



Check S3: Informationen zur Durchführung und Auswertung 2018

Durchführung

Zeitfenster

Die reguläre Durchführung des Checks S3 fand zwischen Montag, 16. April und Freitag, 25. Mai 2018 statt. Für Schülerinnen und Schüler, die während des regulären Zeitfensters nicht am Check S3 teilnehmen konnten, bestand die Möglichkeit, einzelne Tests zwischen dem 28. Mai und dem 1. Juni 2018 nachzuholen.

Anzahl durchgeführter Tests

Während des gesamten Zeitfensters haben 10'988 Schülerinnen und Schüler aus den Kantonen Aargau, Basel-Landschaft und Solothurn am Check S3 teilgenommen.

Die Schülerinnen und Schüler haben in den Fächern Deutsch, Englisch, Französisch, Mathematik sowie Natur und Technik insgesamt 77'604 Online-Tests gelöst.

Für die Beurteilung der Schreibkompetenzen wurden Papier-Tests eingesetzt. Es wurden 10'887 Texte in Deutsch und 9'136 Texte in Französisch verfasst und beurteilt.

Für die Beurteilung der Sprechkompetenzen in Englisch und Französisch wurde ein standardisiertes Verfahren eingesetzt. In Englisch wurden die Sprechkompetenzen von 124, in Französisch von 91 Schülerinnen und Schülern beurteilt. Die Beurteilung wurde von den Lehrpersonen vorgenommen.

Die Durchführung der Online-Tests verlief ohne technische Zwischenfälle. In der Regel lag die Quote für nicht korrekt abgeschlossenen Tests zwischen 0 und 0.5 Prozent.

Auswertung

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 bewerten 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 ausgeschlossen. 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 hatte. Es muss berücksichtigt werden, dass die Zuverlässigkeit eines Testergebnisses mit der Abnahme der Anzahl gelöster



Aufgaben sinkt. Die Zuverlässigkeit eines Testergebnisses wird in der Ergebnisrückmeldung anhand des Vertrauensintervalls angezeigt. Das Vertrauensintervall ist jener Bereich, in dem die Punktzahl mit hinreichender Sicherheit liegt. Das Vertrauensintervall ist bei der grafischen Abbildung der Ergebnisse der Schülerinnen und Schüler in Form des blauen Balkens angegeben.

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.

Berechnung der Testergebnisse im Schreiben

Die Texte im Deutsch Schreiben und Französisch Schreiben wurden von einem Team von Germanistinnen und Germanisten bzw. Romanistinnen und Romanisten beurteilt. Für die Beurteilung der Texte wurde ein standardisiertes Beurteilungsraster eingesetzt. Damit alle Personen die Beurteilungskriterien über die gesamte Beurteilungszeit 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 ein einheitlicher Beurteilungsmassstab erreicht werden. Die verbleibenden Unterschiede in der Korrektur-Strengung der beurteilenden Personen wurden bei der Berechnung der Ergebnisse berücksichtigt (Multifacetten-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 S3 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, 12. Juni 2018