnen und Schüler mit angepassten Lernzielen niedrigere Ergebnisse: in Deutsch -25 Punkte, in Englisch -11 Punkte und in Mathematik -15 Punkte. Die Effektstärke fällt aber in allen Fächern klein bis vernachlässigbar aus (Deutsch  $d = 0.20$ , Englisch  $d = 0.09$ , Mathematik  $d = 0.11$ ). Dies ist auch augenfällig im Vergleich zur Differenz zwischen Schülerinnen und Schülern mit Deutsch als Erstsprache und solchen mit Deutsch als Zweitsprache und wurde so nicht erwartet.

Die Vergleiche nach Lernzielanpassung sollten vorsichtig interpretiert werden, da die Lehrpersonen entscheiden, ob Schülerinnen und Schüler mit angepassten Lernzielen die Tests durchführen sollten. Wie die Online-Befragung der Lehrpersonen zeigt (siehe Abs. 5), nahmen einige Schülerinnen und Schüler aufgrund permanenter oder temporärer Lernzielanpassungen nicht am Check P4 oder an bestimmten Tests teil.

### 2.3 Zusammenhang zwischen den Fächern

Abbildung 3 visualisiert, wie die Ergebnisse in Deutsch, Englisch und Mathematik zusammenhängen. Die blauen Flächen zeigen, wie häufig verschiedene Kombinationen von Ergebnissen in jeweils zwei Fächern vorkommen. Je dunkler die Schattierung, desto mehr Schülerinnen und Schüler erzielen die jeweilige Kombination von Ergebnissen. Zudem sind die Korrelation und der Determinationskoeffizient aufgeführt. Die Korrelation ( $r$ ) beziffert den linearen Zusammenhang zwischen den Ergebnissen in zwei Fächern und reicht von -1 (vollkommener negativer Zusammenhang) bis +1 (vollkommener positiver Zusammenhang). Etwas einfacher zu interpretieren ist der Determinationskoeffizient ( $R^2$ ). Dieser beschreibt das Ausmass der gegenseitigen Abhängigkeit zweier Ergebnisse und kann wie eine Prozentzahl interpretiert werden. Beispielsweise bedeutet der Determinationskoeffizient von  $R^2 = 0.26$  für den Zusammenhang zwischen den Ergebnissen in Deutsch und Englisch, dass die Ergebnisse in Englisch zu 26% durch die Ergebnisse in Deutsch vorhersagbar sind – und umgekehrt.

Abbildung 3: Zusammenhang zwischen den Ergebnissen der drei Fächer



Der stärkste Zusammenhang ist zwischen den Ergebnissen in Deutsch und Mathematik zu finden ( $R^2 = 0.45$ ). Deutsch ist nicht nur die Unterrichtssprache, sondern auch die Sprache, in der die Mathematikaufgaben formuliert sind, was diesen vergleichsweise engen Zusammenhang erklären dürfte. Mathematische Probleme stellen sich meist in einem sprachlichen Kontext, weshalb gute Sprachkompetenzen für das Lösen von Mathematikaufgaben ein Vorteil sind.

Die Zusammenhänge zwischen Deutsch und Englisch sowie zwischen Englisch und Mathematik sind etwas geringer als der Zusammenhang zwischen Deutsch und Mathematik. Auch hier könnten mögliche Erklärungsansätze in generellen sprachlichen Fähigkeiten liegen. Die Zusammenhänge deuten zudem darauf hin, dass die Ergebnisse in einem schulischen Leistungstest nicht nur von den Fähigkeiten der Schülerinnen und Schüler abhängen, sondern auch von anderen Einflussfaktoren, die von dem erfassten Kompetenzbereich nicht vollkommen getrennt werden können.

## 2.4 Unterschiede zwischen Klassen

Abbildung 4 zeigt die Unterschiede zwischen den teilnehmenden Klassen. Die Klassenergebnisse entsprechen dem Mittelwert der Ergebnisse der Schülerinnen und Schüler mit regulären Lernzielen. Für Klassen mit weniger als 5 Ergebnissen wurde kein Mittelwert berechnet. Dadurch wurden 6 kleine Klassen (in Englisch: 7 Klassen) von der Auswertung ausgeschlossen. In den verbleibenden 81 Klassen sind jeweils Ergebnisse von 5 bis 23 Schülerinnen und Schülern verfügbar. Die durchschnittliche Grösse dieser Klassen beträgt 14.3 Schülerinnen und Schüler.

Abbildung 4: Mittelwert und mittlere 50% der Klassenergebnisse, aufsteigend sortiert

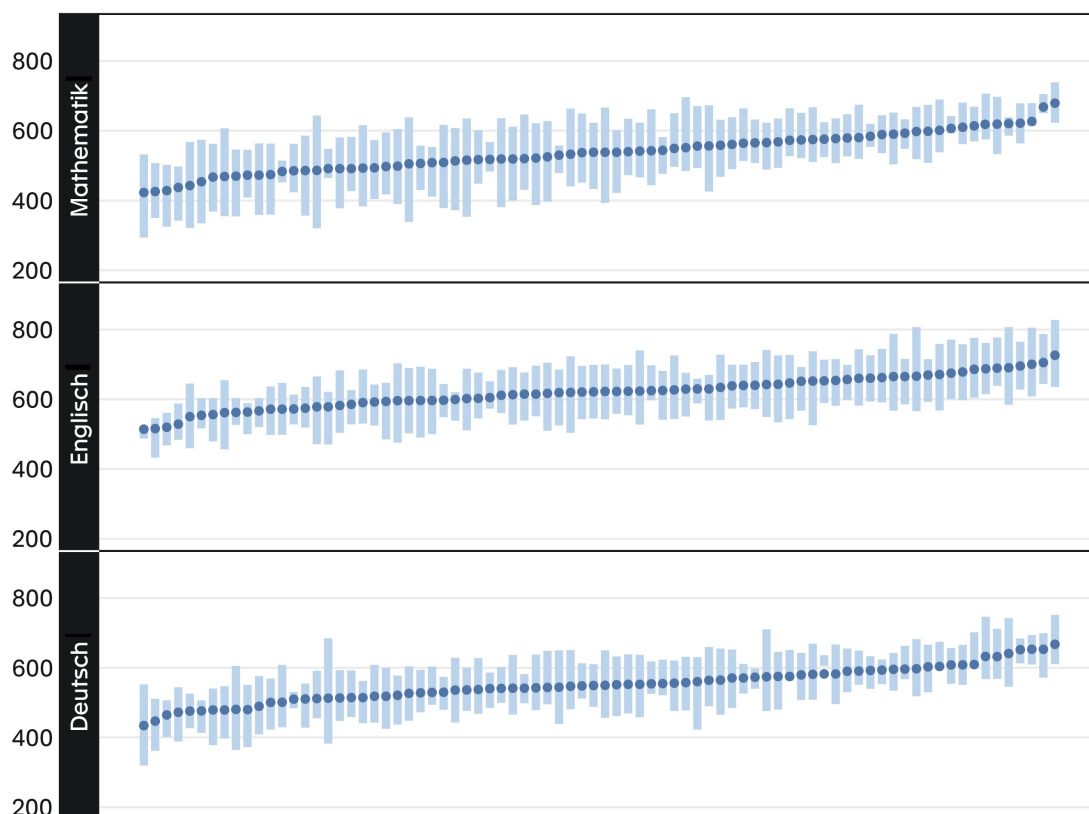


Abbildung 4 zeigt, dass die Unterschiede zwischen den Klassen sehr gross sind. Die Differenz zwischen dem niedrigsten und höchsten Klassenmittelwert beträgt für Deutsch 232 Punkte, für Englisch 212 Punkte und für Mathematik 256 Punkte. Allerdings konnten bei der Analyse der Klassenmittelwerte bedeutsame Lernvoraussetzungen wie die soziale Zusammensetzung der Klasse nicht berücksichtigt werden. Das heisst, es bleibt bislang offen, zu welchen Anteilen die Unterschiede durch Eigenschaften der Schülerinnen und Schüler oder durch Merkmale auf der Ebene der Klasse bzw. der Lehrpersonen oder der Schule zu erklären sind.

Die grossen Unterschiede zwischen den Klassen zeigen, weshalb eine unabhängige Standortbestimmung für die Ergänzung der klassenbezogenen Beurteilung anhand von Noten hilfreich ist. Die Informationen aus dem Check P4 führen zu einer objektiven Beurteilung, die sich nicht am Klassenkontext orientiert, und dadurch zu einer zusätzlichen Quelle für eine faire und ganzheitliche Beurteilung der Schülerinnen und Schüler.

## 2.5 Ergebnisse nach Kompetenzstufen

Zur besseren inhaltlichen Interpretation lassen sich alle Ergebnisse aus *Check Dein Wissen* einer Kompetenzstufe zuordnen. Die Kompetenzstufen reichen für den Check P4 von II bis IX und umschreiben, welche Aufgaben eine Schülerin oder ein Schüler in der Regel sicher lösen kann beziehungsweise welche Kompetenzen sie oder er beherrscht. Aufgaben der niedrigeren Kompetenzstufen kann die Schülerin oder der Schüler ebenfalls sicher lösen. Aufgaben der höheren Kompetenzstufen kann die Schülerin oder der Schüler noch nicht oder nicht sicher lösen.

Die im Lehrplan 21 definierten Grundansprüche und Orientierungspunkte können in den Kompetenzstufen verortet werden. Für die 4. Klasse ist der Orientierungspunkt im 2. Zyklus von Interesse, da dieser Orientierungspunkt definiert, welche Inhalte bis zum Ende der 4. Klasse behandelt werden sollen. Schülerinnen und Schüler, welche die bis zum Orientierungspunkt behandelten Themen beherrschen, erreichen im Check P4 die Kompetenzstufe V.

Im Gegensatz zu den definierten Grundansprüchen, die die Schülerinnen und Schüler bis zum Ende des 2. Zyklus erreichen sollten, handelt es sich beim Orientierungspunkt um einen Anhaltspunkt für die Unterrichtsplanung der Lehrpersonen. Es kann schon deshalb nicht erwartet werden, dass alle Schülerinnen und Schüler die Kompetenzstufe V erreichen, weil der Check P4 im Frühling stattfand und somit vor Ende der 4. Klasse. Vielmehr ist davon auszugehen, dass einzelne Themen, die bis zum Orientierungspunkt vorgesehen sind, zum Zeitpunkt des Check P4 noch nicht behandelt wurden.

Tabelle 5 zeigt, wie viele Schülerinnen und Schüler die Kompetenzstufen II bis IX in den geprüften Kompetenzbereichen erreichten. In Deutsch Lesen erreichen rund 68% der Schülerinnen und Schüler mindestens die Kompetenzstufe V. Die Kompetenzstufe V in Deutsch Lesen beinhaltet, dass Schülerinnen und Schüler verschiedene Arten von einfachen Texten verstehen, deren Themen und Wortschatz ihrem Weltwissen entsprechen (z. B. kurze Sachtexte, kindergerechte Reportagen, Ausschnitte aus Kinderbüchern). Schülerinnen und Schüler auf der Kompetenzstufe V beherrschen auch Kompetenzen der vorhergehenden Stufen und können beispielsweise einfachen diskontinuierlichen Texten (z. B. Einkaufsliste) explizite Informationen entnehmen (Kompetenzstufe III). Sie beherrschen aber noch nicht oder erst teilweise die Kompetenzen der nachfolgenden Kompetenzstufen, wie das Entnehmen von Hauptaussage und Detailinformationen aus anspruchsvolleren Texten (z. B. Kurzgeschichte oder Sachtext mit Diagrammen; Kompetenzstufe VII).

Die Kompetenzstufe V in Deutsch Sprache im Fokus erreichen oder übertreffen rund 66% der Schülerinnen und Schüler. Das bedeutet, dass sie typische Nomen, Verben und Adjektive bestimmen, verschiedene allgemeingültige Rechtschreibregeln anwenden und Kommas in Aufzählungen verwenden. Da sie auch die Kompetenzen der vorhergehenden Stufen beherrschen, können sie unter anderem regelmässige Adjektive steigern (Kompetenzstufe IV) und Wörter wie «Tropfen» oder «Krokodil» korrekt schreiben (Kompetenzstufe III). Sie können aber noch nicht oder nicht sicher die Zeitformen Präsens, Präteritum und Perfekt bestimmen (Kompetenzstufe VI) oder Kommas zwischen Verbgruppen setzen (Kompetenzstufe VII).

Tabelle 5: Verteilung der Kompetenzstufen in den geprüften Kompetenzbereichen

Fach	Kompetenzbereich	Kompetenzstufen (%)							
		II	III	IV	V*	VI	VII	VIII	IX
Deutsch	Lesen	5.7	10.3	16.3	30.6	26.2	8.2	2.7	
	Sprache im Fokus	2.9	9.1	21.9	25.9	28.3	9.7	2.2	
Englisch	Hören		2.0	14.2	28.9	30.3	18.5	5.1	1.1
	Lesen	0.1	1.4	16.7	27.5	24.0	18.8	8.6	2.9
Mathematik	Zahl und Variable	9.8	11.4	20.1	26.9	23.3	6.7	1.7	0.1
	Form und Raum	5.1	12.0	18.8	23.4	26.0	9.1	4.9	0.8
	Grössen,	10.6	9.1	13.1	30.4	22.8	11.1	2.3	0.5
	Funktionen, Daten und Zufall								

\* Die Kompetenzstufe V entspricht dem Erreichen des Orientierungspunktes im Zyklus 2 nach Lehrplan 21

In den Kompetenzbereichen des Fachs Englisch wird die Kompetenzstufe V von auffallend vielen Schülerinnen und Schülern erreicht oder überschritten: rund 84% in Englisch Hören und rund 82% in Englisch Lesen. In Englisch Hören zeichnet sich diese Kompetenzstufe dadurch aus, dass Schülerinnen und Schüler kurze Gespräche über vertraute Themen verstehen können. Zudem können sie grundlegende Informationen aus einfachen, kurzen Hörtexten entnehmen, wenn langsam und deutlich gesprochen wird. Die

Schülerinnen und Schüler können unter anderem auch einfache Fragen zur eigenen oder zu anderen Personen verstehen (Kompetenzstufe IV). Sie können aber noch nicht oder nur teilweise längeren Gesprächen und Hörtexten folgen bzw. vertraute Themen darin erkennen (Kompetenzstufe VIII). In Englisch Lesen beinhaltet die Kompetenzstufe V, dass Schülerinnen und Schüler einfache, kurze Texte und Informationsmaterialien (z. B. Comic, SMS, Rezept) verstehen bzw. grundlegende Aussagen daraus entnehmen. Sie verstehen aber noch nicht oder erst teilweise authentische Texte (z. B. Reportage) und persönliche Texte (z. B. Brief, Blog), selbst wenn diese klar strukturiert sind (Kompetenzstufe VIII).

Die Kompetenzstufe V wird in den Kompetenzbereichen der Mathematik unterschiedlich oft erreicht. In Zahl und Variable erreichen diese Stufe mit rund 59% vergleichsweise wenige Schülerinnen und Schüler. Diese Schülerinnen und Schüler beherrschen unter anderem das schriftliche Addieren und Subtrahieren, können für Addition und Subtraktion Rechenwege notieren und Zahlen bis 1 Million ordnen. Zudem können sie bereits Umkehroperationen nutzen, um Aufgaben in Teilschritten zu lösen (Kompetenzstufe IV), und Lösungen für Ungleichungen bestimmen (Kompetenzstufe III). Sie können aber beispielsweise noch nicht oder nicht sicher Zahlen mit vier Wertziffern addieren und subtrahieren (z. B.  $183 + 2401$ ; Kompetenzstufe VI) oder Dezimalzahlen ordnen, addieren und subtrahieren (Kompetenzstufe VII).

Im Kompetenzbereich Form und Raum erreichen rund 64% der Schülerinnen und Schüler die Kompetenzstufe V. Sie können unter anderem Symmetrieachsen erkennen und Spiegelbilder von Figuren skizzieren. Ebenfalls können sie bereits in einem Koordinatensystem Punkte beschreiben oder Figuren anhand von Koordinaten einzeichnen (Kompetenzstufe IV). Sie können allerdings noch nicht oder nicht sicher mehrere Symmetrieachsen in einer Figur erkennen (Kompetenzstufe VI) oder die Fläche von Rechtecken berechnen (Kompetenzstufe VII).

Im Kompetenzbereich Grössen, Funktionen, Daten und Zufall wird die Kompetenzstufe V von rund 67% der Schülerinnen und Schüler erreicht. Diese Schülerinnen und Schüler können Grössen in andere Masseinheiten umwandeln, damit einfache Rechnungen durchführen und einfache Textaufgaben mit Grössen lösen. Ebenfalls beherrschen sie bereits die Grundoperationen mit Geldbeträgen (Kompetenzstufe IV) und können lineare und nichtlineare Zahlenfolgen fortsetzen (Kompetenzstufe IV). Noch nicht oder nicht sicher können Schülerinnen und Schüler auf der Kompetenzstufe V proportionale Beziehungen anwenden (z. B. Fr. / kg oder l / km; Kompetenzstufe VI) oder Funktionswerte aus Graphen ablesen (Kompetenzstufe VII).

Zusammenfassend ist festzuhalten, dass im Deutsch und in zwei von drei Kompetenzbereichen der Mathematik etwa zwei Drittel der Schülerinnen und Schüler die Kompetenzen bis zum Orientierungspunkt des 2. Zyklus beherrschen oder übertreffen. In Englisch gilt dies sogar für über 80% der Schülerinnen und Schüler. Dies zeigt erneut, dass das Fach Englisch eine besondere Stärke der Schülerinnen und Schüler im Kanton Zug ist. Der Kompetenzbereich Zahl und Variable fällt gegenüber den anderen Kompetenzbereichen – gemessen am Orientierungspunkt – etwas ab. Das ist unerwartet, da bei der ÜGK 2016 die Ergebnisse dieses Kompetenzbereichs sehr nahe bei den Ergebnissen der anderen Kompetenzbereiche im Fach Mathematik lagen.<sup>7</sup>

### 3 Vergleich adaptives und lineares Testdesign

Im Auftrag des Kantons Zug wurden im Rahmen der Pilot-Durchführung des Check P4 zwei Testdesigns (adaptive Tests und lineare Tests) eingesetzt und die Ergebnisse sowie die methodische Qualität verglichen. Die beiden Testdesigns sind in Abschnitt 1.2 detailliert beschrieben. Rund drei Viertel der Tests waren adaptiv, ein zufällig ausgewählter Viertel der Tests wurde linear durchgeführt.

<sup>7</sup> Konsortium ÜGK (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2016: Mathematik 11. Schuljahr*. EDK & SRED. <https://doi.org/10.18747/PHSG-coll3/id/386>



Adaptive Tests haben gegenüber linearen Tests zwei Vorteile. Erstens bearbeiten die Schülerinnen und Schüler weniger Aufgaben, die sie überfordern oder unterfordern. Zweitens kann mit derselben Anzahl Aufgaben eine höhere Messgenauigkeit erreicht werden. Grund dafür ist, dass eine Aufgabe mehr Information zur Schätzung des Ergebnisses beiträgt, wenn ihre Schwierigkeit gut auf die Fähigkeit abgestimmt ist. Im Folgenden wird aufgezeigt, dass diese beiden Vorteile auch auf die adaptiven Tests des Check P4 zutreffen.

### 3.1 Methodische Qualität des adaptiven und des linearen Testdesigns

Zunächst wurde geprüft, ob der vorrangige Zweck des adaptiven Testdesigns – die Passung der Schwierigkeit der Aufgaben zur Fähigkeit der Schülerinnen und Schüler – erfüllt wird. Für einen Grossteil der eingesetzten Aufgaben ist die Schwierigkeit aufgrund früherer Tests bekannt. Bei einigen Aufgaben wurde die Aufgabenschwierigkeit anhand der Daten des Check P4 bestimmt. Zusammen mit den Ergebnissen der Schülerinnen und Schüler kann analysiert werden, welche der gezeigten Aufgaben für die Schülerinnen und Schüler besonders einfach oder schwierig waren. Dabei gilt eine Aufgabe als einfach, wenn der Schüler oder die Schülerin sie mit einer Wahrscheinlichkeit von mehr als 75% lösen konnte, und als schwierig, wenn die Lösungswahrscheinlichkeit unter 25% betrug. Aufgaben mit einer Lösungswahrscheinlichkeit von 25% bis 75% gelten als passend, da sie die Schülerin oder den Schüler weder über- noch unterfordern und viel Information zur Schätzung des Ergebnisses beitragen.

Abbildung 5 zeigt für jeden Test und nach Testdesign (adaptiv oder linear), wie oft eine schwierige (gelb), einfache (blau) oder passende (orange) Aufgabe gezeigt wurde. Für alle Tests ist klar erkennbar, dass durch das adaptive Testdesign häufiger passende Aufgaben angezeigt wurden als bei einem linearen Test. Am grössten ist der Vorteil des adaptiven Testdesigns in Deutsch Lesen und Deutsch Sprache im Fokus (SiF).

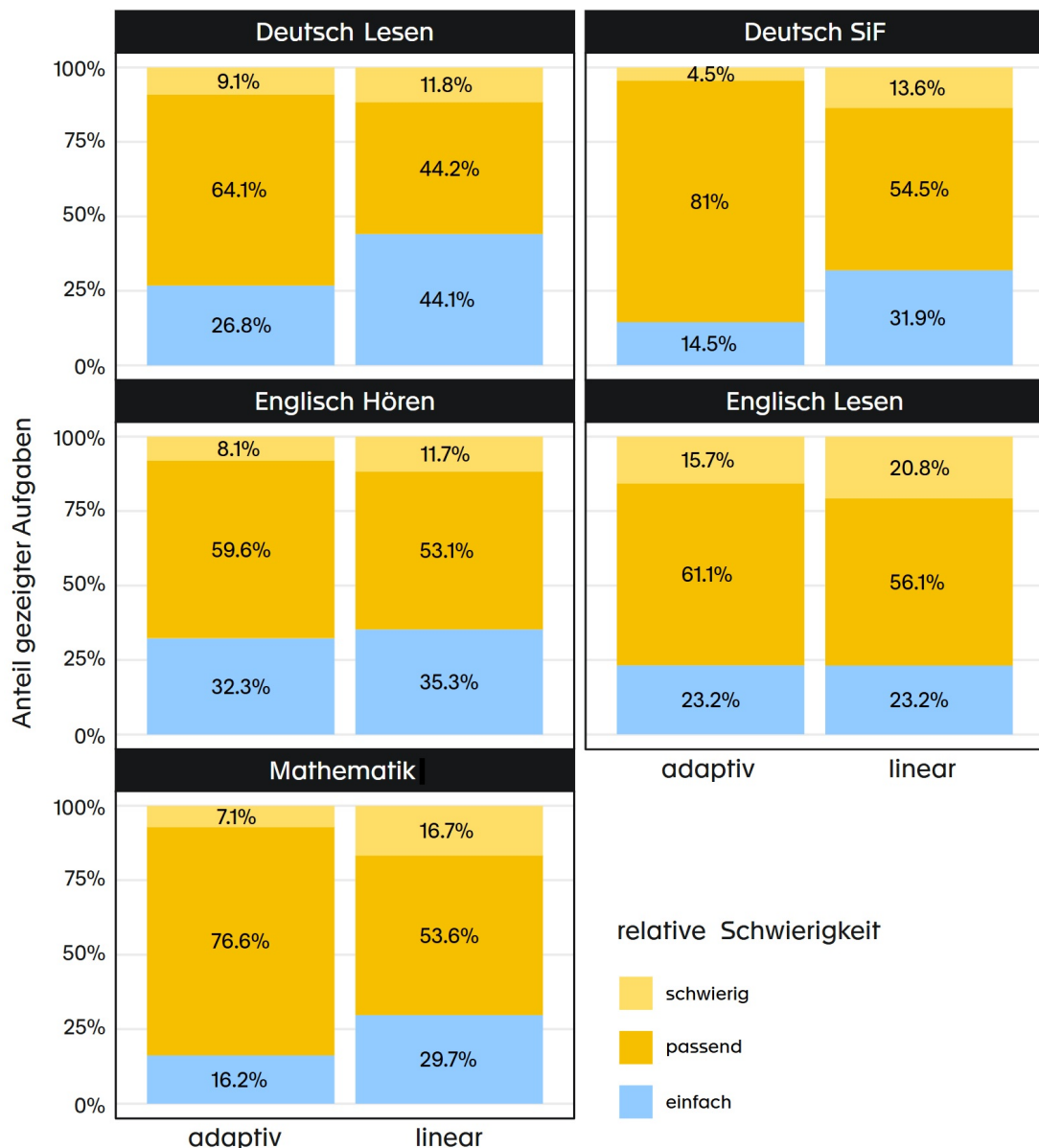
In Englisch ist der Vorteil des adaptiven Testdesigns am kleinsten. Eine mögliche Ursache für dieses Ergebnis ist, dass in den beiden Englisch-Tests – im Unterschied zu den Tests in Deutsch und Mathematik – Aufgaben enthalten waren, deren Schwierigkeit erst nach der Durchführung final bestimmt wurde. Dadurch konnten die Aufgaben weniger optimal auf die unterschiedlich schwierigen Aufgabenblöcke verteilt werden.

Bei den Kompetenzbereichen Lesen und Hören sind die Aufgaben zudem als Gruppen organisiert. Das bedeutet, dass jeweils mehrere Aufgaben pro Testteil zum selben Text oder zum selben Audio eingesetzt werden. Dadurch können die Aufgaben weniger flexibel aufgrund ihrer Schwierigkeit auf die unterschiedlich schwierigen Aufgabenblöcke des adaptiven Testdesigns verteilt werden. Gegen diese Interpretation spricht allerdings, dass dieselbe Einschränkung auch den Test in Deutsch Lesen betrifft.

Nach jeder Durchführung eines Checks werden das Antwortverhalten der Teilnehmenden und die Charakteristika der Aufgaben in allen Tests eingehend analysiert. Anhand dieser weiterführenden Analysen werden bis zur nächsten Durchführung bei Bedarf methodische Verbesserungen umgesetzt. Für die Tests in Englisch Hören und Englisch Lesen des Check P4 bedeutet dies, dass die Passung von Aufgabenschwierigkeit und Fähigkeit der Schülerinnen und Schüler im adaptiven Testdesign zukünftig besser ausgenutzt werden kann.

Des Weiteren ist zu erkennen, dass in allen Tests häufiger einfache als schwierige Aufgaben gezeigt werden. Grund dafür ist, dass beide Testdesigns mit einigen eher einfachen Aufgaben starten, um den Schülerinnen und Schülern den Einstieg in die Prüfungssituation zu erleichtern. Zudem wird beim Testen generell empfohlen, dass die Schülerinnen und Schüler bei einer passenden Aufgabe eine Lösungswahrscheinlichkeit von 60% haben sollten, um sie während des Tests zu motivieren.

Abbildung 5: Anteile der gezeigten Aufgaben, die für die Schülerinnen und Schüler schwierig, passend oder einfach waren, nach Testdesign



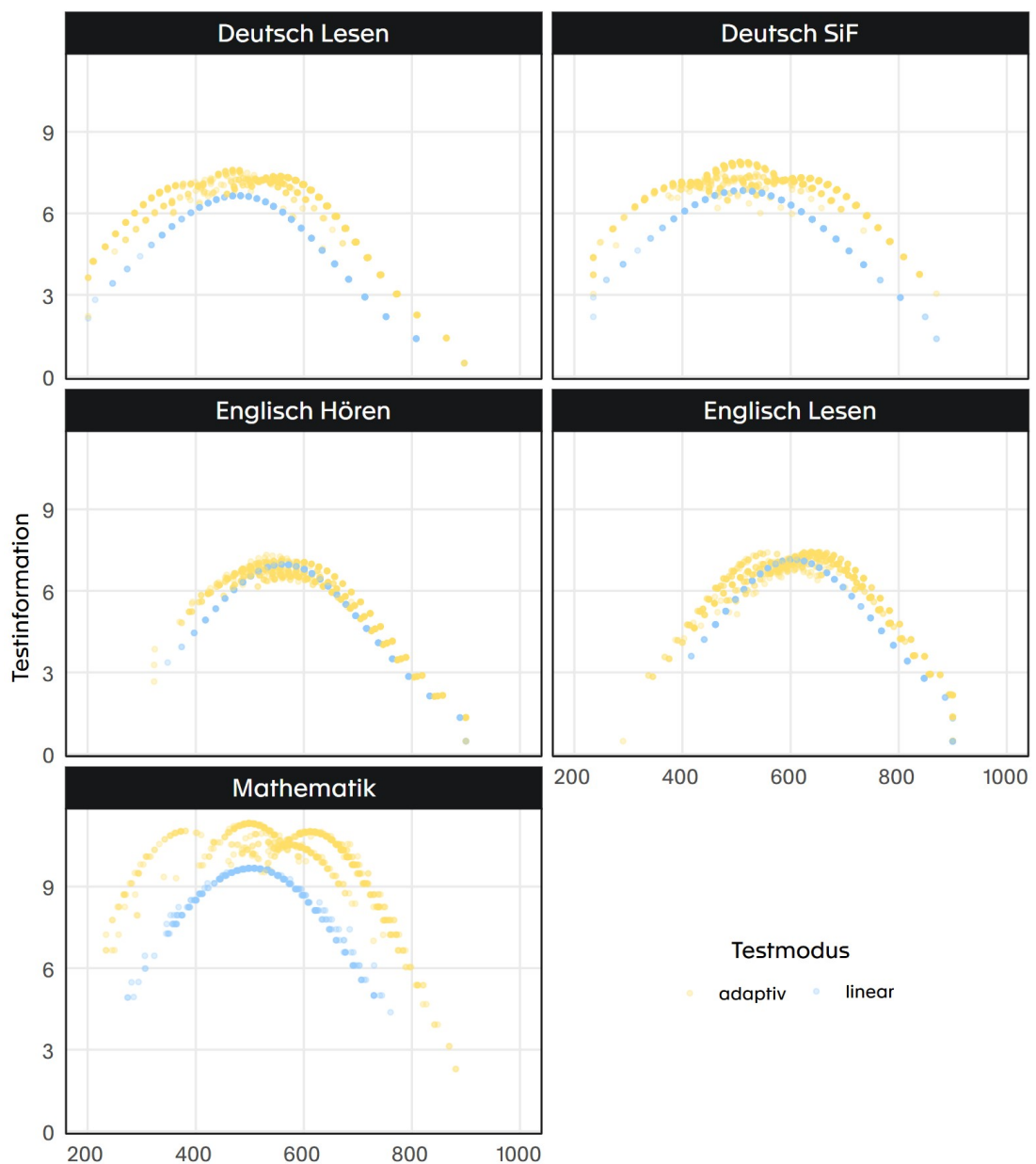
Neben der optimalen Herausforderung der Schülerinnen und Schüler bezweckt das adaptive – im Vergleich zum linearen – Testdesign eine bessere Messgenauigkeit über ein breiteres Fähigkeitsspektrum bei gleicher Testlänge. Um dies zu überprüfen, ist in Abbildung 6 die Testinformation (vertikal) gegen die Ergebnisse (horizontal) aller Schülerinnen und Schüler abgetragen. Die Testinformation ist ein Maß für die Messgenauigkeit bzw. für die aus dem Test erhältliche Information über die Fähigkeit einer geprüften Person. Sie kann verwendet werden, um die Präzision der Fähigkeitsmessung zwischen Tests zu vergleichen.

Die Testinformation variiert in beiden Testdesigns abhängig vom erzielten Ergebnis. Besonders hohe und niedrige Fähigkeiten können weniger genau gemessen werden als Fähigkeiten im mittleren Bereich. Im adaptiven Testdesign (gelb) kommt es zudem vor, dass Schülerinnen und Schüler dasselbe Ergebnis erzielen, sich aber die Testinformation für ihr Ergebnis unterscheidet. Dies veranschaulicht, dass die unterschiedlich

schwierigen Aufgabenblöcke in unterschiedlichen Bereichen optimal messen. Die Ergebnisse des linearen Testdesigns (blau) liegen hingegen auf einer Kurve, da alle Schülerinnen und Schüler dieselben Aufgaben lösten.<sup>8</sup>

Abbildung 6 zeigt, dass die Testinformation aller fünf Tests im adaptiven Testdesign überwiegend höher war als im linearen Testdesign. Der Unterschied ist an den Extremen des Fähigkeitsspektrums am augenfälligsten. Indem sich das adaptive Testdesign anpasst, wird der Bereich, in dem die Fähigkeit präzise gemessen wird, grösser. Auch mittlere Fähigkeiten werden im adaptiven Testdesign präziser gemessen, weil die Schülerinnen und Schüler weniger zu einfache oder zu schwierige Aufgaben lösen müssen.

Abbildung 6: Testinformationen für das adaptive (gelb) und lineare (blau) Testdesign



<sup>8</sup> Abweichungen von der Kurve der Testinformation – v. a. im Mathematik-Test zu sehen – kommen im linearen Testdesign vor, wenn Schülerinnen oder Schüler ganze Testteile auslassen und die Anzahl der bewerteten Aufgaben somit geringer ist.

Lediglich für Englisch Hören und Englisch Lesen sowie sehr vereinzelt in Deutsch Sprache im Fokus ist die Testinformation im adaptiven Testdesign teilweise etwas niedriger als die Testinformation des linearen Testdesigns. In manchen Fällen kann dies daran liegen, dass nicht der ganze Test beantwortet bzw. ein Testteil ausgelassen wurde oder dass Schülerinnen und Schüler beispielsweise durch eine Häufung von Flüchtigkeitsfehlern nicht optimal durch die unterschiedlich schwierigen Aufgabenblöcke geleitet wurden. Für die Tests in Englisch ist – wie weiter oben bereits beschrieben – zu beachten, dass die Zuordnung der Aufgaben zu den unterschiedlich schwierigen Aufgabenblöcken noch optimiert werden kann.

### 3.2 Ergebnisse im adaptiven und linearen Testdesign

Abbildung 7 zeigt die Mittelwerte und Standardabweichungen der beiden Gruppen nach Testdesign. Die Mittelwerte liegen jeweils sehr nahe beieinander. Am grössten ist die Differenz der Mittelwerte der beiden Testdesigns in Mathematik. Mit dem adaptiven Testdesign erzielten die Schülerinnen und Schüler im Mittel 17 Punkte mehr. Im Verhältnis zur Streuung der Ergebnisse ist dieser Unterschied aber als vernachlässigbar gering zu bewerten ( $d = 0.12$ ). Die Mittelwerte und Streuungen nach Test und Testdesign können auch in Tabelle 6 abgelesen werden.

Die Streuung der Ergebnisse um den Mittelwert ist im adaptiven Testdesign mit Ausnahme von Englisch Hören grösser. Dies entspricht den Erwartungen. Im adaptiven Testdesign werden die Aufgaben besser auf die Fähigkeiten der Schülerinnen und Schüler abgestimmt. Dadurch können die Fähigkeiten – insbesondere an den Extremen des Fähigkeitsspektrums – besser gemessen werden, was die Variabilität in den Testergebnissen vergrössern kann. Zudem kommt es auch weniger zu Decken- oder Bodeneffekten durch deutlich zu schwierige oder zu einfache Aufgaben (siehe Abbildung 5).

Abbildung 7: Vergleich der Ergebnisse zwischen adaptivem und linearem Testdesign

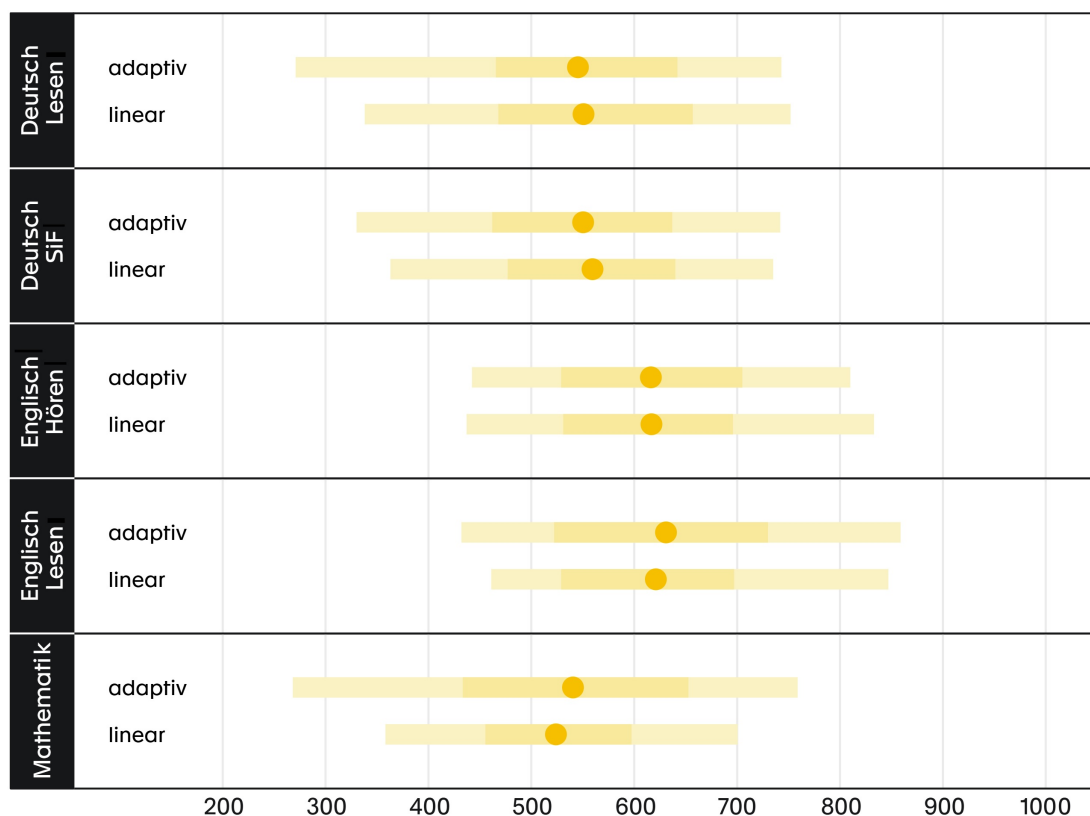


Tabelle 6: Mittelwert (M) und Standardabweichung (SD) der Ergebnisse, separiert nach Testdesign

Kompetenzbereich	Testdesign	Ergebnis		
		<i>M</i>	<i>SD</i>	<i>N</i>
Deutsch Lesen	linear	551	128	296
	adaptiv	545	142	843
Deutsch SiF *	linear	559	122	291
	adaptiv	550	127	846
Englisch Hören	linear	617	117	313
	adaptiv	616	116	846
Englisch Lesen	linear	621	122	293
	adaptiv	631	131	867
Mathematik	linear	524	104	330
	adaptiv	540	153	817

\* SiF: Sprache im Fokus

#### 4 Nutzung von Mindsteps und Ergebnisse im Check P4

Mindsteps ist eine Aufgabensammlung mit über 60'000 kompetenzorientierten Aufgaben, um Potenziale und Lernentwicklung von Schülerinnen und Schülern sichtbar zu machen. *Lernen sichtbar machen* ist die Quintessenz von John Hattie aus über 800 Metaanalysen und 50'000 Studien zu wirksamen Faktoren für den schulischen Lernerfolg.<sup>9</sup> Mindsteps wurde auf der Grundlage der Ergebnisse dieser Studien entwickelt.

Mit Mindsteps können Lehrpersonen regelmässig Standortbestimmungen in den Kompetenzbereichen der Fächer Deutsch, Englisch und Mathematik durchführen, das Erreichen der Kompetenzstufen des Lehrplans 21 überprüfen oder Lernkontrollen zu aktuellen Themen im Unterricht erstellen. Mindsteps ist eine Ergänzung zu den standardisierten und unabhängigen Checks und wird im Unterricht entsprechend den Check-Ergebnissen eingesetzt. Das Format der Aufgaben ist für Mindsteps und die Checks identisch, die Art der Auswahl und Durchführung unterscheidet sich hingegen. Während die Checks nach einheitlichen und standardisierten Regeln zusammengestellt und durchgeführt werden und dadurch zu einer unabhängigen Standortbestimmung führen, dient Mindsteps dem formativen Testen im Unterricht.

Schülerinnen und Schüler haben einen eigenen Zugang zu Mindsteps und können sowohl im Unterricht als auch zuhause Aufgabenserien lösen. Dabei wird zwischen zwei Bereichen unterschieden: Im Bereich *Mind* finden Schülerinnen und Schüler Aufgabenserien, die ihre Lehrpersonen für sie vorbereitet haben. Im Bereich *Steps* können die Schülerinnen und Schüler selbständig Aufgabenserien lösen, die dort zum Üben und Lernen bereitstehen. Die Lehrpersonen haben Einsicht in die Ergebnisse beider Bereiche.

Mindsteps kann im Unterricht oder zuhause genutzt werden. Sowohl der *Mind*- als auch der *Steps*-Bereich umfassen Aufgaben zu allen Kompetenzen, die auch im Check P4 relevant sind. Zudem können sich die Schülerinnen und Schüler über Mindsteps mit dem Check-P4-Aufgabenformat und der Testsituation am Computer vertraut machen. Aus diesen Gründen stellen sich die Fragen, wie oft Mindsteps als Vorbereitung auf den

<sup>9</sup> Hattie, John A. C. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating on achievement*. London & New York: Routledge. <https://doi.org/10.4324/9780203887332>

Check P4 verwendet wurde und inwiefern die Nutzung von Mindsteps mit den Ergebnissen im Check P4 in Zusammenhang steht.

#### 4.1 Nutzung von Mindsteps

Für alle Schülerinnen und Schüler der 4. Klassen im Kanton Zug wurde die Anzahl der bearbeiteten Aufgabenserien in den im Check P4 geprüften Kompetenzbereichen vom 1. August 2023 bis zum 31. Mai 2024 ausgezählt. Dabei wurden nur Aufgabenserien gezählt, in denen Schülerinnen und Schüler mindestens die Hälfte der Aufgaben beantwortet haben und sich dafür pro Aufgabe mindestens 5 Sekunden Zeit nahmen. Grund für dieses Filterkriterium ist, dass Schülerinnen und Schüler auf Mindsteps auch Aufgabenserien ansehen und «durchklicken» können, ohne die Aufgaben tatsächlich zu lösen.

Wie Tabelle 7 zeigt, hat die grosse Mehrheit der Schülerinnen und Schüler mindestens einmal Mindsteps in einem der drei Fächer genutzt. Rund 88% der Schülerinnen und Schüler haben Mindsteps im Fach Mathematik verwendet und 87% in Deutsch. In Englisch ist der Anteil mit 71% etwas niedriger. Die Anzahl der bearbeiteten Aufgabenserien variiert stark und reicht von 0 (alle drei Fächer) bis 68 in Mathematik, 57 in Deutsch und 42 in Englisch. Die meisten Schülerinnen und Schüler haben 3 bis 6 Aufgabenserien pro Fach bearbeitet, wie Tabelle 7 zeigt.

Tabelle 7: Nutzung von Mindsteps im Kanton Zug und Zusammenhang mit den Ergebnissen im Check P4

	<i>Mind</i> -Bereich			<i>Steps</i> -Bereich			Beide Bereiche		
	Nutz.	<i>M</i> Anz.	<i>r</i>	Nutz.	<i>M</i> Anz.	<i>r</i>	Nutz.	<i>M</i> Anz.	<i>r</i>
Deutsch	79.4%	2.4	0.17	54.3%	3.4	0.18	86.8%	5.8	0.24
Englisch	53.1%	1.2	0.02	41.0%	1.7	0.17	71.2%	2.9	0.14
Mathematik	72.3%	2.1	0.10	61.1%	3.8	0.15	87.9%	6.0	0.19

Nutz.: Anteil Schülerinnen und Schüler, der mindestens eine Aufgabenserie bearbeitete

*M*Anz.: Mittlere Anzahl Aufgabenserien pro Schülerin/Schüler

*r*: Korrelation zwischen Anzahl bearbeiteter Aufgabenserien (logarithmiert) und Ergebnis im Check P4

Betrachtet man die Bereiche *Mind* und *Steps* separat, so fällt auf, dass der *Mind*-Bereich in allen Fächern von einem höheren Anteil der Schülerinnen und Schüler genutzt wurde als der *Steps*-Bereich. Zugleich ist aber die mittlere Anzahl bearbeiteter Aufgabenserien im *Steps*-Bereich für alle drei Fächer höher. Dies bedeutet, dass ein Teil der Schülerinnen und Schüler den *Steps*-Bereich sehr intensiv nutzte, während die Aktivität im *Mind*-Bereich ausgeglichener war. Es bleibt offen, ob die Schülerinnen und Schüler aus eigenem Antrieb im *Steps*-Bereich aktiv waren oder ob sie durch die Lehrperson dazu angehalten wurden.

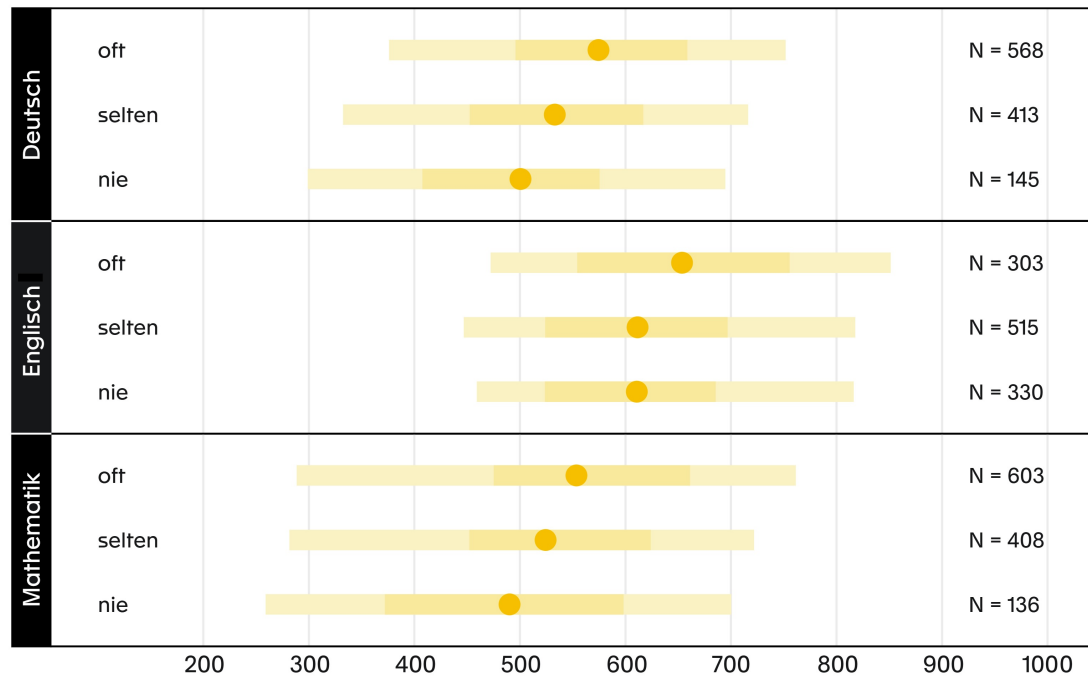
#### 4.2 Zusammenhang zwischen der Nutzung von Mindsteps und den Ergebnissen im Check P4

Tabelle 7 zeigt neben Kennwerten zur Nutzung von Mindsteps auch die Korrelation zwischen der Nutzung und den Ergebnissen im Check P4. Wie erwartet, ist die Korrelation für alle drei Fächer positiv. Das heisst, dass Schülerinnen und Schüler, die Mindsteps häufiger nutzten, etwas bessere Ergebnisse im Check P4 erzielten als jene, die selten oder nie Aufgabenserien auf Mindsteps lösten. Die Korrelationen sind mit  $r = 0.14$  bis  $r = 0.24$  eher klein, aber bedeutsam.

Um die Relevanz dieses Zusammenhangs abzuschätzen, sind in Abbildung 8 die Ergebnisse im Check P4 nach drei Nutzergruppen separiert: Schülerinnen und Schüler, die Mindsteps nie nutzten (keine Aufgabenserie bearbeitet), jene, die Mindsteps selten

nutzten (1 bis 3 Aufgabenserien), und jene, die Mindsteps oft nutzten (mehr als 3 Aufgabenserien). Für die Gruppierung wurden in jedem Fach wieder nur Aufgabenserien gezählt, die mindestens zur Hälfte und nicht übereilt beantwortet wurden.

Abbildung 8: Mittelwerte (orange), mittlere 50% (gelb) und mittlere 90% (hellgelb) der Ergebnisse im Check P4 nach Nutzung von Mindsteps



oft: mehr als 3 Aufgabenserien im jeweiligen Fach  
 selten: 1 – 3 Aufgabenserien im jeweiligen Fach  
 nie: keine Aufgabenserie im jeweiligen Fach

Für alle drei Fächer ist zu erkennen, dass Schülerinnen und Schüler, die Mindsteps oft nutzten, im Mittel bessere Ergebnisse im Check P4 erzielten als die anderen zwei Gruppen. Gegenüber jenen, die Mindsteps selten nutzten, beträgt dieser Vorsprung in Deutsch 41 Punkte ( $d = 0.35$ ), in Englisch 42 Punkte ( $d = 0.36$ ) und in Mathematik 29 Punkte ( $d = 0.21$ ). Im Verhältnis zur Streuung sind dies zwar kleine Effekte. Für Deutsch und Englisch sind die Effekte aber vergleichbar mit dem Vorsprung der Schülerinnen und Schüler mit Erstsprache Deutsch gegenüber jenen mit Deutsch als Zweitsprache im Fach Deutsch (s. Abs. 2.2, Abb. 2).

Gegenüber den Schülerinnen und Schülern, die Mindsteps nie nutzten, haben jene, die oft mit Mindsteps arbeiteten, in Deutsch und Mathematik einen noch grösseren Vorsprung von 74 Punkten ( $d = 0.62$ ) bzw. 63 Punkten ( $d = 0.45$ ). Im Verhältnis zur Streuung ist dies ein mittlerer Effekt für Deutsch, ähnlich dem Vorsprung der Schülerinnen und Schüler mit Deutsch als Zweitsprache in Englisch. Für Mathematik ist der Effekt klein bis mittel, ähnlich dem Vorsprung der Knaben gegenüber Mädchen in diesem Fach (s. Abs. 2.2, Abb. 2).

Einzig in Englisch unterscheiden sich die zwei Gruppen, welche Mindsteps selten und nie nutzten, nicht (Differenz: 1 Punkt,  $d = 0.01$ ). Dies deutet darauf hin, dass in Englisch eine grössere Anzahl Aufgabenserien als in den anderen zwei Fächern notwendig ist, um einen minimalen Effekt im Ergebnis des Check P4 zu erreichen. Grund dafür ist möglicherweise das generell höhere Niveau der Schülerinnen und Schüler in diesem Fach.

Insgesamt ist festzuhalten, dass Schülerinnen und Schüler, die während der 4. Klasse häufiger Aufgabenserien auf Mindsteps lösten, bessere Ergebnisse im Check P4 erzielten. Das Ausmass dieser Unterschiede ist vergleichbar mit Unterschieden zwischen relevanten individuellen Merkmalen wie dem Geschlecht oder der Erstsprache.

Bei der Interpretation dieser Effekte ist allerdings Vorsicht geboten: Ob der Zusammenhang zwischen der Nutzung von Mindsteps und den Ergebnissen im Check P4 auf einer direkten kausalen Beziehung beruht, bleibt ungewiss. Möglich ist nicht nur, dass das Lösen von Aufgabenserien die Leistungen im Check P4 positiv beeinflusste. Es könnte ebenso ein dritter Faktor eine Rolle spielen, welcher sowohl die Nutzung von Mindsteps als auch die Ergebnisse im Check P4 positiv beeinflusst. Beispielsweise könnte das Engagement bzw. die Einstellung der Lehrpersonen der Grund dafür sein, dass Schülerinnen und Schüler sowohl häufiger mit Mindsteps arbeiten als auch besser im Check P4 abschneiden.

## 5 Ergebnisse der Befragung zum Check P4

Die Klassenlehrpersonen der 88 Klassen und die Schulleitungen der 30 Schulen, die am Check P4 2024 beteiligt waren, wurden Ende Juni 2024 per E-Mail eingeladen, an einer anonymen Online-Befragung zum Check P4 teilzunehmen. 46 Personen haben die Umfrage beantwortet, davon 40 Lehrpersonen und 6 Schulleiterinnen und Schulleiter.

Im nächsten Abschnitt werden die Ergebnisse der Befragung zusammengefasst. Davon ausgenommen sind die Rückmeldungen dazu, ob und weswegen einzelne Schülerinnen und Schüler am Check P4 oder an einzelnen Tests nicht teilnahmen. Diese Ergebnisse wurden bereits in Abschnitt 1.3 beschrieben. Die vollständigen Ergebnisse aller geschlossenen Fragen sind in Abschnitt 5.2 zu finden.

### 5.1 Zusammenfassung der wichtigsten Rückmeldungen

Zu Beginn der Online-Befragung wurde die allgemeine Zufriedenheit mit dem Check P4 erfragt. Danach folgten spezifischere Fragen zu verschiedenen Aspekten, wobei einzelne Fragen nur an Lehrpersonen oder nur an Schulleitungen gerichtet wurden. Am Ende wurden die Teilnehmenden um weitere Rückmeldungen gebeten, die sie frei formulieren konnten.

Die Frage nach der Gesamtzufriedenheit («Wie zufrieden sind Sie insgesamt mit dem Check P4?») ergab einen Mittelwert von 4.8 auf einer Skala von 0 bis 10 (SD = 2.6), was eine eher niedrige Zufriedenheit anzeigt. In vergleichbaren Umfragen zu den Checks in anderen Kantonen erzielt dieselbe Frage einen Mittelwert von 7.7. Die eher niedrige Gesamtzufriedenheit steht teilweise im Widerspruch zu den eher positiven Rückmeldungen auf spezifischere Fragen.

Von den befragten Personen berichten 82.1%, dass sie sich gut oder sehr gut über den Check P4 informiert fühlten. Seitens der Lehrpersonen waren vier Arbeitsschritte im Rahmen des Check P4 notwendig (Schülerinnen und Schüler erfassen, Online-Tests planen, Online-Tests durchführen und Online-Tests beaufsichtigen). Davon wurden drei Schritte von 70% der befragten Lehrpersonen als einfach oder sehr einfach bewertet. Das Durchführen der Online-Tests beurteilten 52.5% der Lehrpersonen als einfach, 10% fanden dies schwierig und 37.5% weder schwierig noch einfach. Dieser Schritt wurde somit im Vergleich zu den anderen drei Schritten als anspruchsvoller wahrgenommen.

Die an den Schulen zur Verfügung stehende Infrastruktur für die Durchführung des Check P4 wurde mehrheitlich positiv beurteilt. Den Fragen nach der Verfügbarkeit und Eignung der Geräte für die Online-Tests sowie nach der Stabilität der Internetverbindung und nach der Funktionsfähigkeit des Safe Exam Browsers stimmten 71.7% bis 84.1% der Befragten eher oder ganz zu. Unzufriedenheit mit den Geräten betraf – gemäss den Antworten auf die offene Frage – die Bildschirme, welche nicht immer die ganze Aufgabe und alle Antwortoptionen anzeigen konnten, sowie motorische Herausforderungen beim Scrollen und bei Aufgaben mit Drag-and-Drop. Es bleibt unklar, ob sich dieses Problem auf einen bestimmten Gerätetyp (z. B. Tablets) bezieht.



Der Support des IBE wurde ebenfalls überwiegend positiv beurteilt. Lediglich 14 der befragten Personen hatten Kontakt zum Support. Davon stimmten 13 Personen zu, dass der Support gut erreichbar war und ihr Anliegen zeitnah bearbeitet wurde. 12 der 14 Personen stimmten zu, dass ihre Fragen zufriedenstellend beantwortet wurden.

Ähnlich positiv wurde die Ergebnismeldung zum Check P4 beurteilt. 92.8% der Befragten stimmten zu, dass die Ergebnismeldungen einfach zugänglich und grafisch ansprechend gestaltet sind. Des Weiteren stimmten 85.7% der Befragten zu, dass die Ergebnismeldungen übersichtlich gestaltet sind. Vereinzelt wurde in der offenen Frage kritisiert, dass die Ergebnismeldungen nicht Bezug auf den Orientierungspunkt des 2. Zyklus des Lehrplans 21 nehmen.

Schwieriger zu beurteilen war für die Lehrpersonen der Inhalt des Check P4 bzw. der Tests. 46.2% der befragten Lehrpersonen stimmten zu, dass die Aufgaben dem Niveau des Unterrichts entsprachen, und 38.5% stimmten zu, dass die Schwierigkeit der Aufgaben angemessen war. Des Weiteren stimmten 50% der Lehrpersonen zu, dass die Aufgaben für die Schülerinnen und Schüler verständlich formuliert waren.

Einige Lehrpersonen merkten allerdings am Ende der Befragung an, dass sie die Frage zur Passung des Testinhaltes nicht beantworten konnten: «Das kann ich leider nicht beurteilen, da ich keinerlei Einsicht in die Aufgabenstellungen hatte. Das würde mich aber sehr interessieren.» Dazu passt, dass diese Fragen von rund 30% der Lehrpersonen mit «weder noch» beantwortet wurden. Um die Fairness der Tests und die Vergleichbarkeit über mehrere Durchführungen hinweg zu gewährleisten, ist es wichtig, dass die Fragen vertraulich bleiben. Deshalb erhalten die Lehrpersonen keine Einsicht in die Aufgaben des Check P4.

Rund ein Viertel der befragten Lehrpersonen beurteilte den Inhalt als eher nicht zum Unterricht passend, als eher zu schwierig oder als nicht verständlich formuliert. Bei der offenen Frage wurde teilweise entsprechende Kritik geäußert. Sechs Lehrpersonen merkten an, dass der in den Tests geprüfte Stoff noch nicht behandelt worden sei und/oder die Themen nicht zum verwendeten Lehrmittel gepasst hätten. Vereinzelt wurde auch darauf hingewiesen, dass gewisse Themen erst nach der Durchführung des Check P4 behandelt würden: «Im April ist das Schuljahr noch nicht zu Ende, Geometrie z.B. kommt (in meiner Klasse) erst vor den Sommerferien an die Reihe. Direkte/indirekte Rede ebenfalls. Diese Aufgaben konnten die SuS nicht lösen.» Hierzu ist anzumerken, dass die Online-Tests – insbesondere in den schwierigeren Aufgabenblöcken – bewusst auch Aufgaben enthalten, die über die Anforderungen der vierten Klasse hinausgehen.

Hinweise auf mögliche Gründe für die Diskrepanz zwischen der niedrigen Gesamtzufriedenheit und der mehrheitlich positiven Beurteilung der erfragten Bereiche des Check P4 liefern die Antworten auf die offene Frage am Ende des Fragebogens. Hier wurden mehrere Themen angesprochen, die nicht explizit abgefragt wurden. Der Zeitpunkt der Durchführung und der Ergebnismeldung wurde von acht Personen kritisiert, da die Lehrpersonen gegen Ende des Schuljahres schon viele Prüfungen ansetzen müssen und zugleich die Ergebnisse relativ kurz vor dem Klassen- bzw. Lehrpersonenwechsel zurückgemeldet werden. Ebenfalls haben vier Personen bemängelt, dass der zeitliche Aufwand für die Vorbereitung und Durchführung der Tests zu gross sei.

Vereinzelt wurde auch grundsätzliche Kritik am computergestützten Testen und/oder an standardisierten Tests geäußert. Einige Befragte sind beispielsweise der Meinung, dass die Schülerinnen und Schüler bereits zu viel Zeit am Bildschirm verbringen oder dass ihnen die nötige Medienkompetenz fehlt, um in einem Online-Test gute Leistungen zu zeigen. Andere äussern die Sorge, dass der Check P4 bei den Schülerinnen und Schülern – insbesondere bei den schulisch schwächeren – grossen Stress auslöst oder dass die standardisierten Online-Tests zu wenig Spielraum für Massnahmen zum Nachteilsausgleich zulassen.

Insgesamt zeigt sich, dass die Lehrpersonen und Schulleitungen gut auf den Check P4 vorbereitet waren und die technischen Aspekte der Durchführung gut bewältigen konnten. Weniger positiv wurde die inhaltliche Abstimmung der Tests auf die Lehrmittel und auf die individuelle Unterrichtsplanung beurteilt. Die allgemeine Zufriedenheit wird dadurch eingeschränkt, dass viele Lehrpersonen mit dem Durchführungszeitpunkt des Check P4 unzufrieden sind und den Aufwand für die Durchführung als zu hoch beurteilen.

## 5.2 Vollständige Ergebnisse der geschlossenen Fragen

Nachfolgend sind alle Ergebnisse der Fragen mit geschlossenem Antwortformat aufgeführt.

Was beschreibt Ihre Rolle am besten?

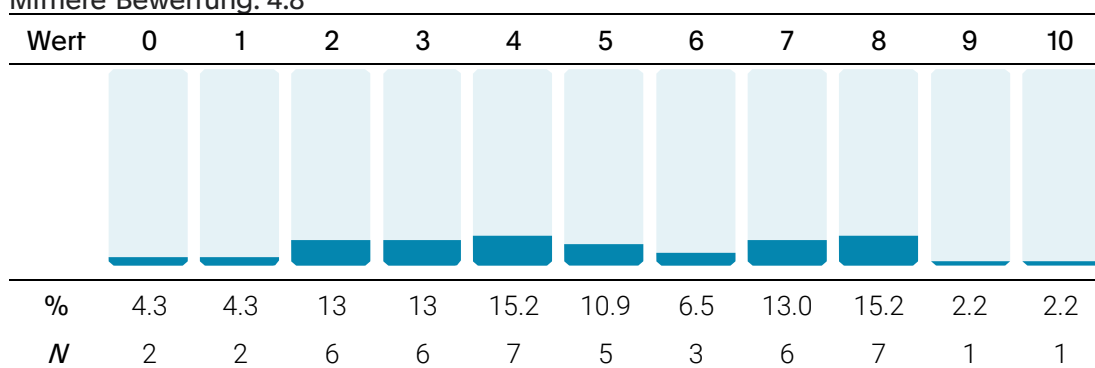
46 von 46 Personen haben diese Frage beantwortet.

Lehrerin/Lehrer	40 Antw.	87%
Schulleiterin/Schulleiter	6 Antw.	13%
Administratorin/Administrator	0 Antw.	0%
andere	0 Antw.	0%

Wie zufrieden sind Sie insgesamt mit dem Check P4?

46 von 46 Personen haben diese Frage beantwortet.

Mittlere Bewertung: 4.8



Wie haben Sie sich über den Check P4 informiert?

46 von 46 Personen haben diese Frage beantwortet (mehrere Antworten pro Person möglich).

Webseite "Check Dein Wissen"	32 Antw.	69.6%
schulinterne Information	27 Antw.	58.7%
kantonales Informationsschreiben	25 Antw.	54.3%
bei Kolleginnen/Kollegen	12 Antw.	26.1%
beim Support	7 Antw.	15.2%
andere	1 Antw.	2.2%

Wie gut haben Sie sich über den Check P4 informiert gefühlt?

46 von 46 Personen haben diese Frage beantwortet

sehr gut informiert	3 Antw.	6.5%
gut informiert	35 Antw.	76.1%
mässig informiert	6 Antw.	13.0%
eher wenig informiert	2 Antw.	4.3%
überhaupt nicht informiert	0 Antw.	0.0%

Wie einfach bzw. schwierig waren die folgenden Arbeitsschritte des Check P4?  
 (Frage wurde nur Lehrpersonen angezeigt.)  
 40 von 40 Lehrpersonen haben diese Frage beantwortet.

	sehr schwierig	schwierig	weder einfach noch schwierig	einfach	sehr einfach
Schülerinnen und Schüler erfassen	0.0%	0.0%	29.7%	37.8%	32.4%
Online-Tests planen	0.0%	17.5%	12.5%	40.0%	30.0%
Online-Tests mit den Schülerinnen und Schülern durchführen	0.0%	10.0%	37.5%	40.0%	12.5%
Online-Tests während der Durchführung beaufsichtigen	0.0%	7.5%	22.5%	37.5%	32.5%

Wie einfach bzw. schwierig waren die folgenden Arbeitsschritte des Check P4?  
 (Frage wurde nur Schulleitung und Administration angezeigt.)  
 5 von 6 Personen (Schulleitung/Administration) haben diese Frage beantwortet.

	sehr schwierig	schwierig	weder einfach noch schwierig	einfach	sehr einfach
Klassen erstellen	0%	0%	75%	25%	0%
Lehrpersonen erfassen	0%	0%	60%	20%	20%
Schülerinnen und Schüler erfassen	0%	0%	75%	25%	0%
Safe Exam Browser installieren	0%	75%	25%	0%	0%

**Wie beurteilen Sie den Inhalt des Check P4?**  
 (Frage wurde nur Lehrpersonen angezeigt.)  
 40 von 40 Lehrpersonen haben diese Frage beantwortet.

	stimme gar nicht zu	stimme nicht zu	weder noch	stimme eher zu	stimme ganz zu
Die Aufgaben entsprachen inhaltlich dem Niveau des Unterrichts.	2.6%	23.1%	28.2%	46.2%	0%
Die Aufgaben waren für die Schülerinnen und Schüler verständlich formuliert.	0%	22.5%	27.5%	42.5%	7.5%
Die Schwierigkeit der Aufgaben war angemessen.	1.6%	28.2%	30.8%	35.9%	2.6%

**Gibt es in Ihrer Klasse Schülerinnen und Schüler, die nicht am Check P4 teilgenommen haben?**

(Frage wurde nur Lehrpersonen angezeigt.)  
 40 von 40 Lehrpersonen haben diese Frage beantwortet.

Nein	27 Antw.	67.5%
Ja	13 Antw.	32.5%

**Aus welchen Gründen haben die Schülerinnen und Schüler nicht am Check P4 teilgenommen?**

(Frage nur gezeigt, wenn vorhergehende Frage mit „Ja“ beantwortet wurde.)  
 13 von 13 Lehrpersonen haben diese Frage beantwortet.

andere	8 Antw.	61.5%
vorübergehende Lernzielanpassungen (vLZA)	4 Antw.	30.8%
überdauernde Lernzielanpassungen (üLZA)	1 Antw.	7.7%
nicht realistischer Nachteilsausgleich (Zeitverlängerung)	0 Antw.	0.0%

**Gibt es in Ihrer Klasse Schülerinnen und Schüler, die einzelne Online-Tests des Check P4 nicht gelöst haben?**

(Frage wurde nur Lehrpersonen angezeigt.)  
 32 von 40 Lehrpersonen haben diese Frage beantwortet.

Nein	19 Antw.	59.4%
Ja	13 Antw.	40.6%

**Aus welchen Gründen haben die Schülerinnen und Schüler einzelne Online-Tests nicht gelöst?**

(Frage nur gezeigt, wenn vorhergehende Frage mit „Ja“ beantwortet wurde.)  
 13 von 13 Lehrpersonen haben diese Frage beantwortet.

andere	9 Antw.	69.2%
vorübergehende Lernzielanpassungen (vLZA)	3 Antw.	23.1%
nicht realistischer Nachteilsausgleich (Zeitverlängerung)	1 Antw.	7.7%
überdauernde Lernzielanpassungen (üLZA)	0 Antw.	0.0%

Wie beurteilen Sie die Infrastruktur an Ihrer Schule in Bezug auf die Durchführung des Check P4?

46 von 46 Personen haben diese Frage beantwortet.

	Stimme gar nicht zu	Stimme nicht zu	Weder noch	Stimme eher zu	Stimme ganz zu
Es waren genügend Geräte vorhanden.	4.4%	15.6%	2.2%	24.4%	53.3%
Die Geräte waren für die Durchführung geeignet.	4.3%	17.4%	6.5%	17.4%	54.3%
Die Internetverbindung war stabil.	0.0%	11.4%	4.5%	31.8%	52.3%
Der Safe Exam Browser funktionierte reibungslos.	2.2%	19.6%	10.9%	37.0%	30.4%

Hatten Sie Kontakt zum Support von Check Dein Wissen?

46 von 46 Personen haben diese Frage beantwortet.

Nein	32 Antw.	69.6%
Ja, telefonisch	11 Antw.	23.9%
Ja, per E-Mail	2 Antw.	4.3%
Ja, telefonisch und per E-Mail	1 Antw.	2.2%

Wie beurteilen Sie das Supportangebot von Check Dein Wissen?

(Frage nur gezeigt, wenn vorhergehende Frage mit „Ja“ beantwortet wurde.)

14 von 14 Personen haben diese Frage beantwortet.

	Stimme gar nicht zu	Stimme nicht zu	Weder noch	Stimme eher zu	Stimme ganz zu
Mein Anliegen wurde zeitnah bearbeitet.	0.0%	0.0%	7.1%	35.7%	57.1%
Meine Anfrage wurde zufriedenstellend beantwortet.	7.1%	0.0%	7.1%	21.4%	64.3%
Der Support war gut erreichbar.	0%	7.1%%	0%	35.7%	57.1%

Wie beurteilen Sie die Ergebnismeldung des Check P4?  
 45 von 46 Personen haben diese Frage beantwortet.

	Stimme gar nicht zu	Stimme nicht zu	Weder noch	Stimme eher zu	Stimme ganz zu
Die Ergebnisse sind einfach zugänglich.	0.0%	0.0%	7.1%	35.7%	57.1%
Die Ergebnismeldungen sind über- sichtlich gestaltet.	7.1%	0.0%	7.1%	21.4%	64.3%
Die Ergebnismeldungen sind grafisch ansprechend gestaltet.	0%	7.1%%	0%	35.7%	57.1%