



## **Check S2: Informationen zur Durchführung und Auswertung 2017**

---

### **Durchführung**

#### **Zeitfenster**

Die reguläre Durchführung des Checks S2 fand zwischen Montag, 20. Februar und Freitag, 31. März 2017 statt. Weil es in der Woche vom 13. bis zum 17. März insbesondere in den Kantonen Basel-Landschaft und Basel-Stadt zu technischen Problemen kam, wurde das Zeitfenster in diesen beiden Kantonen um eine Woche bis Freitag, 7. April verlängert.

Für Schülerinnen und Schüler, die während des regulären Zeitfensters nicht am Check S2 teilnehmen konnten, besteht die Möglichkeit, einzelne Tests zwischen dem 29. Mai und dem 9. Juni nachzuholen. Die Ergebnisse dieser Schülerinnen und Schüler werden am 16. Juni zur Verfügung stehen.

#### **Anzahl durchgeführter Tests**

Während des gesamten Zeitfensters haben 12'990 Schülerinnen und Schüler aus den Kantonen Aargau, Basel-Landschaft, Basel-Stadt und Solothurn am Check S2 teilgenommen.

Die Schülerinnen und Schüler haben in den Fächern Deutsch, Englisch, Französisch, Mathematik sowie Natur und Technik insgesamt 87'336 Online-Tests gelöst.

Für die Beurteilung der Schreibkompetenzen wurden Papier-Tests eingesetzt. Es wurden 12'892 Texte in Deutsch und 12'710 Texte in Englisch verfasst und beurteilt.

Für die Beurteilung der Sprechkompetenzen in Englisch und Französisch wurde ein standardisiertes Verfahren eingesetzt. In Englisch wurden die Sprechkompetenzen von 446, in Französisch von 237 Schülerinnen und Schülern beurteilt. Die Beurteilung wurde von den Lehrpersonen vorgenommen.

#### **Technische Informationen**

Bei der Durchführung der Online-Tests kam es in der Woche vom 13. bis zum 17. März in den Kantonen Basel-Landschaft und Basel-Stadt zu schwerwiegenden Performanceproblemen, von denen je nach Wochentag 5 bis 10 Prozent der Schülerinnen und Schüler betroffen waren. Der Grund für die Performanceprobleme lag darin, dass die Verteilung der Schulen auf die Server nicht wie vorgesehen funktionierte, weil in beiden Kantonen jeweils alle Schulen über eine einzige IP-Adresse ins Netz gelangten. Aufgrund der gleichen IP-Adresse wurden alle Schulen auf den gleichen Server geleitet, statt auf mehrere Server verteilt. Mit einer angepassten Konfiguration im Betrieb konnte das Problem gelöst werden. Abgesehen von diesem schwerwiegenden Problem, das insbesondere die Kantone Basel-Landschaft und Basel-Stadt betraf, verlief die Durchführung der Online-Tests ohne grössere technische Zwischenfälle. In der Regel lag die Quote für nicht korrekt abgeschlossene Tests zwischen 0 und 0.5 Prozent.

### **Auswertung**

#### **Analysen zu den technischen Problemen**

Aufgrund der technischen Probleme in der Woche vom 13. bis zum 17. März wurden insgesamt 2004 Online-Tests wiederholt. Um zu überprüfen, ob sich die technischen Probleme auch nach der Wiederholung der irregulär verlaufenen Online-Tests in den Ergebnissen zeigen, wurden zwei Auswertungen durchgeführt. Erstens wurde überprüft, ob die Aufgaben in der problematischen Woche vom 13. bis 17. März für die Schülerinnen und Schüler schwieriger zu lösen waren als in den übrigen Wochen. Die Auswertung zeigt, dass die Schwierigkeit der Aufgaben zwischen den einzelnen Wochen kaum variierte und dass die problematische Woche vom 13. bis 17. März keine auffällige Ausnahme darstellte. Zweitens wurde überprüft, ob sich die Testergebnisse der Schülerinnen und Schüler zwischen der problematischen Woche vom 13. bis 17. März und den regulären Wochen unterscheiden. Das ist zwar in manchen Kompetenzbereichen der Fall. Die Unterschiede sind aber gering und bewegen sich innerhalb



der zufälligen Schwankungsbreite zwischen allen sechs Wochen. Aufgrund der beiden Analysen zu den technischen Problemen wurden die Ergebnisse der Online-Tests der Woche vom 13. bis zum 17. März von Schülerinnen und Schülern der Kantone Basel-Landschaft und Basel-Stadt gleich wie die übrigen Ergebnisse behandelt.

### **Berechnung der Testergebnisse der Online-Tests**

Die Ergebnisse aller Schülerinnen und Schüler wurden vor der Berechnung in zweifacher Weise kontrolliert.

Erstens wurde überprüft, ob technische Probleme die Funktionalität einer Aufgabe beeinträchtigten. Für die Berechnung der Testergebnisse wurden folglich Aufgaben, deren teststatistischen Gütekriterien ungenügend waren, ausgeschlossen. Das bedeutet, dass einzelne Aufgaben, die für einen Grossteil der Schülerinnen und Schüler entweder unklar waren oder aufgrund von technischen Problemen nicht wie geplant gelöst werden konnten, von der Auswertung ausgeschlossen wurden. Diese Aufgaben zu werten hätte keinen zusätzlichen Gewinn für die Zuverlässigkeit der Ergebnisse gehabt und zwar unabhängig davon, ob die Aufgabe richtig oder falsch gelöst worden ist.

Zweitens wurde überprüft, wie viele Aufgaben von den Schülerinnen und Schülern tatsächlich bearbeitet wurden. Einige Schülerinnen und Schüler haben einzelne Aufgaben nicht bearbeitet, sei dies aufgrund von technischen Problemen oder, weil sie einzelne Aufgaben aus Zeitgründen oder absichtlich nicht bearbeiteten. Wurden alle Aufgaben eines Testteils nicht bearbeitet, dann wurde der entsprechende Testteil für die Berechnung der Testergebnisse nicht einbezogen. Wurde ein Testteil korrekt abgeschlossen, dann wurden nicht bearbeitete Aufgaben als falsch gewertet. Ein Testergebnis wurde allerdings nur dann berechnet, wenn eine Schülerin oder ein Schüler mindestens zwei Testteile eines Online-Tests bearbeitet und mindestens 20 Prozent aller Aufgaben gelöst hat. Bei diesen Testergebnissen muss berücksichtigt werden, dass die Zuverlässigkeit eines Testergebnisses mit der Abnahme der Anzahl gelöster Aufgaben sinkt. Das zeigt sich in der Ergebnismeldung bei der grafischen Umsetzung der Punktzahl und dem Vertrauensintervall, in dem das Ergebnis mit hinreichender Sicherheit liegt. Das Vertrauensintervall ist bei der grafischen Abbildung der Ergebnisse der Schülerinnen und Schüler in Form des blauen Balkens angeben.

Die Gesamtwerte in Deutsch und Mathematik wurden aus den Ergebnissen in den einzelnen Kompetenzbereichen berechnet. Die Gesamtwerte in Englisch und Französisch wurden aus den Ergebnissen in den einzelnen Kompetenzbereichen, aber ohne Ergebnis im Sprechen, berechnet. Die Gesamtwerte wurden nur dann berechnet, wenn es bei den jeweiligen Kompetenzbereichen allesamt gültige Testergebnisse gab, die miteinander verrechnet werden konnten.

### **Berechnung der Testergebnisse im Schreiben**

Die Texte im Deutsch Schreiben und Englisch Schreiben wurden von einem Team von Germanistinnen und Germanisten bzw. Anglistinnen und Anglisten korrigiert. Um die Texte der Schülerinnen und Schüler bewerten zu können, wurde ein standardisiertes Beurteilungsraster eingesetzt. Damit alle Personen die Beurteilungskriterien über die gesamte Korrekturzeit gleich anwenden, wurden anschliessend an eine Schulungsphase täglich fünf bis zehn zufällig ausgewählte Texte von allen Personen beurteilt und die Beurteilung miteinander verglichen. Mit diesem Vorgehen konnte erreicht werden, dass ein einheitlicher Beurteilungsmassstab erreicht werden konnte. Die verbleibenden Unterschiede in der Korrekturstrengung der beurteilenden Personen wurden bei der Berechnung der Ergebnisse berücksichtigt (Multi-facetten-Modell). Entsprechend diesem Verfahren können beim Vergleich der Bewertungsraster je nach Korrektur-Person leicht unterschiedliche Punktzahlen auf der Check-Skala resultieren.

### **Ergebnisse von Schülerinnen und Schülern mit individuellen Lernzielen**

Ein Teil der Schülerinnen und Schüler, die am Check S2 teilnahmen, haben in einem oder in mehreren Fächern individuelle Lernziele. Die Ergebnisse dieser Schülerinnen und Schüler wurden bei der Berechnung der Klassen- und Schulmittelwerte nicht berücksichtigt.

Zürich, 4. Mai 2017