Universität Zürich[UZH]

Institut für Bildungsevaluation
Assoziiertes Institut der Universität Zürich

Evaluation der Aufnahmeprüfungen in die Berufsmaturität

Bericht zuhanden des Departements Bildung, Kultur und Sport des Kantons Aargau

Nina König, Nicole Bayer, Martin J. Tomasik und Urs Moser

Zürich, 23. Juli 2020

Anschrift

Institut für Bildungsevaluation
Assoziiertes Institut der Universität Zürich
Wilfriedstrasse 15
8032 Zürich

Tel.: 043 268 39 60
Fax: 043 268 39 67
www.ibe.uzh.ch

Nina.Koenig@ibe.uzh.ch, Urs.Moser@ibe.uzh.ch

Inhalt

1 Ausgangslage 7

2 Ausbildungswege und Aufnahmeprüfungen 8

2.1 Berufsfachschule mit Berufsmaturität während der Lehre (BM I) 8

2.2 Berufsfachschule mit Berufsmaturität nach der Lehre (BM II) – Ausrichtung Wirtschaft und Dienstleistungen: Typ Wirtschaft 8

2.3 Berufsfachschule mit Berufsmaturität nach der Lehre (BM II) – Alle Ausrichtungen mit Ausnahme von Wirtschaft und Dienstleistungen: Typ Wirtschaft 9

3 Ziele und Fragestellungen 9

4 Datengrundlage 11

5 Methodisches Vorgehen 12

5.1 Überblick zu Fragestellung, methodischem Vorgehen und Ausbildungsweg 12

5.2 Statistische Methoden, Begriffe und Kennwerte 14

6 Ergebnisse 19

6.1 Prüfungsinhalte und Vorkenntnisse 19

6.2 Prüfungsinhalte und Lehrplaninhalte 29

6.3 Prüfungsteile und Prüfungsentscheid 45

6.4 Prüfungsteile und Beurteilungsmassstab 54

6.5 Gütekriterien der Aufnahmeprüfungen 64

6.6 Gütekriterien der Prüfungsaufgaben 73

7 Literaturverzeichnis 82

8 Anhang A 83

9 Anhang B 86

Management Summary

Vorgehen

Mit der Evaluation der Aufnahmeprüfungen in die Berufsmaturität wurden zwei Ziele verfolgt. Zum einen wurde überprüft, wie gut die Inhalte der Aufnahmeprüfungen mit den Vorkenntnissen der Kandidatinnen und Kandidaten übereinstimmen. Zum anderen wurde die Qualität der Aufnahmeprüfungen aus einer psychometrischen Perspektive beurteilt. Gegenstand der Evaluation sind folglich die Inhalte und die Form der Aufnahmeprüfungen in die Maturitätslehrgänge der Berufsfachschulen, nicht aber die Aufnahmeprüfungen für eine Fachmittelschule oder eine Informatikmittelschule.

Zur *Beurteilung der Prüfungsinhalte* wurden einerseits die Vorkenntnisse der Kandidatinnen und Kandidaten durch einen Vergleich der Prüfungsergebnisse mit den Ergebnissen im Check S3 beurteilt. Der Check S3 ist eine unabhängige Standortbestimmung der fachlichen Leistungen am Ende der Sekundarstufe I, die im Bildungsraum Nordwestschweiz regelmässig durchgeführt wird. Andererseits wurden die Inhalte der Aufnahmeprüfungen mit den Inhalten des Lehrplans der Sekundarschule des Kantons Aargau, des Lehrplan 21 und des Check S3 verglichen.

Zur *Beurteilung der Prüfungsform* wurde die Qualität der Aufnahmeprüfungen anhand von psychometrischen Gütekriterien ausgewiesen. Dazu wurde erstens die Bedeutung einzelner Prüfungsteile für das Bestehen der Prüfung ermittelt. Zweitens wurde untersucht, ob die unterschiedlichen Notendurchschnitte der geprüften Fächer gerechtfertigt sind. Drittens wurden die wichtigsten Testgütekriterien (Objektivität, Reliabilität und Validität) für die Aufnahmeprüfungen bestimmt. Viertens wurden die wichtigsten Gütekriterien von Prüfungsaufgaben (Schwierigkeit, Diskrimination und Fairness) ermittelt.

Als Datengrundlage dienten einerseits die Prüfungsergebnisse der Jahre 2017 bis 2019 in Form der Noten der geprüften Fächer und des Prüfungsergebnisses (bestanden/nicht bestanden), andererseits die von den Kandidatinnen und Kandidaten ausgefüllten Prüfungen aus dem Jahr 2019 und darüber hinaus auch die Ergebnisse im Check S3. Die Ergebnisse in den Aufnahmeprüfungen wurden mit den Check-S3-Ergebnissen verknüpft.

Prüfungsinhalte und Vorkenntnisse

Die fachlichen Vorkenntnisse sind für den Prüfungsentscheid von grosser Bedeutung. Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für eine Berufsfachschule zur Erreichung einer Berufsmaturität bestanden haben, verfügen im Durchschnitt über deutlich bessere Vorkenntnisse als solche, die die Prüfung nicht bestanden haben. Besonders wichtig erweisen sich die Vorkenntnisse in Mathematik und Deutsch, für die BM II auch in Französisch, während die Vorkenntnisse in Englisch für den Prüfungsentscheid kaum von Bedeutung sind.

Die Passung zwischen den Vorkenntnissen und den Inhalten der Aufnahmeprüfung für die BM I ist folglich in Mathematik und Deutsch am grössten. Kandidatinnen und Kandidaten, die in diesen Fachbereichen bessere Vorkenntnisse mitbringen, haben eine höhere Chance, die Aufnahmeprüfung zu bestehen. Darüber hinaus erhöhen bessere Fremdsprachenkenntnisse die Chance aber kaum mehr, die Aufnahmeprüfung zu bestehen. Gemessen an den Check-Ergebnissen als externem Massstab ist die Passung zwischen den Vorkenntnissen und den Inhalten der Aufnahmeprüfung für die BM I in Englisch und Französisch als nicht gegeben zu beurteilen. Eine abschliessende Beurteilung zur Passung zwischen den Prüfungsinhalten und den Vorkenntnissen ist für die BM II aufgrund der Datenlage – die Stichprobe ist dafür zu klein – nicht möglich.

Die Prüfungsanforderungen sind hoch. Ein durchschnittliches Ergebnis im Check S3 am Ende der obligatorischen Schulzeit reicht im Schnitt nicht aus, um die Aufnahmeprüfung für die BM I zu bestehen. Die Vorkenntnisse der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, sind am Ende der obligatorischen Schulzeit im Durchschnitt nur leicht geringer als jene von Schülerinnen und Schülern der Bezirksschule.

Prüfungsinhalte und Lehrplaninhalte

Die Inhalte der Prüfungen passen bis auf wenige Ausnahmen vollständig zu den Lehrplaninhalten. Einzig in den Fremdsprachen lassen sich wenige Abweichungen nachweisen. So sind insbesondere für die Aufgaben zum Wortschatz im Lehrplan der Sekundarschule keine Referenzen zu finden. Im künftig massgebenden Lehrplan 21 sind hingegen auch zu den Wortschatzaufgaben Entsprechungen zu finden. Auch die Inhalte des Check S3 decken sich grösstenteils mit den Inhalten der Aufnahmeprüfungen, wobei im Check S3 in den Fremdsprachen zusätzlich der Kompetenzbereich *Hören* erfasst wird und die Bereiche Wortschatz und Grammatik anwendungsorientiert beim Schreiben eines Textes geprüft werden.

Während die inhaltliche Passung zwischen Aufnahmeprüfungen und Lehrplänen als gegeben beurteilt werden kann, ist die Gewichtung der einzelnen Prüfungsinhalte mit den Zielen des Lehrplans und des Sprachunterrichts nicht kompatibel. So werden beispielsweise die Grammatikkenntnisse und der Wortschatz in den Sprachprüfungen deutlich stärker gewichtet als die Lesekompetenzen oder das Hörverständnis, das in den Aufnahmeprüfungen gänzlich fehlt. Aufgrund der marginalen Bedeutung der Lesekompetenzen in der Aufnahmeprüfung könnte der Eindruck entstehen, dass das Lesen für das erfolgreiche Absolvieren eines Maturitätslehrgangs im Vergleich zu Grammatikkenntnissen und Wortschatz von untergeordneter Bedeutung sei. Dies erscheint insbesondere in Anbetracht der vergleichsweise geringen Lesekompetenzen von Schweizer Jugendlichen problematisch (Konsortium PISA.ch, 2019). Der Grund für die geringe Präsenz der Lesekompetenzen in den Sprachprüfungen liegt vermutlich darin, dass die Entwicklung von geeigneten Aufgaben zur Erfassung von Lesekompetenzen wesentlich aufwendiger und schwieriger ist als die Entwicklung von Aufgaben zu Wortschatz und zu Grammatik.

Prüfungsteile und Prüfungsentscheid

Die einzelnen Prüfungsteile sind für den Prüfungsentscheid von unterschiedlicher Bedeutung. Für die allermeisten Kandidatinnen und Kandidaten würde eine deutlich kürzere Prüfung mit nur wenigen Prüfungsteilen – insbesondere ohne gewisse Prüfungsteile in den Fremdsprachen – zum gleichen Prüfungsentscheid führen. Aus einer ökonomischen Perspektive könnte die Prüfung ohne weitreichende Folgen auch in einer verkürzten Form mit weniger Bestandteilen durchgeführt werden. Aus einer pädagogischen Perspektive wäre das allerdings kaum zu verantworten. Vielmehr müsste die Qualität der einzelnen Prüfungsteile bzw. Aufgaben optimiert werden, sodass die einzelnen Kompetenzbereiche valide erfasst werden können. Besser, als die bestehenden Prüfungen zu kürzen, wäre es daher, die Prüfungen so zu überarbeiten, dass sämtliche Kompetenzbereiche, also auch jene in den Fremdsprachen, für den Prüfungsentscheid relevant würden. Dies gelingt durch eine Erhöhung der Anzahl und der Qualität der Aufgaben.

Beurteilungsmassstab nach Fach

Beurteilungsmassstäbe basieren auf dem Fachverständnis bzw. sind sie das Ergebnis von fachspezifischen Anforderungen, die von den Kandidatinnen und Kandidaten erwartet werden. Aufgrund der Ergebnisse im Check S3, der zu einer unabhängigen Abbildung des tatsächlichen Leistungsstands der Schülerschaft am Ende der obligatorischen Schulzeit führt, wurde überprüft, inwieweit die unterschiedlichen Beurteilungsmassstäbe aufgrund der Vorkenntnisse in der Schülerschaft gerechtfertigt – und somit nicht fachspezifisch – sind.

Aufgrund der Verteilung der fachlichen Leistungen in der Schülerschaft sind die unterschiedlichen Beurteilungsmassstäbe zum grossen Teil gerechtfertigt. Der Tendenz nach decken sich die Verteilungen der Prüfungsnoten mit den Verteilungen der tatsächlichen Leistungen. Die vergleichsweise tiefen Notendurchschnitte in Französisch und die vergleichsweise hohen Notendurchschnitte in Englisch der Prüfungen für die BM I werden aufgrund der Ergebnisse im Check S3 bestätigt. Etwas grösser sind die Abweichungen zwischen den Ergebnissen im Check S3 und den Prüfungsnoten in Deutsch und Mathematik. Allerdings kann die eher zu milde bzw. zu strenge Beurteilung in diesen Fächern jeweils nur für die Prüfung in einem Jahr nachgewiesen werden. Die Befunde für die BM I zeigen sich in ähnlicher Weise auch für die BM II. Insgesamt lassen sich die unterschiedlichen Beurteilungsmassstäbe durch die Ergebnisse im Check S3 relativ gut rechtfertigen. Insbesondere der tiefe Notendurchschnitt in Französisch ist aufgrund der tatsächlichen Leistungen in Französisch gerechtfertigt, ebenso wie der eher tiefe Notendurchschnitt in Mathematik. Eher etwas zu milde fällt die Beurteilung in Deutsch aus, eher etwas zu streng in Englisch.

Gütekriterien der Aufnahmeprüfungen

Die anhand der vorliegenden Prüfungen und Lösungsschlüssel durchgeführten Analysen zeigen eine Diskrepanz zwischen den Vorgaben zur Korrektur von Prüfungen und der faktischen Korrektur. Während die Dokumente informativ sind und weitgehend als ausreichend für eine objektive Auswertung beurteilt werden, weisen die korrigierten Prüfungen auf einige Mängel hin. Die Auswertungsobjektivität könnte durch Ergänzungen in den Unterlagen und insbesondere durch eine strengere Kontrolle der Korrekturen ohne grossen Aufwand optimiert werden. Vorwiegend positiv beurteilt wird die Reliabilität der Prüfungen, wobei diese mit geringen Anpassungen ebenfalls noch weiter erhöht werden könnte. Beispielsweise müssten in den Mathematikprüfungen mehr Aufgaben eingesetzt oder zumindest mehr Punkte vergeben werden. Positiv zu werten ist darüber hinaus, dass die Messgenauigkeit in allen Fächern dort am höchsten ist, wo die Prüfung genau messen muss, nämlich im Fähigkeitsbereich der Kandidatinnen und Kandidaten, die nicht eindeutig den Kategorien «bestanden» und «nicht bestanden» zugeordnet werden können. Die Validität der Aufnahmeprüfung in die BM I wird insgesamt als ausreichend beurteilt. Für die BM II liegen für eine abschliessende Beurteilung der Validität nicht genügend Daten vor.

Gütekriterien der Prüfungsaufgaben

Insgesamt sind die Gütekriterien der Prüfungsaufgaben als ausreichend bis sehr gut zu beurteilen. Es zeigt sich, dass die Prüfungsaufgaben jener Kompetenzbereiche, die für den Prüfungsentscheid eine untergeordnete Rolle spielen, tendenziell weniger gut diskriminieren. Insbesondere in den sprachlichen Fächern gibt es einen nicht vernachlässigbaren Anteil an nicht trennscharfen Aufgaben. Dies trifft vor allem für die Binary-Choice-Aufgaben zu, die im Bereich *Lesen* eingesetzt werden. Durch eine gezielte Verbesserung dieser Aufgaben sollten sich die Trennschärfen dieser Aufgaben erhöhen lassen. Zudem werden Binary-Choice-Aufgaben zur Reduktion der Ratewahrscheinlichkeit in der Regel nicht einzeln bewertet, sondern als Gruppen von vier Binary-Choice-Aussagen zu einer übergeordneten Aufgabenstellung.

Die Fairness der (dichotomen) Aufgabenstellungen kann in Bezug auf das Geschlecht als im Grossen und Ganzen gegeben beurteilt werden. Es gibt nur wenige Aufgaben, die bei Kandidatinnen unterschiedlich funktionieren als bei Kandidaten.

1 Ausgangslage

Der Kanton Aargau bildet rund dreimal so viele Jugendliche in der beruflichen Grundbildung aus als in der Allgemeinbildung (Gymnasium und Fachmittelschulen; vgl. Ender et al., 2017, S. 92). Die Attraktivität der beruflichen Grundbildung spiegelt sich auch in der Berufsmaturität, die sich als zusätzliches Bildungsangebot auf der Sekundarstufe II in den letzten Jahren eindrücklich etabliert hat. Während beispielsweise die gymnasiale Maturitätsquote in den Jahren 2000 und 2010 rund 15 Prozent und im Jahr 2016 rund 16 Prozent betrug, stieg die berufliche Maturitätsquote von rund 7 Prozent im Jahr 2000 auf rund 13 Prozent im Jahr 2010 und auf rund 17 Prozent im Jahr 2016 (vgl. Bundesamt für Statistik, 2017).

Die Berufsmaturität wird erlangt, indem zusätzlich zur beruflichen Grundbildung mit Eidgenössischem Fähigkeitszeugnis (EFZ) ein Maturitätslehrgang für den Erwerb einer erweiterten Allgemeinbildung abgeschlossen wird. Dieser Lehrgang kann in einer Berufsfachschule parallel zur beruflichen Grundbildung (BM I) oder nach Abschluss der beruflichen Grundbildung (BM II) absolviert werden. Folgende Ausrichtungen werden dabei unterschieden:

- Technik, Architektur, Life Sciences
- Natur, Landschaft und Lebensmittel
- Wirtschaft und Dienstleistungen: Typ Wirtschaft
- Wirtschaft und Dienstleistungen: Typ Dienstleistungen
- Gestaltung und Kunst
- Gesundheit und Soziales

Für den Besuch eines beruflichen Maturitätslehrgangs ist im Kanton Aargau in vielen Fällen eine Aufnahmeprüfung abzulegen. Während für die BM I für alle Ausrichtungen die gleiche Aufnahmeprüfung durchgeführt wird, unterscheidet sich für die BM II die Aufnahmeprüfung der Ausrichtung «Wirtschaft und Dienstleistungen: Typ Wirtschaft» von der Aufnahmeprüfung der übrigen Ausrichtungen. Für die Koordination der Aufnahmeprüfungen ist die Abteilung Berufsbildung und Mittelschule des Departements Bildung, Kultur und Sport zuständig. Im Rahmen dieser Aufgabe wurde der Stoffplan der Aufnahmeprüfungen überprüft und dabei unter anderem im Bereich Fremdsprachen Handlungsbedarf festgestellt. Aus diesem Grund wurden die Aufnahmeprüfungen einer externen Beurteilung unterzogen.

Mit der Evaluation der Aufnahmeprüfungen in einen beruflichen Maturitätslehrgang im Kanton Aargau wurde insbesondere überprüft, ob die Anforderungen der Aufnahmeprüfungen mit Blick auf die Vorkenntnisse der Kandidatinnen und Kandidaten angemessen sind. Zugleich wurden die Aufnahmeprüfungen der Jahre 2017 bis 2019 einer Qualitätsanalyse aus inhaltlicher und psychometrischer Perspektive unterzogen.

Gegenstand der Beurteilung waren die Aufnahmeprüfungen für eine Berufsfachschule zur Erreichung einer Berufsmaturität, nicht aber die Aufnahmeprüfungen für eine Fachmittelschule, die zu einer Fachmaturität führen, oder die Aufnahmeprüfungen für eine Informatikmittelschule, die ebenfalls in einer Berufsmaturität münden.

2 Ausbildungswege und Aufnahmeprüfungen

2.1 Berufsfachschule mit Berufsmaturität während der Lehre (BM I)

Aufnahmebedingungen
Die Aufnahme in eine Berufsfachschule während der Lehre ist für Schülerinnen und Schüler der Bezirksschule prüfungsfrei, wenn sie am Ende der Abschlussklasse oder am Ende des 1. Semesters der Abschlussklasse in den Fächern einen Notendurchschnitt von mindestens 4.4 erzielt haben oder über einen gleichwertigen Schulabschluss verfügen.[1] Sind diese Bedingungen nicht erfüllt, muss eine Aufnahmeprüfung absolviert werden.

Schülerinnen und Schüler der Sekundarschule werden prüfungsfrei in eine Berufsfachschule aufgenommen, wenn sie am Ende der Abschlussklasse oder am Ende des 1. Semesters der Abschlussklasse in den Fächern einen Notendurchschnitt von mindestens 5.3 erzielt haben oder über einen gleichwertigen Schulabschluss verfügen. Sind diese Bedingungen nicht erfüllt, muss eine Aufnahmeprüfung absolviert werden.

Prüfungsform
Die Aufnahmeprüfung besteht aus schriftlichen Prüfungen in den Fächern Deutsch (90 Minuten), Französisch (45 Minuten), Englisch (45 Minuten) und Mathematik (60 Minuten). Der Durchschnitt der Prüfungsnoten Französisch und Englisch ergibt die Fremdsprachennote. Die Noten werden in sämtlichen Fächern (Fachnote Deutsch, Fachnote Mathematik und Fachnote Fremdsprachen) auf eine halbe Note gerundet.

Die Aufnahmeprüfung gilt als bestanden, wenn der Durchschnitt der drei gerundeten Fachnoten Deutsch, Mathematik und Fremdsprachen mindestens 4.0 beträgt und höchstens eine dieser drei Fachnoten unter 4.0 liegt. Der Durchschnitt der drei Fachnoten wird auf eine Dezimalstelle gerundet.

Prüfungsinhalt
Im Kanton Aargau wird in allen Berufsfachschulen die gleiche Prüfung durchgeführt. Der Prüfungsinhalt basiert auf dem Lehrplan der Abschlussklassen der Sekundarschule.

2.2 Berufsfachschule mit Berufsmaturität nach der Lehre (BM II) – Ausrichtung Wirtschaft und Dienstleistungen: Typ Wirtschaft

Aufnahmebedingungen
Die Aufnahme in eine Berufsfachschule für den Typ Wirtschaft der Ausrichtung Wirtschaft und Dienstleistungen ist nach der Lehre für gelernte Berufsleute mit eidgenössischem Fähigkeitszeugnis prüfungsfrei, wenn sie einen Abschluss als Kaufmann/Kauffrau EFZ E-Profil mit einer schulischen Gesamtnote von mindestens 4.7 erreicht oder im Zeugnis für das 5. Semester einen Notendurchschnitt von mindestens 4.7 in den Fächern Wirtschaft und Gesellschaft (Gewichtung: 2/5),

[1] Informationen zur Berechnung des Notendurchschnitts bei Schülerinnen und Schülern der Bezirks- und Sekundarschule finden sich im Anhang der kantonalen Verordnung über die Berufsmaturität an Berufsfachschulen (V Berufsmaturität BFS).

Deutsch (Gewichtung: 1/5), Französisch (Gewichtung: 1/5) und Englisch (Gewichtung: 1/5) erzielt haben. Für das Fach Englisch zählt die Zeugnisnote des 4. Semesters. Sind diese Bedingungen nicht erfüllt, muss eine Aufnahmeprüfung absolviert werden.

Prüfungsform
Für die Zulassung zur Berufsfachschule für den Typ Wirtschaft der Ausrichtung Wirtschaft und Dienstleistungen werden die folgenden Fächer schriftlich geprüft: Deutsch (90 Minuten), Französisch (60 Minuten), Englisch (60 Minuten) und Finanz- und Rechnungswesen (120 Minuten). Der Durchschnitt der Noten in Englisch und Französisch ergibt die Fremdsprachennote. Die Noten werden in sämtlichen Fächern (Fachnote Deutsch, Fachnote Finanz- und Rechnungswesen und Fachnote Fremdsprachen) auf eine halbe Note gerundet.

Die Aufnahmeprüfung gilt als bestanden, wenn der Durchschnitt der gerundeten Fachnote Fremdsprachen und der Fachnoten Deutsch und Finanz- und Rechnungswesen mindestens 4.0 beträgt und höchstens eine dieser drei Fachnoten unter 4.0 liegt. Der Durchschnitt der drei Fachnoten wird auf eine Dezimalstelle gerundet.

Prüfungsinhalt
Der Prüfungsinhalt richtet sich nach dem geltenden Rahmenlehrplan für die Berufsmaturität des Staatssekretariats für Bildung, Forschung und Innovation (SBFI) und basiert auf dem Stoff, der bis zum Abschluss der beruflichen Grundbildung Kaufmann/Kauffrau EFZ E-Profil behandelt wurde.

2.3 Berufsfachschule mit Berufsmaturität nach der Lehre (BM II) – Alle Ausrichtungen mit Ausnahme von Wirtschaft und Dienstleistungen: Typ Wirtschaft

Aufnahmebedingungen
Für die Aufnahme in eine Berufsfachschule nach der Lehre für eine Ausrichtung, die nicht Wirtschaft und Dienstleistungen des Typs Wirtschaft entspricht, müssen gelernte Berufsleute mit einem eidgenössischen Fähigkeitszeugnis (EFZ) eine Aufnahmeprüfung bestehen.

Prüfungsform
Es werden die folgenden Fächer geprüft: Deutsch (90 Minuten), Französisch (45 Minuten), Englisch (45 Minuten) und Mathematik (60 Minuten). Der Durchschnitt der Noten in Englisch und Französisch ergibt die Fachnote Fremdsprachen. Die Noten werden in sämtlichen Fächern (Fachnote Deutsch, Fachnote Mathematik und Fachnote Fremdsprachen) auf eine halbe Note gerundet.

Die Aufnahmeprüfung gilt als bestanden, wenn der Durchschnitt der Fachnoten Fremdsprachen, Deutsch und Mathematik mindestens 4.0 beträgt und höchstens eine dieser drei Fachnoten unter 4.0 liegt. Der Durchschnitt der drei Fachnoten wird auf eine Dezimalstelle gerundet.

Prüfungsinhalt
In allen Berufsfachschulen wird die gleiche Prüfung durchgeführt. Der Prüfungsinhalt basiert auf dem Lehrplan der Abschlussklassen der Sekundarschule.

3 Ziele und Fragestellungen

Mit der Evaluation der Aufnahmeprüfungen in die Berufsmaturität wurden zwei Ziele verfolgt. Zum einen sollte überprüft werden, wie gut die Inhalte der Aufnahmeprüfungen mit den Vorkenntnissen der Kandidatinnen und Kandidaten übereinstimmen. Zum anderen sollte die Qualität

der Aufnahmeprüfungen aus einer psychometrischen Perspektive beurteilt werden. Gegenstand der Evaluation sind folglich die Prüfungsinhalte und die Prüfungsform.

Zur *Beurteilung der Prüfungsinhalte* wurden einerseits die Vorkenntnisse der Kandidatinnen und Kandidaten durch einen Vergleich der Prüfungsergebnisse mit den Ergebnissen im Check S3 beurteilt. Der Check S3 ist eine unabhängige Standortbestimmung der fachlichen Leistungen am Ende der Sekundarstufe I, die im Bildungsraum Nordwestschweiz regelmässig durchgeführt wird. Mit dem Check S3 werden die Vorkenntnisse der Kandidatinnen und Kandidaten schultypen- und klassenunabhängig beurteilt, wodurch ein geeignetes externes Mass zur Beurteilung der Prüfungsanforderungen im Vergleich zu den Vorkenntnissen am Ende der obligatorischen Schulzeit vorliegt.

Andererseits wurden die Inhalte der Aufnahmeprüfungen mit den Inhalten des Lehrplans der Sekundarschule des Kantons Aargau, des Lehrplan 21 und des Check S3 verglichen. Mit Ausnahme der Aufnahmeprüfung für den Lehrgang «BM II, Ausrichtung Wirtschaft und Dienstleistungen: Typ Wirtschaft» basieren alle Prüfungen auf dem Lehrplan der Abschlussklassen der Sekundarschule. Der Prüfungsinhalt für den Lehrgang «BM II, Ausrichtung Wirtschaft und Dienstleistungen: Typ Wirtschaft» basiert auf dem Stoff, der bis zum Abschluss der beruflichen Grundbildung Kaufmann/Kauffrau EFZ E-Profil behandelt wurde. Für die Inhalte der Aufnahmeprüfung in die BM II Typ «Wirtschaft» gilt deshalb der Rahmenlehrplan für die Berufsmaturität des Staatssekretariats für Bildung und Forschung (SBFI). Der Rahmenlehrplan enthält Stundentafeln zu den einzelnen Fachbereichen, jedoch keine konkreten Angaben zu Inhalten. Aus diesem Grund bezieht sich der Vergleich der Prüfungsinhalte und der Lehrplaninhalte ausschliesslich auf die Aufnahmeprüfung in die BM I und in die BM II Typ «andere». Mit *BM II Typ «andere»* werden im Folgenden die Lehrgänge der BM II ohne den Typ «Wirtschaft» bezeichnet.

Zur Beurteilung der *Prüfungsform* werden die Prüfungen und die Prüfungsergebnisse beigezogen. Zu den Prüfungsergebnissen gehören der Prüfungsentscheid (bestanden/nicht bestanden), die Punktzahlen und Noten in den einzelnen Prüfungsteilen sowie die Lösungen der einzelnen Prüfungsaufgaben. Auch für die Beurteilung der Prüfungsform wurden zum Teil die Check-S3-Ergebnisse der Kandidatinnen und Kandidaten beigezogen.

Die folgenden spezifischen Fragestellungen beziehen sich auf den Prüfungsinhalt (PI) und die Prüfungsform (PF):

Prüfungsinhalt

PI1 Entsprechen die Prüfungsinhalte den Vorkenntnissen der Kandidatinnen und Kandidaten?

PI2 Unterscheiden sich die Inhalte der Aufnahmeprüfungen von den Inhalten des Lehrplans?

Prüfungsform

PF1 Welche Bedeutung haben die einzelnen geprüften Fächer bzw. Prüfungsteile (bspw. *Verfassen eines Textes*, *Leseverständnis*) für den Erfolg bei der Aufnahmeprüfung?

PF2 Sind die unterschiedlichen Notendurchschnitte der geprüften Fächer gerechtfertigt?

PF3 Erfüllen die vorliegenden Aufnahmeprüfungen die wichtigsten Gütekriterien von Prüfungen (Objektivität, Reliabilität und Validität)?

PF4 Erfüllen die einzelnen Aufgaben die wichtigsten Gütekriterien von Prüfungsaufgaben (Schwierigkeit, Diskrimination und Fairness)?

4 Datengrundlage

Für die Beantwortung der Fragestellungen werden verschiedene Datenquellen genutzt.

Prüfungsergebnisse
Die Prüfungsergebnisse wurden einerseits in Verbindung mit den Ergebnissen im Check S3 zur Beurteilung der Passung zwischen den Aufnahmeprüfungen und den Vorkenntnissen genutzt (Fragestellung PI1), andererseits für die Beurteilung der Prüfungsform, insbesondere bezüglich der Bedeutung einzelner Prüfungsfächer und Prüfungsteile für den Erfolg bei der Aufnahmeprüfung (Fragestellung PF1). Einbezogen wurden die Prüfungsergebnisse der Jahrgänge 2017, 2018 und 2019 für beide Ausbildungswege (BM I und BM II). Für die Analysen werden folgende Informationen genutzt:

- Aufnahmeprüfung Prüfungsentscheid (bestanden/nicht bestanden)
- Prüfungsnoten (Mathematik, Deutsch, Französisch, Englisch)

Ergebnisse im Check S3
Die Ergebnisse im Check S3 wurden einerseits in Verbindung mit den Prüfungsergebnissen zur Beurteilung der Passung zwischen den Aufnahmeprüfungen und den Vorkenntnissen genutzt (Fragestellung PI1), andererseits für die Beurteilung der Notendurchschnitte in den geprüften Fächern (Fragestellung PF2). Einbezogen wurden die Ergebnisse der Jahre 2016, 2017, 2018 und 2019 als Mass für die Vorkenntnisse der Schülerinnen und Schüler am Ende der obligatorischen Schulbildung.

Für die BM konnte die Verbindung der Daten nur anhand der Ergebnisse im Check S3 des Jahres 2016 und der Aufnahmeprüfung 2019 gemacht werden, weil nur in dieser Konstellation mindestens drei Jahre zwischen dem Check S3 und der Aufnahmeprüfung liegen. Dabei gilt es zu beachten, dass der Check S3 im Kanton Aargau im Jahr 2016 freiwillig durchgeführt wurde. Es standen Daten von 2'020 Schülerinnen und Schülern zur Verfügung, was rund einem Drittel der Population entspricht.

Aufnahmeprüfungen und Lösungsschlüssel
Für die Beurteilung der Qualität der Prüfungen aus inhaltlicher Perspektive (Fragestellung PI2) wurden die Aufnahmeprüfungen und Lösungsschlüssel sowie weitere Informationen zu den Prüfungen der Jahrgänge 2017, 2018 und 2019 genutzt.

Rohdaten der Prüfungen
Für die Beurteilung der Qualität der Prüfungen aus psychometrischer Perspektive (Fragestellung PF3) sowie für die Beurteilung der Prüfungsaufgaben aus psychometrischer Perspektive (Fragestellung PF4) wurden Rohdaten in Form der Antworten der Kandidatinnen und Kandidaten zu den einzelnen Aufgaben genutzt. Dazu wurden die Antworten zu den einzelnen Prüfungsaufgaben der Kandidatinnen und Kandidaten des Jahrgangs 2019 (BM I und BM II) elektronisch erfasst. Für die Analysen wurde die Information genutzt, ob die jeweilige Aufgabe richtig oder falsch gelöst wurde.

Verbindung von Prüfungsergebnissen und Check-Ergebnissen
Für eine aussagekräftige Überprüfung der Passung zwischen den Prüfungsanforderungen und den Vorkenntnissen wurden die Prüfungsergebnisse mit den Ergebnissen im Check S3 verbunden (Fragestellung PI1). Diese Verbindung wurde anhand des Namens und des Vornamens sowie des Geburtsdatums der jeweiligen Kandidatin oder des jeweiligen Kandidaten hergestellt.

5 Methodisches Vorgehen

Für die Beantwortung der Fragen wurde mehrheitlich ein empirisch quantitativer Zugang gewählt. Im Folgenden wird zuerst das methodische Vorgehen zur Beantwortung der einzelnen Fragestellungen kurz skizziert. Darauf folgt eine ausführliche Beschreibung der verwendeten statistischen Begriffe und Kennwerte sowie der wichtigsten statistischen Methoden.

5.1 Überblick zu Fragestellung, methodischem Vorgehen und Ausbildungsweg

Prüfungsinhalte und Vorkenntnisse
Die Beurteilung, wie gut die Prüfungsinhalte den Vorkenntnissen der Kandidatinnen und Kandidaten entsprechen, erfolgte ausschliesslich auf empirischer Basis anhand der Ergebnisse im Check S3. Dazu wurden die Prüfungsergebnisse (bestanden/nicht bestanden) mit den Ergebnissen in den fachlichen Leistungstests des Check S3 verglichen. Es wurde mit Hilfe von deskriptiven Analysen überprüft, wie stark sich die Check-Ergebnisse von erfolgreichen und nicht erfolgreichen Kandidatinnen und Kandidaten einer Berufsfachschule unterscheiden bzw. wie hoch die Anforderungen der Aufnahmeprüfungen gemessen an den Check-Ergebnissen sind. Weiter wurde mit logistischen Regressionsanalysen untersucht, wie gut die Prüfungsergebnisse durch die Check-Ergebnisse erklärt bzw. vorhergesagt werden können.

Für die BM I konnten die Daten der Aufnahmeprüfungen der Jahrgänge 2017, 2018 und 2019 sowie die entsprechenden Ergebnisse im Check S3 dieser Jahre genutzt werden. Für die BM II können nur die Daten der Aufnahmeprüfung 2019 und die Ergebnisse im Check S3 des Jahres 2016 genutzt werden, weil nur in dieser Konstellation mindestens drei Jahre zwischen dem Check S3 und der Aufnahmeprüfung liegen.

Prüfungsinhalte und Lehrplaninhalte
Für die Beurteilung der Übereinstimmung von Prüfungsinhalten und Lehrplaninhalten wurde eine Inhaltsanalyse durchgeführt. Die Inhalte der Aufnahmeprüfungen für die BM I und BM II Typ «andere» wurden zum einen mit den Inhalten des Lehrplans der Sekundarschule des Kantons Aargau und des Lehrplan 21, zum anderen mit den Inhalten des Check S3 verglichen.

Die Inhalte der Aufnahmeprüfungen für die BM II Typ «Wirtschaft» konnten aufgrund fehlender inhaltlicher Angaben im Rahmenlehrplan des SBFI nicht beurteilt werden.

Bedeutung einzelner Prüfungsfächer und Prüfungsteile
Die Beurteilung der Bedeutung der einzelnen geprüften Fächer sowie einzelner Prüfungsteile (bspw. *Verfassen eines Textes*, *Leseverständnis*) für den Prüfungserfolg erfolgte anhand von logistischen Regressionsanalysen. Mit diesen Analysen wurde überprüft, inwieweit der Prüfungsentscheid (bestanden/nicht bestanden) von den einzelnen geprüften Fächern oder den einzelnen Prüfungsteilen abhängt.

Beurteilungsmassstab der geprüften Fächer
Zur Beantwortung der Frage, ob unterschiedliche Notendurchschnitte der geprüften Fächer gerechtfertigt sind, wurden die Notendurchschnitte der geprüften Fächer mit den Ergebnissen im Check S3 verglichen. Der Vergleich erfolgte anhand der Verteilungen der Prüfungsnoten und der Check-Ergebnisse. Die Check-Ergebnisse ergeben sich aufgrund eines unabhängigen Leistungstests und spiegeln das Leistungsspektrum der Schülerinnen und Schüler am Ende der obligatorischen Schulbildung unabhängig vom besuchten Schultyp. Durch den Vergleich der Verteilungen

der Prüfungsergebnisse mit den Check-Ergebnissen wurde anhand eines externen Kriteriums festgestellt, ob bei den Aufnahmeprüfungen je nach Prüfungsfach unterschiedlich strenge Beurteilungsmassstäbe angewendet wurden und inwiefern die unterschiedlichen Beurteilungsmassstäbe gemessen an den zu erwartenden Leistungen am Ende der Volksschule gerechtfertigt sind.

Gütekriterien der Aufnahmeprüfungen

Zu den wichtigsten Qualitätsmerkmalen einer Aufnahmeprüfung zählen Objektivität, Reliabilität und Validität.

Prüfungen werden als objektiv bezeichnet, wenn das Ergebnis nicht von den Personen, welche die Prüfung durchführen, auswerten und interpretieren, abhängig ist. Die Objektivität der Prüfungen wurde aufgrund einer Analyse von Aufgaben und Lösungsschlüsseln beurteilt. Dabei ging es vorwiegend um die Beurteilung der Auswertungsobjektivität. Insbesondere wurde die Frage beantwortet, ob einzelne Aufgaben stets auf die gleiche Weise korrigiert wurden.

Prüfungen werden als reliabel bezeichnet, wenn sie möglichst genau und ohne Messfehler messen. Die Reliabilität bzw. die Messgenauigkeit der Prüfung wurde anhand der Rohdaten der Aufnahmeprüfungen 2019 bestimmt. Zum einen wurden Reliabilitätskoeffizienten der Klassischen Testtheorie (KTT) berechnet (interne Konsistenz nach Cronbach und Split-Half-Reliabilität). Zum anderen wurden Reliabilitätsmasse berechnet, die auf der Item Response Theory (IRT) basieren, beispielsweise die Verteilung des Standardfehlers über das gesamte Fähigkeitsspektrum. Zudem wurde überprüft, ob die Messgenauigkeit tatsächlich beim Schwellenwert (bestanden/nicht bestanden) am höchsten ist, was idealerweise der Fall wäre.

Prüfungen werden als valide bezeichnet, wenn sie auch wirklich das messen, was sie messen sollten. Für die Beurteilung der Validität der Prüfungen wurden die Ergebnisse in den geprüften Fachbereichen miteinander und mit den Ergebnissen im Check S3 verglichen und so die konvergente und die divergente Validität bestimmt.

Gütekriterien der Prüfungsaufgaben

Für die Beurteilung der Gütekriterien der Prüfungsaufgaben wurden ebenfalls die Rohdaten der Aufnahmeprüfungen 2019 genutzt. Diese Rohdaten wurden einer sogenannten Itemanalyse aufgrund der Klassischen Testtheorie und der Item Response Theory unterzogen. So konnten für jede Aufgabe Gütekriterien wie die Aufgabenschwierigkeit oder die Trennschärfe ausgewiesen werden. Ausserdem wurde untersucht, ob die Aufgaben für weibliche und männliche Kandidierende gleich schwierig und somit fair sind (Analyse von Differential Item Functioning).

Zusammenfassung

Tabelle 1 gibt einen Überblick über die Fragestellungen, die Methoden und die verwendeten Daten. Die Datengrundlage wird für die Ausbildungswege getrennt dargestellt.

Tabelle 1: Fragestellung, Methode und Daten nach Ausbildungsweg

Fragestellung (kursiv) und Methode	Daten BM I	Daten BM II
PI1 *Prüfungsinhalte und Vorkenntnisse* Abgleich Prüfungsergebnisse mit Ergebnissen im Check S3	Prüfungsergebnisse und Ergebnisse im Check S3 der Jahre 2017, 2018 und 2019 in elektronischer Form	Prüfungsergebnisse im Jahr 2019 und Ergebnisse des Check S3 im Jahr 2016 in elektronischer Form
PI2 *Prüfungsinhalte und Lehrplaninhalte* Inhaltsanalyse	Aufnahmeprüfungen und Lösungsschlüssel der Jahre 2017, 2018 und 2019 in Papierform	Aufnahmeprüfungen und Lösungsschlüssel der Jahre 2017, 2018 und 2019 in Papierform
PF1 *Bedeutung von Prüfungsfächern und Prüfungsteilen* Logistische Regression	Prüfungsergebnisse der Jahre 2017, 2018, 2019 in elektronischer Form	Prüfungsergebnisse der Jahre 2017, 2018, 2019 in elektronischer Form
PF2 *Beurteilungsmassstab der geprüften Fächer* Vergleich der Verteilungen der Prüfungsnoten und der Ergebnisse im Check S3	Prüfungsergebnisse und Ergebnisse im Check S3 der Jahre 2017, 2018 und 2019 in elektronischer Form	Prüfungsergebnisse im Jahr 2019 und Ergebnisse des Check S3 im Jahr 2016 in elektronischer Form
PF3 *Gütekriterien der Aufnahmeprüfungen* Beurteilung der Objektivität Beurteilung der Reliabilität Beurteilung der Validität	Aufnahmeprüfungen und Lösungsschlüssel der Jahre 2017, 2018 und 2019 in Papierform Aufnahmeprüfungen 2019 in bearbeiteter Form (Rohdaten der Prüfungen)	Aufnahmeprüfungen und Lösungsschlüssel der Jahre 2017, 2018 und 2019 in Papierform Aufnahmeprüfungen 2019 in bearbeiteter Form (Rohdaten der Prüfungen)
PF4 *Gütekriterien der Prüfungsaufgaben* Itemanalyse nach KKT und IRT	Aufnahmeprüfungen 2019 in bearbeiteter Form (Rohdaten der Prüfungen)	Aufnahmeprüfungen 2019 in bearbeiteter Form (Rohdaten der Prüfungen)

5.2 Statistische Methoden, Begriffe und Kennwerte

Korrelation

Mit dem Korrelationskoeffizienten r, der zwischen –1 und +1 variieren kann, wird der Zusammenhang zwischen zwei Merkmalen beschrieben (vgl. bspw. Bortz & Schuster, 2010, S. 156–170). Ein positiver Korrelationskoeffizient steht für einen positiven Zusammenhang im Sinne von «je mehr X desto mehr Y», ein negativer Koeffizient steht für einen negativen Zusammenhang im Sinne von «je mehr X desto weniger Y». Je höher der absolute Wert von r ist, desto stärker ist der Zusammenhang zwischen zwei Merkmalen.

Statistische Signifikanz
Zusammenhänge zwischen geprüften Kompetenzen und dem Prüfungserfolg werden anhand statistischer Testverfahren auf Signifikanz hin geprüft. Nur statistisch signifikante Unterschiede oder Zusammenhänge dürfen interpretiert werden. Ist ein Unterschied nicht statistisch signifikant, dann kann nicht mit ausreichender Sicherheit ausgeschlossen werden, dass Zusammenhänge oder Unterschiede nur zufällig zustande gekommen sind (vgl. bspw. Hirsig, 2001, S. 55). In diesem Bericht erfüllen statistisch signifikante Ergebnisse die Bedingung, dass die Irrtumswahrscheinlichkeit bei weniger als 5 Prozent liegt ($p < .05$).

Effektstärke
Zur Beurteilung der Bedeutung eines Unterschieds wird jeweils die Effektstärke d angegeben. Die Effektstärke d ist ein Mass zur Beschreibung der Grösse beziehungsweise der Stärke eines Unterschieds zwischen zwei statistischen Kennwerten (z.B. zwischen zwei Gruppenmittelwerten). Sie steht in Ergänzung zur statistischen Signifikanzangabe. Die Effektstärke d berechnet sich durch die Differenz zweier Gruppenmittelwerte im Verhältnis zur Streuung innerhalb der beiden Gruppen und entspricht somit einer standardisierten Mittelwertdifferenz, die in Standardabweichungen ausdrückt, wie gross der Unterschied zwischen zwei Gruppen ist. Effekte von $d = .20$ werden als klein, Effekte von $d = .50$ als mittel und Effekte von $d = .80$ als gross angesehen (vgl. Cohen, 1988).

Standardisierung durch z-Transformation
Testergebnisse, die aus unterschiedlichen Erhebungen stammen und gegebenenfalls auf unterschiedlichen Skalen erhoben wurden, können nicht ohne Weiteres miteinander verglichen werden. Erreicht eine Kandidatin in der Englischprüfung die Note 4.5 und im Englisch-Check 801 Punkte, ist nicht klar, ob sie in der Prüfung oder im Check besser abgeschnitten hat. Um diesen Vergleich zu ermöglichen, können die Werte standardisiert (z-transformiert) werden (vgl. bspw. Bortz, 1993, S. 45). Bei einer z-Transformation wird von den Werten der Kandidatinnen und Kandidaten zunächst der Mittelwert der gesamten Gruppe abgezogen. Der transformierte Gruppenmittelwert liegt dadurch bei 0 und es lässt sich einfach beurteilen, ob das Ergebnis einer Kandidatin über (positiver Wert) oder unter (negativer Wert) dem Gruppendurchschnitt liegt. In einem zweiten Schritt wird die individuelle Abweichung vom Mittelwert mit der Streuung innerhalb der gesamten Gruppe in Beziehung gesetzt, indem durch die Standardabweichung dividiert wird. Die standardisierten Werte werden als z-Werte bezeichnet und messen die Anzahl Standardabweichungen vom Mittelwert. Die z-Werte streuen in der Regel in Form einer Normalverteilung zwischen –3 und +3 um den Mittelwert $M = 0$.

Objektivität
Eine Prüfung wird als objektiv bezeichnet, wenn das Ergebnis nicht von den Personen, welche die Prüfung durchführen, auswerten und interpretieren, abhängig ist. Im Rahmen der vorliegenden Evaluation wurde die Auswertungsobjektivität beurteilt. Dazu wurden zum einen die Lösungsschlüssel einer kritischen Beurteilung unterzogen und zum anderen die Korrektur und Beurteilung der Prüfungsaufgaben auf Konsistenz und Genauigkeit überprüft.

Reliabilität
Mit der Reliabilität wird die Messgenauigkeit einer Prüfung angegeben. Eine Prüfung ist dann reliabel, wenn sie möglichst genau und ohne Messfehler misst bzw. der Messfehler möglichst gering ist. Das Ausmass an Reliabilität wird über Reliabilitätskoeffizienten angegeben.

Bei der am häufigsten verwendeten Methode zur Berechnung der Reliabilität wird ein Test in Einzelaufgaben zerlegt und der Mittelwert der Interkorrelationen aller Teilaufgaben berechnet. Dieser

Koeffizient entspricht dem sogenannten Cronbach-Alpha (vgl. Cronbach, 1951). Der Koeffizient kann Werte zwischen 0 und 1 annehmen. Ein Testinstrument sollte nur dann verwendet werden, wenn ein Wert von $\alpha = .70$ oder mehr erreicht wird. Da Prüfungen zur Beurteilung einer individuellen Eignung eingesetzt werden, sollte der Wert deutlich höher als $\alpha = .70$ sein. Grund für eine relativ tiefe Reliabilität kann die Anzahl der Aufgaben sein. Je mehr Aufgaben ein Test umfasst, desto höher wird die Reliabilität. Die Reliabilität hängt zudem auch davon ab, ob ein Test eher heterogen oder homogen ist. Je ähnlicher die Aufgaben sind, desto einfacher ist es, eine hohe Reliabilität in einer Prüfung zu erreichen.

Eine weitere gängige Methode zur Berechnung der Reliabilität von Prüfungen, die nur einmal eingesetzt werden, ist die Berechnung der Testhalbierungs-Reliabilität (Split-Half-Reliabilität). Dazu wird eine Prüfung in zwei äquivalente Hälften geteilt und die Korrelation r der beiden Hälften berechnet. Da die Reliabilität mit der Testlänge steigt, werden die Koeffizienten mit der Spearman-Brown prophecy formula auf die doppelte Testlänge hochgerechnet. Für das Aufteilen in zwei Testhälften gibt es unterschiedliche Ansätze. Hier wird die Odd-Even-Methode angewendet, welche nach der Reihenfolge der Aufgaben in der Prüfung immer abwechslungsweise eine Aufgabe der ersten oder der zweiten Hälfte zuordnet. Ein guter Test sollte eine Reliabilität von $r = .80$ aufweisen. Werte von .80 bis .90 werden als «gut» und Werte von .90 bis 1.00 als «hervorragend» eingestuft (vgl. bspw. Weise, 1975, S. 219).

Validität

Die Validität gibt an, ob mit der Prüfung auch wirklich das Merkmal gemessen wird, das gemessen werden sollte.

Durch den Vergleich der Prüfungsinhalte mit den Vorkenntnissen der Kandidatinnen und Kandidaten bzw. mit den Inhalten der Lehrpläne wird die inhaltliche Validität im Sinne eines Expertenurteils geprüft.

Eine weitere Möglichkeit, die Validität einer Prüfung auszuweisen, bietet die Konstruktvalidität. Diese ist gegeben, wenn von den Prüfungsergebnissen tatsächlich auf die Fähigkeiten geschlossen werden kann. Dazu kann überprüft werden, wie gut die Prüfungsergebnisse beispielsweise im Prüfungsteil Mathematik mit den Mathematikergebnissen im Check S3 übereinstimmen (konvergente Validität). Umgekehrt kann überprüft werden, ob das Prüfungsergebnis in Mathematik mit Prüfungsergebnissen in anderen Kompetenzbereichen, die nichts mit Mathematik zu tun haben, nicht bzw. nur schwach korrelieren (divergente Validität).

Zur Beurteilung der Konstruktvalidität wurde eine Multimethod-Multitrait-Analyse (MTMM-Analyse) durchgeführt. Damit ein Messinstrument im Sinne der MTMM-Analyse als valide beurteilt werden kann, müssen sowohl die konvergente als auch die divergente Validität erfüllt sein (vgl. bspw. Amelang & Schmidt-Atzert, 2006, S. 160).

Schwierigkeit einer Aufgabe

Die Schwierigkeit einer Aufgabe kann durch die Lösungshäufigkeit beziehungsweise den Anteil an Prüfungsteilnehmenden, der die Aufgabe richtig gelöst oder eine gewisse Teilpunktzahl erreicht hat, dargestellt werden.

Eine Lösungshäufigkeit zwischen 75 und 100 Prozent deutet auf eine (sehr) einfache Aufgabe hin, eine Lösungshäufigkeit von 50 bis 75 Prozent auf eine eher einfache Aufgabe, eine Lösungshäufigkeit zwischen 25 und 50 Prozent auf eine eher schwierige Aufgabe und eine Lösungshäufigkeit von weniger als 25 Prozent auf eine (sehr) schwierige Aufgabe.

Für die Berechnung der Schwierigkeit einer Aufgabe werden für nicht bearbeitete Aufgaben – gleich wie für falsch gelöste Aufgaben – 0 Punkte vergeben.

Damit die Leistungen der Kandidatinnen und Kandidaten gut differenziert und beschrieben werden können, sollten in einem Test sowohl einfache wie auch schwierige Aufgaben eingesetzt werden. Extrem einfache oder extrem schwierige Testaufgaben sind hingegen eher zu vermeiden. Die mittlere Schwierigkeit des Tests liegt idealerweise bei 50 Prozent.

Trennschärfe einer Aufgabe

Die Trennschärfe gibt an, wie gut die Aufgaben zwischen guten und schwachen Kandidatinnen und Kandidaten trennen beziehungsweise wie stark das Ergebnis bei einer Aufgabe mit dem Gesamtpunktwert der Prüfung korreliert. Eine Aufgabe ist trennscharf, wenn zwischen dem Aufgabenpunktwert und dem Gesamtpunktwert der Prüfung, von der sie Teil ist, eine hinreichende Korrelation besteht. Ein hoher Trennschärfekoeffizient bedeutet, dass erfolgreiche Kandidatinnen und Kandidaten die Aufgabe richtig lösen und nicht erfolgreiche die Aufgabe falsch lösen. Ein niedriger Trennschärfekoeffizient (um 0) bedeutet, dass erfolgreiche und nicht erfolgreiche Kandidatinnen und Kandidaten die Aufgabe gleich häufig richtig oder falsch lösen. Eine solche Aufgabe liefert keine Information über die Eignung eines Kandidaten oder einer Kandidatin. Ein negativer Trennschärfekoeffizient weist darauf hin, dass die Aufgabe entweder falsch korrigiert wurde oder fehlerhaft ist.

Differential Item Functioning

Anhand des Differential Item Functioning (DIF) kann analysiert werden, ob die Prüfungsaufgaben fair sind. Fair ist eine Prüfungsaufgabe, wenn sie von unterschiedlichen Gruppen – beispielsweise von weiblichen und männlichen Kandidierenden – als gleich schwierig empfunden wird.

Logistische Regressionsanalyse

Logistische Regressionsanalysen werden dann angewendet, wenn eine binäre abhängige Variable, das heisst eine Variable mit nur zwei Ausprägungen, vorliegt (z.B. bestanden/nicht bestanden). Im Fokus dieser Methode steht das Verhältnis der Eintrittswahrscheinlichkeit eines bestimmten Ereignisses (z.B. Prüfung bestanden) zur Gegenwahrscheinlichkeit desselben Ereignisses (z.B. Prüfung nicht bestanden). Dieses Wahrscheinlichkeitsverhältnis wird als Chance (odds) bezeichnet. Im Jahr 2017 bestanden beispielsweise 54 Prozent der Kandidatinnen und Kandidaten die Aufnahmeprüfung in einen beruflichen Maturitätslehrgang BM I. Die Gegenwahrscheinlichkeit, das heisst die Wahrscheinlichkeit, die Prüfung nicht zu bestehen, betrug somit 46 Prozent.

Mit der logistischen Regressionsanalyse kann darüber hinaus überprüft werden, wie gut der Prüfungsentscheid durch einzelne Variablen (z.B. einzelne Prüfungsnoten) vorausgesagt werden kann. In Absatz 6.1 wird beispielsweise überprüft, wie gut der Prüfungsentscheid (bestanden/nicht bestanden) anhand der Ergebnisse im Check S3 erklärt werden kann. Die Regressionsanalyse zeigt auf, mit welcher Wahrscheinlichkeit eine Kandidatin oder ein Kandidat aufgrund der Ergebnisse im Check S3 der Gruppe «Prüfung bestanden» zugeordnet wird.

Diese Wahrscheinlichkeit kann wiederum in eine binäre Entscheidung übersetzt werden, indem ein Schwellenwert bestimmt wird, der die vorhergesagten Werte den Gruppen 0 («nicht bestanden») und 1 («bestanden») zuordnet. Wird der Schwellenwert beispielsweise bei 50 Prozent angesetzt, wird ein vorausgesagter Wert von 43 Prozent der Gruppe «nicht bestanden» zugeteilt. Wird der Schwellenwert aber bereits bei 40 Prozent angesetzt, wird ein vorausgesagter Wert von 43 Prozent der Gruppe «bestanden» zugeordnet. Die Wahl des Schwellenwertes beeinflusst die Trefferquote und die Art der Fehler, die bei der Zuordnung gemacht werden.

Wenn verschiedene logistische Regressionsmodelle miteinander verglichen werden, kann die Modellgüte anhand der erklärten Varianz berechnet werden. Der hier verwendete Koeffizient entspricht dem korrigierten R^2 nach McFadden, er korrigiert für die Anzahl Variablen, die in einem Modell enthalten sind. Dadurch können Modelle unterschiedlicher Länge miteinander verglichen werden. Je höher die erklärte Varianz, desto besser kann der Prüfungsentscheid durch das Modell abgebildet werden.

Sensitivität und Spezifität

Wenn beispielsweise das Bestehen einer Prüfung anhand der Ergebnisse im Check S3 vorhergesagt werden soll, dann interessiert die Trefferquote. Das Regressionsmodell erzielt einen Treffer, wenn es eine erfolgreiche Kandidatin oder einen erfolgreichen Kandidaten der Gruppe «bestanden» zuordnet oder wenn es eine nicht erfolgreiche Kandidatin oder einen nicht erfolgreichen Kandidaten der Gruppe «nicht bestanden» zuordnet. Diese Fälle werden in Tabelle 2 als «richtig positiv» und «richtig negativ» benannt. Jedes Modell macht aber auch Fehler. Es kann sein, dass es Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, nicht als solche erkennt («falsch negativ») oder aber, dass das Modell für eine Kandidatin einen positiven Prüfungsentscheid vorhersagt, obwohl diese die Aufnahmeprüfung nicht bestanden hat («falsch positiv»). Aus diesen vier Gruppen – richtig positiv, richtig negativ, falsch positiv und falsch negativ – lassen sich nun Masse der Güte eines Modells berechnen. Von Interesse sind hier zwei Kennzahlen, nämlich Sensitivität und Spezifität. Die *Sensitivität* beschreibt den Anteil jener Kandidatinnen und Kandidaten, der die Aufnahmeprüfung bestanden hat und der vom Modell als «bestanden» erkannt wird. Die *Spezifität* beschreibt den Anteil jener Kandidatinnen und Kandidaten, der die Aufnahmeprüfung nicht bestanden hat und der vom Modell als «nicht bestanden» vorhergesagt wird.

Anhand dieser Gütemasse wird nun beispielhaft dargelegt, wie die Wahl des Schwellenwertes ebendiese beeinflusst. Aus Sicht der Chancengleichheit könnte gefordert werden, dass wirklich alle Kandidatinnen und Kandidaten, die das Potenzial haben, eine Berufsmatura zu erlangen, diese Chance auch wirklich bekommen. Die Sensitivität sollte in diesem Fall also 100 Prozent betragen, was mit einem Schwellenwert von 0 erreicht werden könnte. Mit einem Schwellenwert von 0 ordnet das Modell jeden Kandidaten und jede Kandidatin der Gruppe «bestanden» zu. In der Vierfeldertafel würden nur die Felder A und C gefüllt, alle erfolgreichen Kandidatinnen und Kandidaten würden also als solche erkannt werden. Auf der anderen Seite würden allerdings von den tatsächlich «nicht erfolgreichen» Kandidatinnen und Kandidaten keine erkannt werden. Die Spezifität läge folglich bei 0. Umgekehrt läge die Spezifität bei 1 und die Sensitivität bei 0, wenn der Schwellenwert bei 1 angesetzt würde.

Tabelle 2: Vierfeldertafel

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	A richtig positiv	B falsch negativ	Sensitivität A / (A + B)
nicht bestanden	C falsch positiv	D richtig negativ	Spezifität D / (C + D)

Für die Wahl des Schwellenwertes werden Sensitivität und Spezifität also gegeneinander abgewogen. Je tiefer der Schwellenwert angesetzt wird, desto höher ist die Sensitivität und desto tiefer ist die Spezifität. Für diese Analyse wird der Schwellenwert jeweils so angesetzt, dass die Gesamttref-

ferquote maximiert wird. Das Modell soll die Entscheidung der Aufnahmeprüfung also möglichst genau nachbilden. Dadurch werden Sensitivität und Spezifität gleich hoch gewichtet. Unterschiedliche Modelle können auch anhand ihrer Sensitivität und Spezifität miteinander verglichen werden.

6 Ergebnisse

6.1 Prüfungsinhalte und Vorkenntnisse

Zur Beurteilung, ob die Prüfungsinhalte den Vorkenntnissen der Kandidatinnen und Kandidaten entsprechen, wurde überprüft, welche Bedeutung die fachlichen Kompetenzen am Ende der obligatorischen Schulzeit für das Bestehen einer Aufnahmeprüfung in einen beruflichen Maturitätslehrgang haben. Dazu wurden die Ergebnisse im Check S3 von Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, mit denjenigen von Kandidatinnen und Kandidaten verglichen, die die Aufnahmeprüfung nicht bestanden haben. Für diesen Vergleich konnten von 79 Prozent der Kandidatinnen und Kandidaten einer Aufnahmeprüfung in die BMS I (N = 425) und von 12 Prozent der Kandidatinnen und Kandidaten einer Aufnahmeprüfung in die BMS II (N = 42) die Check-S3-Ergebnisse genutzt werden. Der Vergleich wird zuerst anhand der Mittelwerte der beiden Gruppen und anhand der Verteilung der Check-Ergebnisse dargestellt. Danach wird die Bedeutung der fachlichen Kompetenzen für das Bestehen der Aufnahmeprüfung ermittelt und beschrieben.

Mittelwerte

Tabelle 3 enthält die Gruppenmittelwerte von Kandidatinnen und Kandidaten für die BM I im Check S3 nach dem Prüfungsentscheid (bestanden/nicht bestanden). Die Ergebnisse im Check S3 werden auf einer Skala von 0 bis 1200 Punkten abgebildet. Die Unterschiede zwischen den Mittelwerten der beiden Gruppen sind für die vier Fächer statistisch signifikant. Kandidatinnen und Kandidaten, die die Aufnahmeprüfung in die BM I bestanden haben, erreichen im Check S3 bessere Ergebnisse als Kandidatinnen und Kandidaten, die die Aufnahmeprüfung nicht bestanden haben. Zur Beurteilung der Bedeutung der Unterschiede zwischen den beiden Gruppen sind in Tabelle 3 auch die Effektstärken angegeben. Die Unterschiede sind in Mathematik als gross (d = .84), in Deutsch (d = .69) und Französisch (d = .47) als mittel und in Englisch als klein (d = .27) einzustufen.

Tabelle 3: Mittelwerte der Kandidatinnen und Kandidaten für die BM I im Check S3 nach Prüfungsentscheid

	Prüfungsentscheid			
	nicht bestanden	bestanden	t	d
Deutsch	842	865	–7.08***	.69
Englisch	890	902	–2.79**	.27
Französisch	812	830	–4.69***	.47
Mathematik	851	880	–8.60***	.84

* p < .05, ** p < .01, *** p < .001

Tabelle 4 enthält die Gruppenmittelwerte der Kandidatinnen und Kandidaten für die BM II im Check S3 2016 nach Prüfungsentscheid (bestanden/nicht bestanden). Die Unterschiede zwischen den Mittelwerten sind lediglich in den Fächern Französisch und Mathematik statistisch signifikant. Kandidatinnen und Kandidaten, die die Aufnahmeprüfung in die BM II bestanden haben, erreichen im Check S3 in Französisch und Mathematik bessere Ergebnisse als Kandidatinnen und Kandidaten, die die Aufnahmeprüfung nicht bestanden haben. Die Unterschiede zwischen den beiden Gruppen sind sowohl in Französisch ($d = 1.02$) als auch in Mathematik ($d = .81$) als gross einzustufen.

Tabelle 4: Mittelwerte der Kandidatinnen und Kandidaten für die BM II im Check S3 nach Prüfungsentscheid

	Prüfungsentscheid			
	nicht bestanden	bestanden	*t*	*d*
Deutsch	838	857	–1.57	.51
Englisch	873	894	–1.40	.46
Französisch	787	836	–2.90**	1.02
Mathematik	838	871	–2.55*	.81

* $p < .05$, ** $p < .01$, *** $p < .001$

Zur Einordnung der Ergebnisse innerhalb der gesamten Schülerschaft enthält Tabelle 5 die Mittelwerte im Check S3 aller Schülerinnen und Schüler im Kanton Aargau für die Jahre 2016 und 2019 nach Schultyp (Realschule, Sekundarschule und Bezirksschule). Für die BM I dienen die Ergebnisse im Check S3 2019 und für die BM II die Ergebnisse im Check S3 2016 als Vergleichsgrundlage.

Die Check-Ergebnisse der Kandidatinnen und Kandidaten für die BM I liegen zwischen den Mittelwerten der Sekundarschule und der Bezirksschule (vgl. Tabelle 5). Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, erreichen im Durchschnitt rund 10 bis 20 Punkte weniger als die Schülerinnen und Schüler der Bezirksschule. Kandidatinnen und Kandidaten, die die Prüfung nicht bestanden haben, erreichen im Durchschnitt rund 5 bis 15 Punkte mehr als die Schülerinnen und Schüler der Sekundarschule.

Tabelle 5: Mittelwerte der Schülerinnen und Schüler des Kantons Aargau im Check S3 nach Schultyp

	Durchführungsjahr					
	2016			2019		
	Real	Sek	Bez	Real	Sek	Bez
Deutsch	776	834	896	783	832	884
Englisch	817	865	921	837	881	925
Französisch	747	798	859	771	809	852
Mathematik	784	836	904	787	837	891

Auch die Check-Ergebnisse der Kandidatinnen und Kandidaten für die BM II bewegen sich grösstenteils zwischen den Mittelwerten der Sekundarschule und der Bezirksschule (vgl. Tabelle 5). Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, erreichen im Durchschnitt rund

20 bis 40 Punkte weniger als die Schülerinnen und Schüler der Bezirksschule. Die Mittelwerte der Kandidatinnen und Kandidaten, die die Prüfung nicht bestanden haben, liegen in Deutsch, Englisch und Mathematik leicht über, in Französisch leicht unter dem Mittelwert der Schülerinnen und Schüler der Sekundarschule.

Tabelle 6 enthält den Anteil an Schülerinnen und Schülern des Kantons Aargau, deren Punktzahl im Check S3 gleich hoch oder höher ist als der Mittelwert der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung in einen beruflichen Maturitätslehrgang bestanden haben. Dieser Anteil wird für die drei Schultypen separat ausgewiesen.

Zwischen 57 Prozent (Englisch) und 71 Prozent (Französisch) der Bezirksschülerinnen und -schüler erreichen eine mindestens so hohe Punktzahl wie die Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für die BM I bestanden haben. Die Anteile der Schülerinnen und Schüler der Sekundar- und Realschule liegen deutlich tiefer. Zwischen 15 Prozent (Deutsch) und 25 Prozent (Französisch) der Sekundarschülerinnen und -schüler erreichen eine mindestens so hohe Punktzahl wie die Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für die BM I bestanden haben. Bei den Realschülerinnen und -schülern bewegen sich diese Anteile zwischen 1 Prozent (Deutsch) und 5 Prozent (Englisch und Französisch).

Für die BM II liegen diese Anteile tendenziell etwas höher. Zwischen 68 Prozent (Französisch) und 87 Prozent (Deutsch) der Bezirksschülerinnen und -schüler erreichen eine mindestens so hohe Punktzahl wie die Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für die BM II bestanden haben. Die Anteile der Schülerinnen und Schüler der Sekundar- und Realschule liegen wiederum deutlich tiefer. Zwischen 18 Prozent (Französisch) und 26 Prozent (Deutsch) der Sekundarschülerinnen und -schüler erreichen eine mindestens so hohe Punktzahl wie die Kandidatinnen und Kandidaten, die die Aufnahmeprüfung in die BM II bestanden haben. Bei den Realschülerinnen und -schülern bewegen sich diese Anteile zwischen 1 Prozent (Deutsch) und 8 Prozent (Englisch).

Tabelle 6: Anteil an Schülerinnen und Schülern, deren Punktzahl im Check S3 gleich oder höher ist als der Mittelwert der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, nach Schultyp

	beruflicher Maturitätslehrgang					
	BM I[1]			BM II[2]		
	Real	Sek	Bez	Real	Sek	Bez
Deutsch	1%	15%	70%	1%	26%	87%
Englisch	5%	21%	57%	8%	25%	76%
Französisch	5%	25%	71%	2%	18%	68%
Mathematik	2%	14%	61%	2%	20%	78%

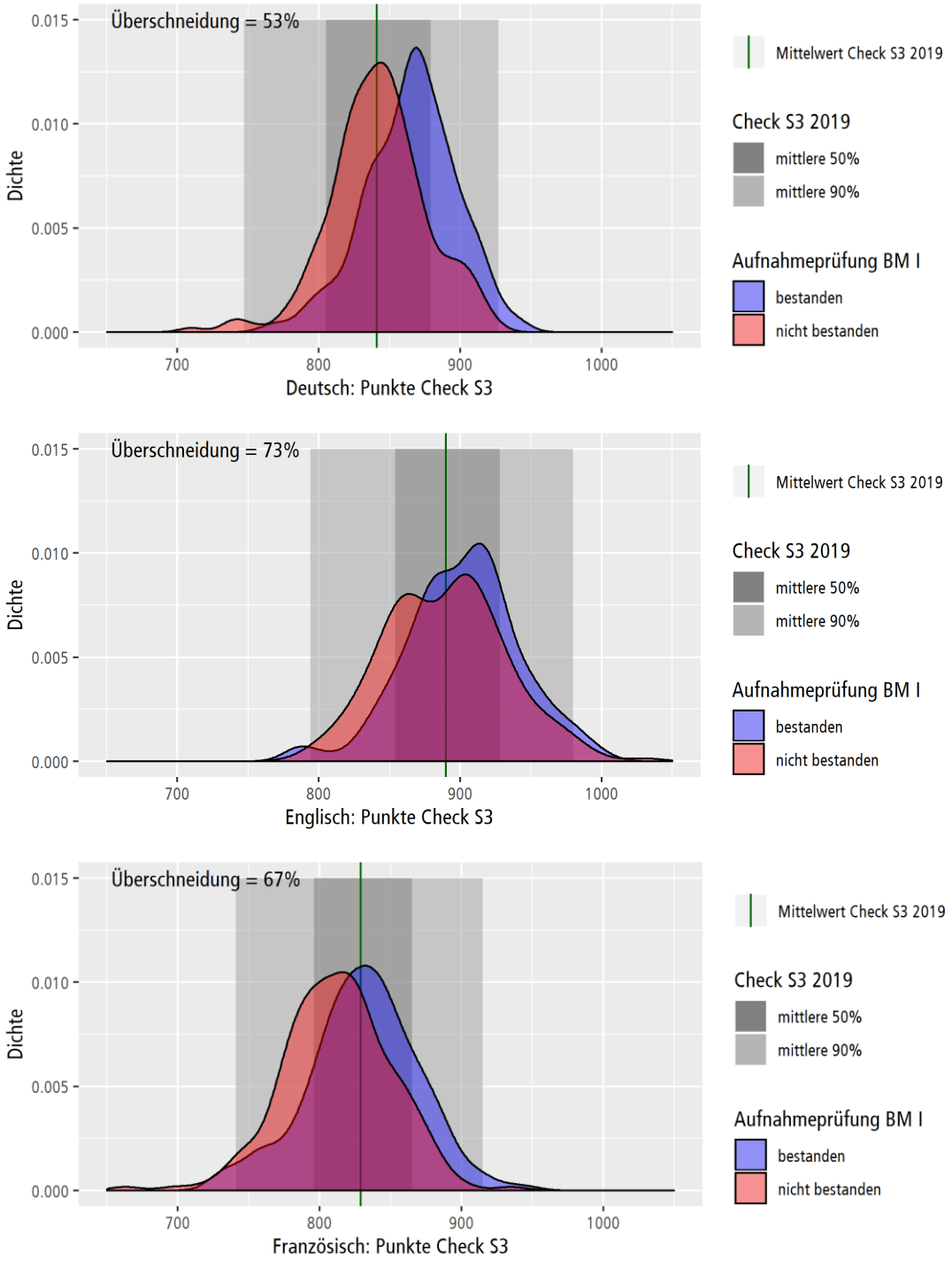
[1] Für die BM I werden die Checks S3 der Jahre 2016 bis 2019 berücksichtigt.

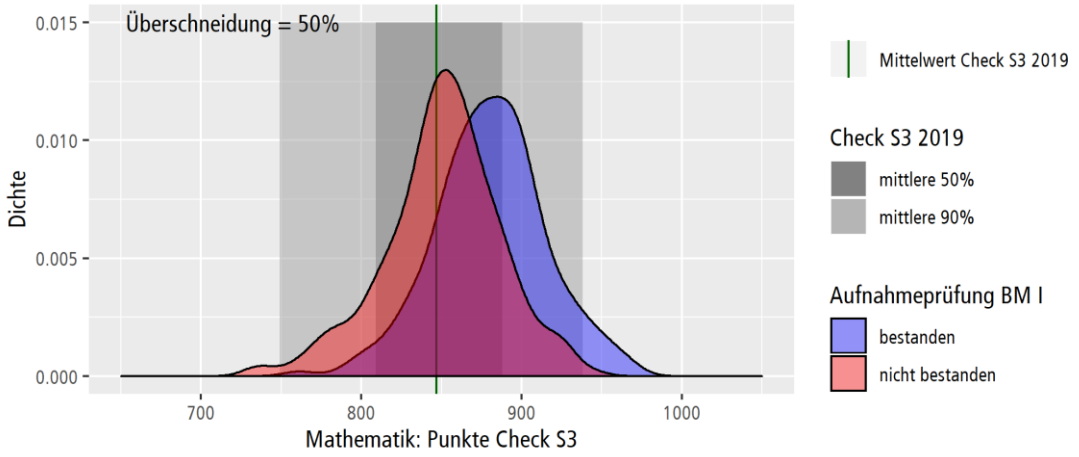
[2] Für die BM II wird nur der Check S3 2016 berücksichtigt.

Verteilungen

In Abbildung 1 sind die Verteilungen der Punktzahlen der Kandidatinnen und Kandidaten für die BM I nach Prüfungsentscheid dargestellt (blau = bestanden, rot = nicht bestanden). Zur Einordnung der Ergebnisse innerhalb der gesamten Schülerschaft sind die Bereiche der mittleren 50 Pro-

Abbildung 1: Verteilungen der Punktzahlen der Kandidatinnen und Kandidaten für die BM I im Check S3 nach Prüfungsentscheid





zent und der mittleren 90 Prozent der Punktzahlen aller Schülerinnen und Schüler des Kantons Aargau im Check S3 2019 grau hinterlegt. Die grüne Senkrechte illustriert den Mittelwert der gesamten Schülerschaft im Kanton Aargau. Zudem ist oben links angegeben, wie stark sich die Kurven der Kandidatinnen und Kandidaten nach Prüfungsentscheid (bestanden/nicht bestanden) überschneiden. Die Prozentangabe entspricht dem Anteil an Prüfungsergebnissen, der im Überschneidungsbereich liegt. Sie kann als Effektstärke interpretiert werden. Je kleiner die Prozentzahl ist, desto deutlicher unterscheiden sich die beiden Gruppen in den Punktzahlen.

Die Verteilungen zeigen, dass die erfolgreichen Kandidatinnen und Kandidaten mehrheitlich vergleichsweise sehr gute Ergebnisse im Check S3 erreichen. Die Mittelwerte der erfolgreichen Kandidatinnen und Kandidaten liegen in Deutsch, Englisch sowie in Mathematik deutlich und in Französisch leicht über den Mittelwerten aller Schülerinnen und Schüler, die im Kanton Aargau am Check S3 2019 teilgenommen haben. In Mathematik erreicht rund die Hälfte der erfolgreichen Kandidatinnen und Kandidaten im Check S3 Ergebnisse, die innerhalb der besten 25 Prozent aller Ergebnisse im Check S3 im Kanton Aargau liegen. Die Mittelwerte der nicht erfolgreichen Kandidatinnen und Kandidaten liegen in Deutsch und Englisch nahe beim Mittelwert im Kanton Aargau. In Französisch liegt der Mittelwert jener Kandidatinnen und Kandidaten, die die Aufnahmeprüfung nicht bestanden haben, unterhalb des kantonalen Durchschnitts. Im Gegensatz dazu liegt in Mathematik auch der Mittelwert nicht erfolgreicher Kandidatinnen und Kandidaten leicht über dem Mittelwert im Kanton Aargau.

Am stärksten überschneiden sich die Verteilungen der Punktzahlen nach Prüfungsentscheid in den Fremdsprachen Englisch (73 Prozent) und Französisch (67 Prozent). In Deutsch (53 Prozent) und in Mathematik (50 Prozent) sind die Überschneidungen deutlich geringer. Diese Reihenfolge zeigt sich auch in den Effektstärken (vgl. Tabelle 3). Das Bestehen der Aufnahmeprüfung für die BM I hängt am stärksten mit den Vorkenntnissen in Mathematik und am schwächsten mit den Vorkenntnissen in Englisch zusammen.

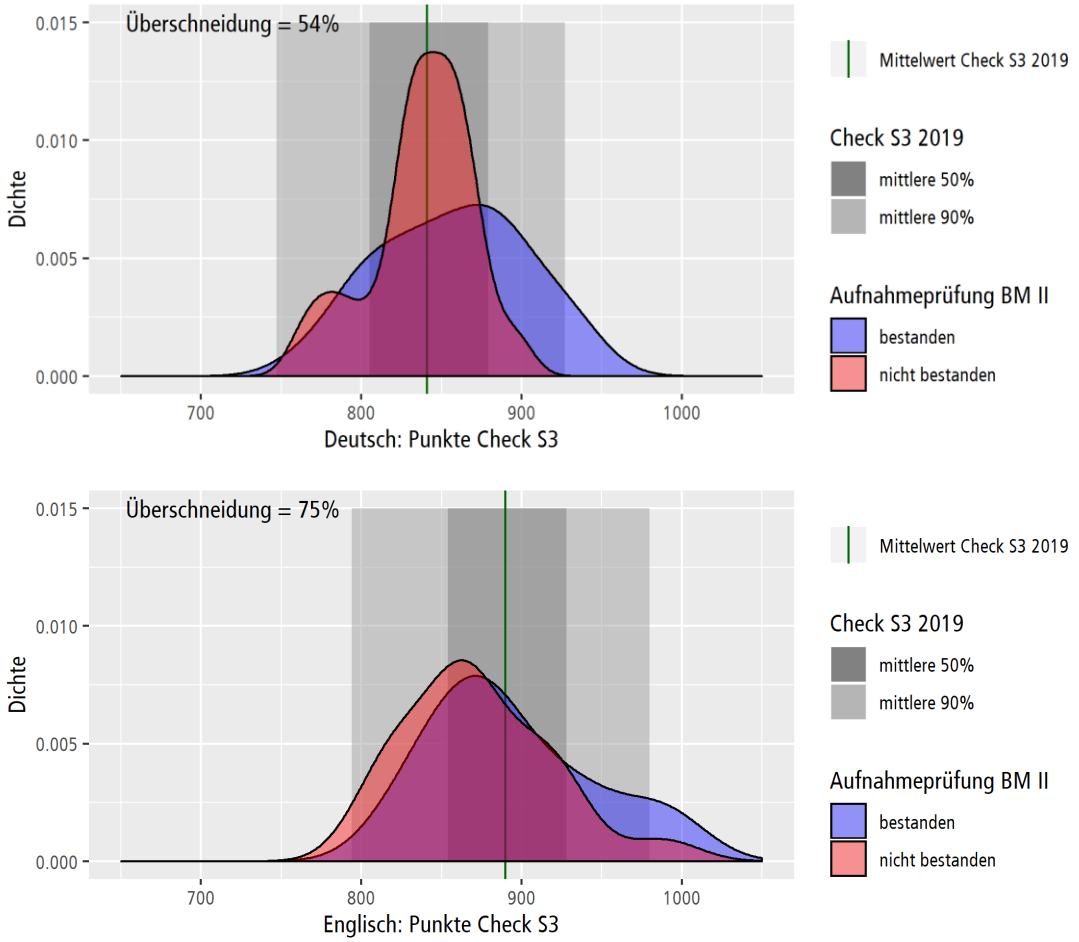
In Abbildung 2 sind die Verteilungen der Punktzahlen der Kandidatinnen und Kandidaten für die BM II nach Prüfungsentscheid dargestellt (blau = bestanden, rot = nicht bestanden). Zudem sind zur Einordnung der Ergebnisse die gleichen Parameter wie in Abbildung 1 aufgeführt.[2] Die Vertei-

[2] Die Teilnahme am Check S3 2016 war im Kanton Aargau freiwillig. Der Anteil an Bezirksschülerinnen und Bezirksschülern war 2016 vergleichsweise niedrig. Die kantonalen Ergebnisse im Check S3 2016 bilden folglich nicht die

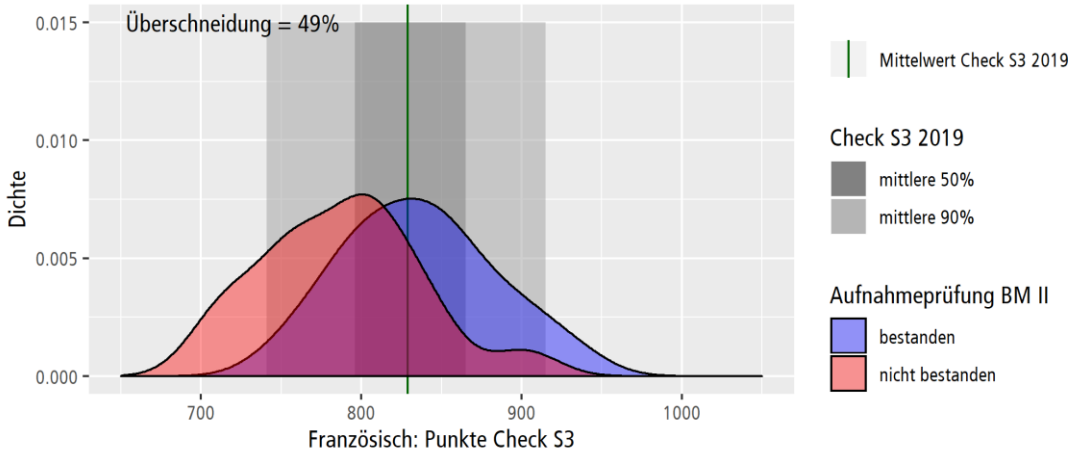
lungen zeigen, dass Kandidatinnen und Kandidaten, die die Prüfung für die BM II bestanden haben, in Mathematik mehrheitlich sehr gute Ergebnisse im Check S3 erreichen. Ihr Mittelwert liegt deutlich über dem Gesamtmittelwert des Kantons Aargau, rund die Hälfte erreicht in Mathematik eine Punktzahl, die innerhalb der besten 25 Prozent aller Schülerinnen und Schüler des Kantons Aargau liegt. In Deutsch schneiden Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für die BM II bestanden haben, leicht besser ab als der Kanton Aargau, in Englisch und Französisch etwa gleich gut. Die Mittelwerte der Kandidatinnen und Kandidaten, die die Prüfung nicht bestanden haben, liegen in Deutsch nahe beim Gesamtmittelwert des Kantons Aargau, in den anderen Fächern darunter.

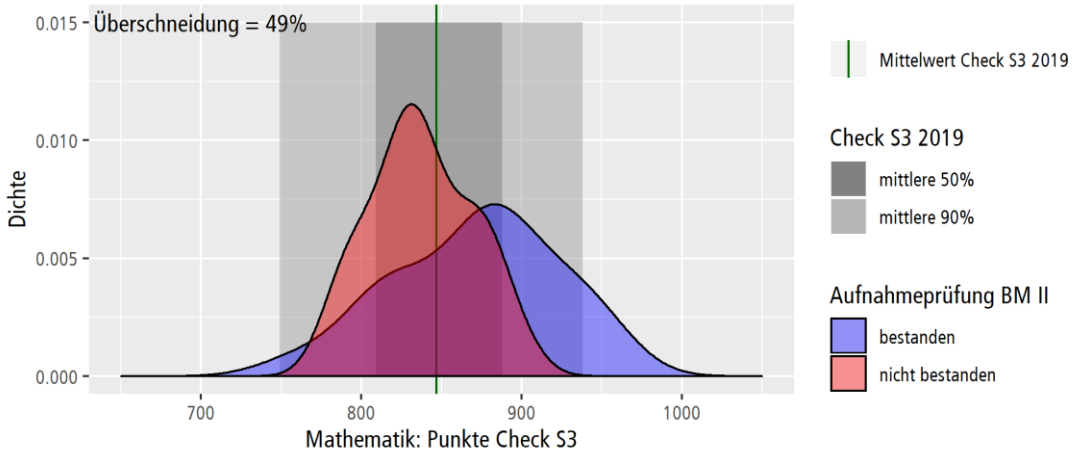
Am stärksten überschneiden sich die Verteilungen der Punktzahlen nach Prüfungsentscheid in Englisch (75 Prozent). In Deutsch (54 Prozent) sowie in Französisch und Mathematik (je 49 Prozent) sind die Überschneidungen deutlich geringer. Diese Reihenfolge zeigt sich in den Effektstärken nur bedingt (vgl. Tabelle 4). Das Bestehen der Aufnahmeprüfung für die BM II hängt am stärksten mit den Vorkenntnissen in Französisch und Mathematik und am schwächsten mit den Vorkenntnissen in Englisch zusammen.

Abbildung 2: Verteilungen der Punktzahlen der Kandidatinnen und Kandidaten für die BM II im Check S3 nach Prüfungsentscheid



Gesamtpopulation des Kantons Aargau ab. Deswegen werden auch hier die Ergebnisse des Check S3 2019 als Vergleichshintergrund herbeigezogen.





Die Kandidatinnen und Kandidaten für die BM I schneiden im Check S3 – mit Ausnahme von Französisch – etwas besser ab als die Kandidatinnen und Kandidaten für die BM II. Bei diesem Vergleich ist jedoch zu beachten, dass die Check-Ergebnisse der Kandidatinnen und Kandidaten für die BM II aus dem Jahr 2016 stammen. In diesem Jahr war die Teilnahme am Check S3 im Kanton Aargau freiwillig. Dadurch können Verzerrungen durch Selektionseffekte nicht ausgeschlossen werden.

Bedeutung fachlicher Kompetenzen für den Prüfungserfolg

Für die Überprüfung der Bedeutung der Vorkenntnisse für das Bestehen der Aufnahmeprüfung wurden die Check-Ergebnisse in den einzelnen Kompetenzbereichen der vier Fächer genutzt. Mit einer Regressionsanalyse wurde der Zusammenhang zwischen den Punktzahlen und dem Prüfungsentscheid (bestanden/nicht bestanden) berechnet.[3] Aufgrund der geringen Anzahl Check-Ergebnisse von Kandidatinnen und Kandidaten für die BM II wurde die Regressionsanalyse nur mit den Punktzahlen der Kandidatinnen und Kandidaten für die BM I berechnet.

Mit der Regressionsanalyse lässt sich nachweisen, wie gut der Prüfungsentscheid durch die Ergebnisse in den Kompetenzbereichen vorhergesagt werden kann. Je mehr Kompetenzbereiche in das Modell miteinbezogen werden, desto genauer kann der Prüfungsentscheid tendenziell nachmodel-

[3] Die Kompetenzbereiche *Englisch Schreiben* und *Französisch Schreiben* werden im Check S3 nicht gleichzeitig, sondern über die Jahre alternierend geprüft. Deswegen wurden sie für die logistischen Regressionen nicht berücksichtigt.

liert werden. Deswegen werden zunächst alle Kompetenzbereiche in die Analyse einbezogen.[4] In einem zweiten Schritt wird überprüft, wie gut sich der Prüfungsentscheid mit jenen Kompetenzbereichen vorhersagen lässt, die sich in der ersten Analyse als statistisch signifikant erwiesen haben.

Die Ergebnisse der Regressionsanalyse mit allen Kompetenzbereichen (Modell 1) sind in Tabelle 7 enthalten. Die Punktzahlen in den Kompetenzbereichen *Deutsch Schreiben*, *Deutsch Sprache im Fokus*, *Mathematik Zahl und Variable* sowie *Mathematik Form und Raum* hängen statistisch signifikant mit dem Prüfungsentscheid zusammen. Die Punktzahlen in den übrigen Kompetenzbereichen hängen nur zufällig mit dem Prüfungsentscheid zusammen, sind also dafür kaum von Bedeutung.

Ebenfalls in Tabelle 7 enthalten sind die Ergebnisse der Regressionsanalyse mit den vier statistisch signifikanten Kompetenzbereichen *Deutsch Schreiben*, *Deutsch Sprache im Fokus*, *Mathematik Zahl und Variable* sowie *Mathematik Form und Raum* (Modell 2). Auch in diesem Modell sind die Zusammenhänge zwischen den vier Kompetenzbereichen und dem Prüfungsentscheid statistisch signifikant. Zudem bewegen sich die Koeffizienten beider Modelle in derselben Grössenordnung. Das lässt darauf schliessen, dass die Zusammenhänge robust sind.

Tabelle 7: Zusammenhang zwischen dem Prüfungsentscheid (bestanden/nicht bestanden) und den Punktzahlen in den Kompetenzbereichen des Check S3

	Modell 1	Modell 2
Konstante	–44.750***	–40.985***
Deutsch		
Lesen	–.005	
Sprache im Fokus	.009**	.009**
Schreiben	.013***	.012***
Englisch		
Hören	–.001	
Lesen	.004	
Französisch		
Hören	.004	
Lesen	.000	
Mathematik		
Zahl und Variable	.013**	.014***
Form und Raum	.013**	.012***
Grössen, Funktionen, Daten und Zufall	.002	
McFaddens korrigiertes R^2	.19	.20

* $p < .05$, ** $p < .01$, *** $p < .001$

[4] Mit Ausnahme von Natur & Technik.

Der Vergleich der erklärten Varianz (R^2) zeigt, dass der Prüfungsentscheid mit beiden Modellen nahezu gleich gut vorhergesagt werden kann. Das bedeutet, dass vor allem die Vorkenntnisse in Deutsch und Mathematik für das Bestehen der Aufnahmeprüfung eine Rolle spielen, während die Vorkenntnisse in den Fremdsprachen kaum eine Rolle spielen bzw. nicht von Bedeutung sind.

Spezifität und Sensitivität

Die Vorhersage des Prüfungsentscheids durch die Vorkenntnisse der Kandidatinnen und Kandidaten lässt sich durch die Berechnung von Spezifität und Sensitivität konkretisieren. Tabelle 8 enthält die Angaben zur Trefferquote. Werden sämtliche Kompetenzbereiche einbezogen (Modell 1), dann lässt sich mit den Check-Ergebnissen von 72 Prozent der Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, und von 73 Prozent der Kandidatinnen und Kandidaten, die die Prüfung nicht bestanden haben, der Prüfungsentscheid richtig vorhersagen (Sensitivität = .72; Spezifität = .73). Für 281 (rund 73 Prozent) der insgesamt 388 Kandidatinnen und Kandidaten kann der Prüfungsentscheid aufgrund der Vorkenntnisse folglich korrekt vorhergesagt werden. 107 Kandidatinnen und Kandidaten (rund 27 Prozent) werden falsch zugeordnet. 13 Prozent, die die Prüfung bestanden haben, hätten aufgrund der Check-Ergebnisse die Prüfung nicht bestanden, rund 14 Prozent, die die Prüfung nicht bestanden haben, hätten die Prüfung aufgrund der Check-Ergebnisse bestanden.

Tabelle 8: Vierfeldertafel Modell 1 mit sämtlichen Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	141	52	Sensitivität = .72
nicht bestanden	55	140	Spezifität = .73

Tabelle 9 enthält die Trefferquote, wenn nur vier der insgesamt zehn Kompetenzbereiche einbezogen werden (Modell 2). Mit den Punktzahlen in den vier Kompetenzbereichen lässt sich von jeweils 72 Prozent der Kandidatinnen und Kandidaten, die die Prüfung bestanden haben bzw. die die Prüfung nicht bestanden haben, der Prüfungsentscheid richtig vorhersagen (Sensitivität = .72; Spezifität = .72). Für 279 (rund 72 Prozent) der insgesamt 388 Kandidatinnen und Kandidaten kann der Prüfungsentscheid aufgrund der Vorkenntnisse folglich korrekt vorhergesagt werden. 109 Kandidatinnen und Kandidaten (rund 28 Prozent) werden falsch zugeordnet. Rund 14 Prozent, die die Prüfung bestanden haben, hätten aufgrund der Check-Ergebnisse die Prüfung nicht bestanden und weitere rund 14 Prozent, die die Prüfung nicht bestanden haben, hätten die Prüfung aufgrund der Check-Ergebnisse bestanden.

Tabelle 9: Vierfeldertafel Modell 2 mit vier Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	139	54	Sensitivität = .72
nicht bestanden	55	140	Spezifität = .72

Die Ergebnisse zeigen deutlich, dass die Vorkenntnisse in den Fremdsprachen für den Prüfungsentscheid nahezu irrelevant sind. Nur zwei der insgesamt 388 Kandidatinnen und Kandidaten werden

anders zugeordnet, wenn zusätzlich zu den Vorkenntnissen in Deutsch und Mathematik auch die Vorkenntnisse in Englisch und Französisch berücksichtigt werden.

Fazit

Die fachlichen Vorkenntnisse sind für den Prüfungsentscheid von grosser Bedeutung. Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für eine Berufsfachschule zur Erreichung einer Berufsmaturität bestanden haben, verfügen im Durchschnitt über deutlich bessere Vorkenntnisse als solche, die die Prüfung nicht bestanden haben. Besonders wichtig erweisen sich die Vorkenntnisse in Mathematik und Deutsch für die BM I sowie in Französisch und Mathematik für die BM II, während die Vorkenntnisse in Englisch für den Prüfungsentscheid kaum von Bedeutung sind.

Die Prüfungsanforderungen sind hoch. Ein durchschnittliches Ergebnis im Check S3 am Ende der obligatorischen Schulzeit reicht im Schnitt nicht aus, um die Aufnahmeprüfung für die BM I zu bestehen. Die Vorkenntnisse der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, sind am Ende der obligatorischen Schulzeit im Durchschnitt nur leicht geringer als jene von Schülerinnen und Schülern der Bezirksschule. Insgesamt sind die Vorkenntnisse der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für die BM I bestanden haben, leicht höher als die Vorkenntnisse der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung für die BM II bestanden haben. Weil die Teilnahme am Check S3 2016 freiwillig und die untersuchte Stichprobe der BM II relativ klein war ($N = 42$), können Verzerrungen durch Selektionseffekte allerdings nicht ausgeschlossen werden. Ob die Ergebnisse der Kandidatinnen und Kandidaten der BM II auch dann tiefer liegen würden, wenn die Stichprobe die Population vollständig abdecken würde, ist nicht klar.

Der Zusammenhang zwischen den Vorkenntnissen und dem Prüfungsentscheid für die BM I ist in Deutsch und Mathematik besonders eng. Der Prüfungsentscheid kann am besten mit je zwei Kompetenzbereichen in Deutsch (*Sprache im Fokus* und *Schreiben*) und Mathematik (*Zahl und Variable* sowie *Form und Raum*) vorhergesagt werden. Die Kompetenzbereiche *Lesen* sowie *Grössen, Funktionen, Daten und Zufall* liefern keine zusätzlichen Informationen, um den Prüfungsentscheid zu erklären. Ebenso hängen die Kompetenzbereiche *Hören* und *Lesen* in den Fremdsprachen Englisch und Französisch nur zufällig mit dem Prüfungsentscheid zusammen. Dass die Vorkenntnisse in Englisch und Französisch nahezu irrelevant sind, zeigt sich in der Vorhersage des Prüfungsentscheids anschaulich: Nur zwei der insgesamt 388 Kandidatinnen und Kandidaten für die BM I werden dem Prüfungsentscheid anders zugeordnet, wenn zusätzlich zu den Vorkenntnissen in Deutsch und Mathematik auch die Vorkenntnisse in Englisch und Französisch berücksichtigt werden.

Dementsprechend ist die Passung zwischen den Vorkenntnissen und den Inhalten der Aufnahmeprüfung für die BM I in Mathematik und Deutsch am grössten. Kandidatinnen und Kandidaten, die in diesen Fachbereichen bessere Vorkenntnisse mitbringen, haben eine höhere Chance, die Aufnahmeprüfung zu bestehen. Darüber hinaus erhöhen bessere Fremdsprachenkenntnisse die Chance kaum mehr, die Aufnahmeprüfung zu bestehen. Gemessen an den Check-Ergebnissen als externem Massstab ist die Passung zwischen den Vorkenntnissen und den Inhalten der Aufnahmeprüfung für die BM I in Englisch und Französisch nicht gegeben. Eine abschliessende Beurteilung zur Passung zwischen den Prüfungsinhalten und den Vorkenntnissen ist für die BM II aufgrund der Datenlage – die Stichprobe ist dafür zu klein – nicht möglich.

Es ist davon auszugehen, dass die geringe Passung der Prüfungsinhalte mit den Vorkenntnissen der Kandidatinnen und Kandidaten in den Fremdsprachen unter anderem mit der unterschiedlichen Gewichtung der einzelnen Kompetenzbereiche im Check S3 bzw. in den Aufnahmeprüfungen zusammenhängt. Während im Check S3 in den Fremdsprachen die Lese- und Hörkompetenzen

ausführlich geprüft werden, werden diese in den Aufnahmeprüfungen nur marginal bzw. gar nicht geprüft. Im Gegensatz dazu werden Wortschatz und Grammatik im Check S3 anwendungsorientiert beim Schreiben eines Textes erfasst, während sie in den Aufnahmeprüfungen der BM I und der BM II ausführlich geprüft werden (vgl. Absatz 6.2).

6.2 Prüfungsinhalte und Lehrplaninhalte

Aufnahmeprüfung BM I und BM II Typ «andere»
Die Inhalte der Aufnahmeprüfungen für die BM I und für die BM II Typ «andere» basieren auf dem Lehrplan der Abschlussklassen der Sekundarschule des Kantons Aargau. Für die Inhalte der Aufnahmeprüfung für die BM II Typ «Wirtschaft» gilt der Rahmenlehrplan für die Berufsmaturität des SBFI. Der Rahmenlehrplan enthält Stundentafeln zu den einzelnen Fachbereichen, jedoch keine konkreten Angaben zu Inhalten oder Kompetenzbeschreibungen. Aus diesem Grund bezieht sich der Vergleich der Prüfungsinhalte und der Lehrplaninhalte ausschliesslich auf die Aufnahmeprüfung für die BM I und für die BM II Typ «andere».

Gegenstand der Inhaltsanalyse sind Aufnahmeprüfungen, die ausschliesslich auf dem Lehrplan der Abschlussklassen der Sekundarschule des Kantons Aargau basieren. Ab dem Schuljahr 2022/23 gilt für die Abschlussklassen der Sekundarschule der Lehrplan 21. Für die Überprüfung der Passung der Prüfungsinhalte mit den Lehrplaninhalten wurden die Prüfungsinhalte für die BM I und für die BM II Typ «andere» der Jahre 2017 bis 2019 deshalb a) mit den Inhalten des Lehrplans für die Abschlussklassen der Sekundarschule des Kantons Aargau, b) mit den Inhalten des Lehrplan 21 sowie c) mit den Inhalten des Check S3 verglichen. Die Leistungstests des Check S3 werden auf der Grundlage des Lehrplan 21 entwickelt.

Deutsch

Die Deutschprüfung für die BM I umfasst zwei Teile: *Textverständnis und Sprachbetrachtung* sowie *Verfassen eines Textes*. Tabelle 10 enthält die Inhalte der Deutschprüfung mit den passenden Inhalten des Lehrplans der Sekundarschule, des Lehrplan 21 und des Check S3. Die erste Spalte enthält jeweils den Prüfungsinhalt (Titel, Aufgabenformat) und in Klammern die Anzahl Punkte, die für die entsprechenden Aufgaben vergeben wurden. Diese kann über die Jahre variieren. In der zweiten und dritten Spalte werden die Referenzen zum Lehrplan der Sekundarschule des Kantons Aargau sowie zum Lehrplan 21 festgehalten. In der vierten Spalte wird mit einem «x» angegeben, ob die Inhalte im Check S3 vorkommen.

In der Deutschprüfung für die BM I können insgesamt 100 Punkte erreicht werden. Für den Prüfungsteil *Textverständnis und Sprachbetrachtung* werden jeweils 40, für den Prüfungsteil *Verfassen eines Textes* jeweils 60 Punkte vergeben. Im Prüfungsteil *Textverständnis und Sprachbetrachtung* werden für das Textverständnis zwischen 6 und 10 Punkte vergeben. 14 Punkte werden jeweils für Wortschatzaufgaben vergeben (Synonyme/Antonyme, Fremdwörter, Sprichwörter und Redewendungen). Die restlichen 16 bis 20 Punkte werden für Aufgaben zur Grammatik (inkl. Satzzeichensetzung) vergeben.

Im Lehrplan der Sekundarschule können sowohl für die Aufgaben im Bereich *Textverständnis und Sprachbetrachtung* (bzw. Grammatik) als auch für diejenigen im Bereich *Verfassen eines Textes* Referenzen nachgewiesen werden. Einzig für die Aufgaben zum Wortschatz finden sich im Lehrplan der Sekundarschule keine Inhalte. Im Lehrplan 21 kann hingegen auch zu den Wortschatzaufgaben (Antonyme/Synonyme sowie Fremdwörter) eine Referenz ausgewiesen werden. In diesem Bereich passen die Prüfungsinhalte zwar nicht zum aktuell massgebenden Lehrplan der Se-

kundarschule, jedoch – zumindest bedingt – zum Lehrplan 21, der in den nächsten Jahren auf der Sekundarstufe I eingeführt wird.

Tabelle 10: Vergleich der Prüfungsinhalte der Deutschprüfung mit den Inhalten der Lehrpläne und des Check S3

Aufnahmeprüfung	Lehrplan		Check S3
	Sekundarschule Kanton Aargau	Lehrplan 21	
Lesetext: Offene Fragen (3–6 Punkte)	Anspruchsvollere Texte für sich lesen und verstehen Gefühle, Stimmungen und Meinungen wahrnehmen, verschiedenen Fragen nachgehen Unterschiedliche Textsorten kennen Haltungen, Wertungen und Hintergründe von Personen erkennen (Lesen)	Die Schülerinnen und Schüler … können wesentliche Informationen aus diskontinuierlichen Sachtexten für den eigenen Wissensaufbau entnehmen, indem sie auch Zusammenhänge zwischen Text und Abbildungen herstellen (z.B. Artikel aus Internet, Anleitung). (D.2.B.1.h) können einen übersichtlich strukturierten Text als Ganzes verstehen sowie zentrale Elemente erkennen und mit der eigenen Lebenswelt in Verbindung bringen (z.B. Artikel aus Jugendzeitschrift). (D.2.B.1.h)	
Lesetext: Richtig-/Falsch-Aussagen (3–4 Punkte)			x
Synonyme und Antonyme (Gegenteile) (4–6 Punkte)		können mit Unterstützung die Bedeutung von unbekannten Wörtern aus dem Kontext oder mit geeigneten Hilfsmitteln (z.B. Wörterbuch, Sachbuch, Internet) erschliessen und differenzieren damit ihren rezeptiven Wortschatz aus. (D.2.B.1.h)	
Fremdwörter (4–6 Punkte)			
Redewendungen, Sprichwörter (3–6 Punkte)			
Satzzeichen (Kommas) (2–5 Punkte)	Einfache und zusammengesetzte Sätze erkennen (Grammatik) Zeichensetzung bei Aufzählungen, in der direkten Rede, bei Propositionen und Einschüben (Rechtschreibung 1./2. Klasse und 3./4. Klasse)	können folgende Rechtschreibregeln in dafür konstruierten Übungen anwenden: … Komma bei Aufzählungen, Komma zwischen übersichtlichen Verbgruppen. (D.5.E.1.d) … Komma bei infinitivischen Verbgruppen, bei Einschüben und Relativsätzen. (D.5.E.1.g)	x
Wortarten (2017 nur Pronomen) (4–7 Punkte)	Verb, Nomen, Adjektiv, Pronomen, Partikel nach formalen Kriterien ordnen und ihre Funktionen vergleichen (Grammatik)	können Nomen, Verb und Adjektiv mithilfe formaler Kriterien sowie Pronomen mithilfe der umfassenden Pronomenliste bestimmen sowie den Rest als Partikel benennen. (D.5.D.1.f)	x
Wortfamilien (0–3 Punkte)	Über Wörter, Wortwendungen, ihre Bedeutung und ihre Verwandtschaft nachdenken (Sprache erforschen)	können Wortstamm (Stamm-Morphem) bestimmen und Wortfamilien bilden. (D.5.D.1.c)	x

Aufnahme-prüfung	Lehrplan		Check S3
	Sekundarschule Kanton Aargau	Lehrplan 21	
Die vier Fälle (Satzglieder bestimmen) (3–5 Punkte)			
Zeitformen (3–7 Punkte)	Die vier Fälle mündlich und schriftlich möglichst korrekt gebrauchen (Grammatik) In Sätzen verbale Teile und Satzglieder erkennen (Grammatik)	können in typischen Fällen Nominativ, Akkusativ, Dativ und Genitiv mithilfe der Ersatzprobe bestimmen. (D.5.D.1.e) können mithilfe eines Verbenfächers Satzglieder bestimmen. (D.5.D.1.f)	x
Verfassen eines Textes (60 Punkte)	Die Zeitformen erkennen und mündlich und schriftlich anwenden (Grammatik)	können Präsens, Präteritum, Perfekt und Futur sowie Infinitiv und Personalform bestimmen. (D.5.D.1.f)	x
	Formen von adressatenbezogenem und zweckgebundenem Schreiben anwenden Mit Texten bewusst Wirkungen erzielen Zu verschiedenen Themen eine eigene Haltung oder einen eigenen Standpunkt ausdrücken und begründen (Texte schreiben)	kennen vielfältige Textmuster (z.B. Erzählung, Argumentation, Zeitungsbericht, Geschäftsbrief, Blog-Beitrag, Lernjournal, Flyer, Präsentationsfolien), um sie für das eigene Schreiben nutzen zu können. (D.4.B.1.h) können das eigene Repertoire an Strategien beim Planen ihrer Texte selbstständig, situationsangemessen und auf das Schreibziel ausgerichtet einsetzen. (D.4.C.1.h) können ihre Gedanken und Ideen im Text in eine verständliche und sinnvolle Abfolge bringen und eine gezielte Wirkung erzeugen. (D.4.D.1.e) können textstrukturierende Mittel (z.B. Titel, Absätze) und textverknüpfende Mittel (z.B. Pronomen, Partikel) beim Entwerfen gezielt setzen, um den Text klarer zu strukturieren. (D.4.D.1.f) können in Überarbeitungsprozessen Mittel zur Leserführung gezielt einsetzen, um den Text leserfreundlicher zu gestalten (z.B. Überleitung, Wiederaufnahme). (D.4.E.1.g) können selbstständig auf Papier oder am Computer ihre Texte sprachformal überarbeiten. (D.4.F.1.g)	x

Der Vergleich der Inhalte der Aufnahmeprüfung für die BM I mit den Inhalten des Check S3 zeigt, dass in den Bereichen *Verfassen eines Textes* und *Sprachbetrachtung* (Grammatik) ähnliche Inhalte erfasst werden. Dagegen wird im Check S3 das *Textverständnis* (Leseverstehen) ausführlicher erfasst als in der Prüfung für die BM I. Zudem wird im Check S3 bewusst auf die separate Erfassung der Wortschatzkompetenzen verzichtet, weil die Gefahr besteht, dass mit Wortschatzaufgaben weniger die Deutschkompetenz und vielmehr die soziale Herkunft erfasst wird (vgl. Barthel, 2019).

Die Deutschprüfung für die BM II Typ «andere» besteht nur aus dem Prüfungsteil *Verfassen eines längeren Textes*. Die Kandidatinnen und Kandidaten müssen jeweils eine Erörterung zu einem von drei vorgegebenen Themen schreiben. Die inhaltliche Passung mit den Lehrplänen und dem Check S3 ist Tabelle 10 (unterste Zeile) zu entnehmen. Für eine valide Beurteilung der Passung dieses Prüfungsteils sind allerdings nicht nur die Inhalte, sondern vor allem auch die Beurteilungskriterien beizuziehen. Tabelle 11 gibt einen Überblick über die in den Prüfungen für die BM I und für die BM II Typ «andere» angewendeten Beurteilungskriterien, dies im Vergleich zu den Beurteilungskriterien, die im Rahmen des Check S3 genutzt werden.

Tabelle 11: Vergleich der Beurteilungskriterien der Deutschprüfung mit den Beurteilungskriterien des Check S3

BM I		BM II Typ «andere»		Check S3	
Kriterien	Punkte	Kriterien	Punkte	Kriterien	Punkte
Inhalt/Aufbau (2017: Logik, Tiefe, Kreativität, Struktur, Kohärenz, Stringenz; 2018: Logik, Tiefe, Kreativität, Struktur, Kohärenz, Darstellung; 2019: Logischer Aufbau, Tiefe, Kreativität, Kohärenz, Darstellung)	25	Inhalt, Aufbau, Kohärenz	6	Inhalt: Auftragserfüllung und Aussagekraft (inkl. Kreativität)	9
		Darstellung, Gliederung	2	Textaufbau und Textzusammenhang: logischer Zusammenhang (Kohärenz)	3
				Textaufbau und Textzusammenhang: Textaufbau (äussere Gliederung – Abschnitte)	1
Stil/Sprache (Sprachliche Präzision, Wortschatz, Syntax, Kohäsion)	25	Wortwahl, Grammatik, Stil	6	Sprachangemessenheit, Schreibstil und Ästhetik (Satzbau, Wortwahl, Schreibstil, sprachliches Wagnis)	12
				Textaufbau und Textzusammenhang: Sprachlicher Zusammenhang (Kohäsion)	3
Formale Korrektheit (Grammatik, Orthografie, Interpunktion)	10	Rechtschreibung, Satzzeichen	6	Sprachrichtigkeit: Rechtschreibung	3
				Sprachrichtigkeit: Grammatik	3
				Sprachrichtigkeit: Satzbau (Korrektheit, Vollständigkeit)	3
				Sprachrichtigkeit: Satzzeichen	3
Total	60	Total	20	Total	40

Die Prüfung für die BM I unterscheidet drei übergreifende Kriterien, die Prüfung für die BM II Typ «andere» deren vier. Die Texte im Rahmen des Check S3 werden ebenfalls anhand von vier übergreifenden Kriterien beurteilt (das vollständige Kriterienraster zur Bewertung der Texte des Check S3 befindet sich im Anhang A). Die Unterschiede zwischen den drei Beurteilungsrastern sind ge-

ring. Eine relevante Abweichung lässt sich in der Prüfung für die BM II Typ «andere» nachweisen, bei der die Grammatik gemeinsam mit der Wortwahl und dem Stil ein übergreifendes Kriterium bildet, während die Grammatik in den Prüfungen für die BM I und im Check S3 zur formalen Korrektheit bzw. zur Sprachrichtigkeit gezählt wird.

Englisch

Die Englischprüfung für die BM I umfasst drei Teile: *Reading Comprehension and Vocabulary*, *Grammar and Structures* sowie *Writing*. Die Englischprüfung für die BM II Typ «andere» umfasst einen Teil, der Aufgaben zu *Reading Comprehension and Vocabulary* und zu *Grammar and Structures* vereint. Tabelle 12 enthält die Inhalte beider Englischprüfungen mit den passenden Inhalten des Lehrplans der Sekundarschule, des Lehrplan 21 und des Check S3. Die erste Spalte enthält jeweils den Prüfungsinhalt (Titel, Aufgabenformat) und in Klammern die Anzahl Punkte, die für die entsprechenden Aufgaben vergeben werden. Diese kann zwischen den BM-Typen und über die Jahre variieren. In der zweiten und dritten Spalte werden die Referenzen zum Lehrplan der Sekundarschule des Kantons Aargau sowie zum Lehrplan 21 festgehalten. In der vierten Spalte wird mit einem «x» angegeben, ob die Inhalte im Check S3 vorkommen.

Tabelle 12: Vergleich der Prüfungsinhalte der Englischprüfung mit den Inhalten der Lehrpläne und des Check S3

	Lehrplan		
Aufnahmeprüfung	Sekundarschule Kanton Aargau	Lehrplan 21	Check S3
Reading: Richtig-/Falsch-Aussagen (7–8 Punkte)	Kurze Texte zu aktuellen oder vertrauten Themen verstehen, in denen Meinungen und Haltungen dargelegt werden, resp. für oder gegen etwas argumentiert wird (Lesen B1.2)	Die Schülerinnen und Schüler … können Texte im Wesentlichen verstehen, wenn das Thema vertraut ist (z.B. Auszug aus einem Jugendbuch, Songtext, unkomplizierter Sachtext). (FS1E.2.A.1.f, B1.2) können in einfachen argumentativen Texten die zugrunde liegende Meinung oder Haltung erkennen (z.B. Blogeintrag). (FS1E.2.A.1.f, B1.2)	x
Vocabulary (3–4 Punkte)		verfügen über ein ausreichendes Repertoire an Wörtern und Wendungen, das ihnen erlaubt, Texte zu unterschiedlichen Themen der eigenen Lebenswelt zu verfassen sowie an alltäglichen Gesprächen aktiv teilzunehmen. (FS1E.5.B.1.d)	

Aufnahmeprüfung	Lehrplan		Check S3
	Sekundarschule Kanton Aargau	Lehrplan 21	
Choose the correct answer (5–6 Punkte)		können beim Sprechen und Schreiben auch einige komplexere grammatische Strukturen ausreichend korrekt verwenden (z.B. Hilfs- und Modalverben, verschiedene Formen der Negation, Demonstrativ- und Reflexivpronomen). (FS1E.5.D.1.d) können in Texten auch komplexere Strukturen erkennen und für das Verständnis nutzen (z.B. Zeitform der Verben, häufige Modalverben, Konjunktionen zur Begründung, Possessiv- und Reflexivpronomen). (FS1E.5.D.1.d)	
Fill in the correct verbs (7–8 Punkte)		können in Texten auch komplexere Strukturen erkennen und für das Verständnis nutzen (z.B. Zeitform der Verben, häufige Modalverben, Konjunktionen zur Begründung, Possessiv- und Reflexivpronomen). (FS1E.5.D.1.d) können unter Anleitung komplexere grammatische Strukturen untersuchen, mit anderen Sprachen vergleichen und Regeln ableiten (z.B. Zeitform, einfache Konjugation). (FS1E.5.D.2.b)	
Make questions about the underlined words (2–4 Punkte)		können in Texten einfache grammatische Strukturen erkennen und für das Verständnis nutzen (z.B. Wortfolge im Satz, Frage mit Fragewort oder Intonation, Konjugationsformen). (FS1E.5.D.1.c)	
Make negative sentences (2–4 Punkte)		können beim Sprechen und Schreiben auch einige komplexere grammatische Strukturen ausreichend korrekt verwenden (z.B. Hilfs- und Modalverben, verschiedene Formen der Negation, Demonstrativ- und Reflexivpronomen). (FS1E.5.D.1.d)	
Writing (10 Punkte)	Einen persönlichen, detaillierten Brief schreiben (über Erfahrungen, Gefühle, Ereignisse berichten) (Schreiben B1.1) Einfache Texte über vertraute Themen zusammenfassen und die eigene Meinung dazu schreiben (Schreiben B1.1)	können einfache Texte über vertraute Themen zusammenfassen und ihre Meinung dazu äussern (z.B. Buch- oder Filmbesprechung). (FS1E.4.A.1.e, B1.1) können in einfachen zusammenhängenden Texten über ein Ereignis berichten (z.B. Reisebericht, Experiment, kultureller Anlass). (FS1E.4.A.1.e, B1.1)	x[1]

[1] Englisch Schreiben wird im Check S3 alternierend mit Französisch Schreiben geprüft.

In der Englischprüfung für die BM I können insgesamt 40 Punkte erreicht werden. 7 bis 8 Punkte werden für *Reading Comprehension and Vocabulary* und 10 Punkte für *Writing* vergeben. Für die zwei Bereiche können im Lehrplan der Sekundarschule Referenzen nachgewiesen werden. Die übrigen 22 bis 23 Punkte werden für Inhalte des Bereichs *Grammar and Structures* vergeben, die nicht explizit im Lehrplan der Sekundarschule ausgewiesen werden. Im Kompetenzbereich *Sprache im Fokus* (Grammatik, Wortschatz) des Lehrplan 21 können hingegen auch zu diesen Inhalten Referenzen nachgewiesen werden, wobei die Inhalte in Grammatik wie in Wortschatz anwendungsorientiert beim Verstehen von (Hör-)Texten oder bei der eigenen Sprachproduktion aufgeführt sind. Gleich wie für Deutsch zeigt sich für Englisch, dass sich die Prüfungsinhalte zwar nicht vollständig mit den Inhalten des aktuellen Lehrplans decken, im Lehrplan 21 jedoch für alle Prüfungsinhalte zumindest eine annähernde Referenz nachgewiesen werden kann.

Im Check S3 werden Wortschatz und Grammatik anwendungsorientiert beim Schreiben eines Textes erfasst. Darüber hinaus wird mit dem Check S3 auch das Hörverstehen erhoben. Leseverständnis und Hörverständnis werden mit mindestens 30 Teilaufgaben erhoben und als eigenständige Kompetenzbereiche ausgewiesen.

Französisch

Die Französischprüfung umfasst vier Teile: *Compréhension de texte*, *Vocabulaire* (en contexte), *Grammaire*, *Production écrite*. Tabelle 13 enthält die Inhalte der Französischprüfung mit den passenden Inhalten des Lehrplans der Sekundarschule, des Lehrplan 21 und des Check S3. Die erste Spalte enthält jeweils den Prüfungsinhalt (Titel, Aufgabenformat) und in Klammern die Anzahl Punkte, die für die entsprechenden Aufgaben vergeben werden. Diese kann zwischen den BM-Typen und über die Jahre variieren. In der zweiten und dritten Spalte werden die Referenzen zum Lehrplan der Sekundarschule des Kantons Aargau sowie zum Lehrplan 21 festgehalten. In der vierten Spalte wird mit einem «x» angegeben, ob die Inhalte im Check S3 vorkommen.

In der Prüfung für die BM I konnten im Jahr 2017 maximal 65 Punkte erzielt werden, in den Jahren 2018 und 2019 maximal 80 Punkte. In der Prüfung für die BM II Typ «andere» konnten jeweils maximal 65 Punkte erzielt werden. Etwas mehr als ein Drittel der Punkte wird jeweils für *Grammaire* vergeben, je rund ein Viertel der Punkte für *Compréhension de texte* und für *Vocabulaire* sowie zehn Punkte für *Production écrite*. Für *Compréhension de texte*, *Grammaire* und *Production écrite* lassen sich im Lehrplan der Sekundarschule Referenzen nachweisen, für die Aufgaben zu *Vocabulaire* hingegen nicht. Trotzdem werden rund ein Viertel der Punkte für diesen Prüfungsinhalt vergeben. Die Referenzen in *Grammaire* beziehen sich meistens nicht auf die Abschlussklasse der Sekundarschule, sondern auf die beiden Jahre zuvor. Anders sieht der Vergleich mit dem Lehrplan 21 aus. Die meisten Inhalte des Bereichs *Grammaire* können dem Kompetenzbereich *Sprache im Fokus* (Grammatik, Wortschatz) des Lehrplan 21 zugeordnet werden.

Wie in Englisch werden auch in Französisch im Check S3 Wortschatz und Grammatik anwendungsorientiert beim Schreiben eines Textes erfasst. Darüber hinaus wird mit dem Check S3 auch das Hörverstehen erhoben. Leseverständnis und Hörverständnis werden mit mindestens 30 Teilaufgaben erhoben und als eigenständige Kompetenzbereiche ausgewiesen.

Tabelle 13: Vergleich der Prüfungsinhalte der Französischprüfung mit den Inhalten der Lehrpläne und des Check S3

Aufnahmeprüfung	Lehrplan		Check S3
	Sekundarschule Aargau	Lehrplan 21	
Lesetext: Multiple-Choice-Fragen (0–3 Punkte)	Einfache authentische Texte lesen und verstehen (3. Klasse) Beim Lesen verschiedener Textsorten Strategien zur Informationsentnahme entwickeln und anwenden	Die Schülerinnen und Schüler … können Texte im Wesentlichen verstehen, wenn das Thema vertraut ist (z.B. Auszug aus einem Jugendbuch, Liedtext, unkomplizierter Sachtext). (FS2F.2.A.1.f, B1.2) können in einfachen argumentativen Texten die zugrunde liegende Meinung oder Haltung erkennen (z.B. Blogeintrag). (FS2F.2.A.1.f, B1.2)	x
Lesetext: Vrai/Faux inkl. justification, falls faux, korrekten Satz aus Text dazuschreiben (12–13 Punkte)	Unbekanntes selbstständig, aus dem Kontext und mit Hilfe eines Wörterbuchs erschliessen (detailliertes und globales Leseverstehen, Sachtexte: Anweisungen, Rezepte) (4. Klasse)		x
Lesetext: Offene Fragen (0–8 Punkte)			
Complétez les phrases par un mot du texte (Vocabulaire) (5–9 Punkte)		verfügen über ein ausreichendes Repertoire an Wörtern und Wendungen, das ihnen erlaubt, Texte zu unterschiedlichen Themen der eigenen Lebenswelt zu verfassen sowie an alltäglichen Gesprächen aktiv teilzunehmen. (FS2F.5.B.1.d)	
Trouvez le contraire (5–10 Punkte)			
Wortfamilie (5 Punkte)			

Aufnahmeprüfung	Lehrplan		Check S3
	Sekundarschule Aargau	Lehrplan 21	
Verben konjugieren (10–12 Punkte)	Verbe pronominal: présent, impératif (2. Klasse), passé composé (3. Klasse) passé composé/participe passé (formation) (2. Klasse) Futur composé: aller + infinitif (2. Klasse) Verbe: verbes irréguliers, accord participe (3. Klasse) Imparfait: formation (3. Klasse)	können in Texten einfache grammatische Strukturen erkennen und für das Verständnis nutzen (z.B. Wortfolge im Satz, Frage mit Fragewort oder Intonation, Konjugationsformen). (FS2F.5.D.1.c) können unter Anleitung komplexere grammatische Strukturen untersuchen, mit anderen Sprachen vergleichen und Regeln ableiten (z.B. Zeitform, einfache Konjugation). (FS2F.5.D.2.b)	
Passé composé oder imparfait (0–6 Punkte)	Les temps du récit: passé composé – imparfait, plus-que-parfait (compétence de base) (3. Klasse)	können in Texten auch komplexere Strukturen erkennen und für das Verständnis nutzen (z.B. Zeitform der Verben, häufige Modalverben, Konjunktionen zur Begründung, Possessiv- und Reflexivpronomen). (FS2F.5.D.1.d) können unter Anleitung komplexere grammatische Strukturen untersuchen, mit anderen Sprachen vergleichen und Regeln ableiten (z.B. Zeitform, einfache Konjugation). (FS2F.5.D.2.b)	
Textteile durch Pronomen ersetzen (5–6 Punkte)	Complément et pronom direct/indirect, la place du pronom: COD, COI (3. Klasse) pronom tonique: moi, toi, etc. (1. Klasse) Pronom indirect: y/en (4. Klasse)	können in Texten einfache grammatische Strukturen erkennen und für das Verständnis nutzen (z.B. Wortfolge im Satz, Frage mit Fragewort oder Intonation, Konjugationsformen). (FS2F.5.D.1.c)	
Adjektive (3–5 Punkte)	Adjectif qualificatif: accord irrégulier (2. Klasse)		

Aufnahmeprüfung	Lehrplan		Check S3
	Sekundarschule Aargau	Lehrplan 21	
Structure de la phrase / Formulez les questions en français standard (Fragen stellen) (4–5 Punkte)	Interrogation: par intonation, avec «est-ce que», par inversion (compétence réceptive) (2. Klasse)	können beim Sprechen und Schreiben einfache grammatische Strukturen verwenden, machen dabei aber teilweise noch Fehler (z.B. Wortfolge in einfachen Aussagesätzen und Fragen, Verben mit zwei Objekten, unregelmässige Pluralformen, häufige Präpositionen). (FS2F.5.D.1.c) können in Texten einfache grammatische Strukturen erkennen und für das Verständnis nutzen (z.B. Wortfolge im Satz, Frage mit Fragewort oder Intonation, Konjugationsformen). (FS2F.5.D.1.c) können unter Anleitung einzelne grammatische Strukturen erforschen, gegebenenfalls mit anderen Sprachen vergleichen und darüber auf Deutsch Vermutungen anstellen (z.B. Singular-Pluralform, bestimmter/unbestimmter Artikel, Wortstellung in einem Fragesatz). (FS2F.5.D.2.a)	
Bestimmter und unbestimmter Artikel (0–5 Punkte)	Article: défini, indéfini, singulier, pluriel (1. Klasse) Article partitif: «de» partitif (2. Klasse) Préposition: à/de + article (2. Klasse)	können beim Sprechen und Schreiben einfache grammatische Strukturen verwenden, machen dabei aber teilweise noch Fehler (z.B. Wortfolge in einfachen Aussagesätzen und Fragen, Verben mit zwei Objekten, unregelmässige Pluralformen, häufige Präpositionen). (FS2F.5.D.1.c) können unter Anleitung einzelne grammatische Strukturen erforschen, gegebenenfalls mit anderen Sprachen vergleichen und darüber auf Deutsch Vermutungen anstellen (z.B. Singular-Pluralform, bestimmter/unbestimmter Artikel, Wortstellung in einem Fragesatz). (FS2F.5.D.2.a)	
Production écrite (7–10 Punkte)	Den Grundwortschatz korrekt schreiben Kurze eigene Texte schreiben und mit geeigneten Hilfsmitteln überarbeiten Im Erkennen und Anwenden sprachlicher Gesetzmässigkeiten grössere Sicherheit erlangen (Texte schreiben, 4. Klasse)	können einfache Texte über vertraute Themen zusammenfassen und ihre Meinung dazu äussern (z.B. Buch- oder Filmbesprechung). (FS2F.4.A.1.e, B1.1) können in einfachen zusammenhängenden Texten über ein Ereignis berichten (z.B. Reisebericht, Experiment, kultureller Anlass). (FS2F.4.A.1.e, B1.1)	x[1]

[1] Französisch Schreiben wird im Check S3 alternierend mit Englisch Schreiben geprüft.

Mathematik

Die Aufnahmeprüfungen in Mathematik werden nicht in Kompetenzbereiche unterteilt. Tabelle 14 enthält die Inhalte der Mathematikprüfung mit den passenden Inhalten des Lehrplans der Sekundarschule, des Lehrplan 21 und des Check S3. Die erste Spalte enthält jeweils den Prüfungsinhalt (Titel, Aufgabenformat) und in Klammern die Anzahl Punkte, die für die entsprechenden Aufgaben vergeben werden. Diese kann zwischen den BM-Typen und über die Jahre variieren. In der zweiten und dritten Spalte werden die Referenzen zum Lehrplan der Sekundarschule des Kantons Aargau sowie zum Lehrplan 21 festgehalten. In der vierten Spalte wird mit einem «x» angegeben, ob die Inhalte im Check S3 vorkommen.

In den Mathematikprüfungen werden insgesamt nur 19 bis 21 Punkte vergeben. Mit Ausnahme des Bereichs *Zahlenfolgen beschreiben und als Terme verallgemeinern* (Aufnahmeprüfung für die BM I 2017) und der Inhalte *kleinstes gemeinsames Vielfaches (kgV)* und *grösster gemeinsamer Teiler (ggT)*, die jedoch nur in einzelnen Prüfungen eingesetzt wurden, können für alle Inhalte bzw. Aufgaben Referenzen im Lehrplan der Sekundarschule des Kantons Aargau nachgewiesen werden. Der Lehrplan 21 enthält ebenfalls bis auf die Aufgaben zu den Inhalten *kgV* und *ggT* Referenzen zu allen Inhalten der Mathematikprüfung. Die Aufgaben in den Prüfungen verteilen sich in der Regel zu einem Drittel auf die drei Kompetenzbereiche des Lehrplan 21 (*Zahl und Variable*, *Form und Raum* sowie *Grössen, Funktionen, Daten und Zufall*). Eine Ausnahme stellt die Prüfung für die BM II Typ «andere» im Jahr 2019 dar. In dieser Prüfung wurden jeweils 8 Punkte für die Kompetenzbereiche *Zahl und Variable* sowie *Form und Raum*, jedoch nur 4 Punkte für *Grössen, Funktionen, Daten und Zufall* vergeben. Bis auf die Konstruktionsaufgaben zum Kompetenzbereich *Form und Raum* sind im Check S3 vergleichbare Aufgaben wie in den Prüfungen enthalten.

Tabelle 14: Vergleich der Prüfungsinhalte der Mathematikprüfung mit den Inhalten der Lehrpläne und des Check S3

Aufnahmeprüfung	Lehrplan		Check S3
	Sekundarschule Aargau	Lehrplan 21	
		Die Schülerinnen und Schüler …	x
Terme vereinfachen, kürzen, zerlegen (2–4 Punkte)	Das Potenzgesetz der Multiplikation verstehen und anwenden (4. Klasse)	können das Distributivgesetz bei Termumformungen anwenden (z.B. $a \cdot (b + c) = a \cdot b + a \cdot c = ab + ac$). (MA.1.A.4.i)	
	(Quadratwurzel von Zahlen und Termen kennen lernen (4. Klasse))	Erweiterung: können Polynome addieren und subtrahieren (z.B. $3(a^2 + 2b) - 2(a^2 + b) = a^2 + 4b$). (MA.1.A.4.j)	
	Einfache Terme umformen (3. und 4. Klasse)	Erweiterung: können Terme ausmultiplizieren und ausklammern (Faktorzerlegung). (MA.1.A.4.j)	
	Grundoperationen mit einfachen Termen ausführen (3. und 4. Klasse)	Erweiterung: können Terme mit Variablen umformen bzw. sinnvoll vereinfachen (ausklammern, ausmultiplizieren, kürzen und Vorzeichenregeln). (MA.1.A.4.j)	
		können Terme mit Variablen addieren und subtrahieren (z.B. $a + 2a + b + 3b + \frac{1}{4} + \frac{3}{8} = 3a + 4b + \frac{5}{8}$). (MA.1.A.4.k)	

Aufnahme-prüfung	Lehrplan		Check S3
	Sekundarschule Aargau	Lehrplan 21	
Gleichungen (2–4 Punkte)	Lineare Gleichungen und Ungleichungen mit einer Variablen mit Hilfe von Äquivalenzumformungen auflösen (4. Klasse) Bruchgleichungen mit einer Variablen lösen (und in Sachaufgaben anwenden) (4. Klasse)	Erweiterung: können lineare Gleichungen mit einer Variablen mit Äquivalenzumformungen lösen (z.B. $5x + 3 = 7$). (MA.1.A.4.j) können lineare Gleichungssysteme mit 2 Unbekannten lösen. (MA.1.A.4.m)	x
Grössen umwandeln (0–2 Punkte)	Sich in der Systematik der SI-Einheiten orientieren und in Sachaufgaben anwenden (4. Klasse)	können Grössen (Geld, Längen, Gewicht bzw. Masse, Zeit, Volumen [l]) schätzen, bestimmen, vergleichen, runden, mit ihnen rechnen, in benachbarte Masseinheiten umwandeln und in zweifach benannten Einheiten schreiben. (MA.3.A.2.h) können Flächeninhalte und Volumen [m^3] in einer geeigneten Masseinheit schätzen und in benachbarte Masseinheiten umwandeln. (MA.3.A.2.i)	x
Sachrechnen (2–6 Punkte)	Sich in der Systematik der SI-Einheiten orientieren und in Sachaufgaben anwenden (4. Klasse) Aus einfachen Sachzusammenhängen Gleichungen und Ungleichungen ableiten (3. Klasse) Lineare Gleichungen und Ungleichungen mit einer Variablen mit Hilfe von Äquivalenzumformungen auflösen (4. Klasse) In Sachrechnungen Sicherheit im Umgang mit Prozenten und Prozentrechnungen gewinnen (3. Klasse) Verschiedene Anwendungsbereiche von Prozentrechnungen kennen (3. Klasse)	Erweiterung: können Gleichungen sprachlich deuten (z.B. $x = y + 1 \rightarrow x$ ist um 1 grösser als y) und Textgleichungen umsetzen. (MA.1.A.4.j)	x
Konstruktionen (1,5–3 Punkte)	Dreiecke konstruieren (Inkreis etc., 4. Klasse) Kongruenzabbildungen kennen und konstruktiv ausführen (Spiegelungen etc., 3. Klasse)	können Senkrechte, Winkelhalbierende und Mittelsenkrechte mit dem Geodreieck zeichnen. (MA.2.C.2.h) können Winkelhalbierende, Mittelsenkrechte und gleichseitiges Dreieck mit Zirkel und Lineal konstruieren. (MA.2.C.2.h) können Figuren und geometrische Beziehungen skizzieren und Zeichnungen mit Geodreieck und Zirkel oder dynamischer Geometriesoftware ausführen (z.B. ein Parallelogramm mit a, b und h_a zeichnen oder konstruieren). (MA.2.C.2.i)	

Aufnahme-prüfung	Lehrplan		Check S3
	Sekundarschule Aargau	Lehrplan 21	
Berechnungen (Fläche, Höhe etc.) (2–3 Punkte)	Flächen von Trapezen bestimmen (3. Klasse)	können den Flächeninhalt von Drei- und Vierecken berechnen. (MA.2.A.3.g) können Kantenlängen, Seitenflächen und Volumen von Quadern berechnen. (MA.2.A.3.g)	x
Volumen u.Ä. (2 Punkte)	Oberflächen und Volumen von geraden Zylindern, Pyramiden, Kugeln und Kegeln bestimmen (4. Klasse)	können Längen und Flächeninhalte mithilfe des Satzes von Pythagoras berechnen. (MA.2.A.3.h) können Umfang und Flächeninhalt von Kreisen berechnen. (MA.2.A.3.i) können Kantenlängen, Flächen und Volumen an geraden Prismen und Zylindern berechnen. (MA.2.A.3.i) können Strecken, Flächen und Volumen an Pyramiden, Kegeln und Kugeln berechnen. (MA.2.A.3.j) können Winkel aufgrund von Winkelsummen, Satz von Thales, Ähnlichkeit und Kongruenz bestimmen. (MA.2.A.3.j)	x
Zahlenfolgen beschreiben und als Terme verallgemeinern (0–2 Punkte) kgV und ggT (0–2 Punkte)		können lineare Figurenfolgen in einen Term übertragen (z.B. die Anzahl benötigter Hölzchen, um eine Reihe von n gleichseitigen Dreiecken zu legen, als 2n + 1). (MA.1.C.2.i)	x

Aufnahmeprüfung BM II Typ «Wirtschaft»

Für die Inhalte der Aufnahmeprüfung in die BM II Typ «Wirtschaft» gilt der Rahmenlehrplan für die Berufsmaturität des SBFI. Dieser Lehrplan enthält Stundentafeln zu den einzelnen Fachbereichen, jedoch keine konkreten Angaben zu Inhalten oder Kompetenzbeschreibungen. Aus diesem Grund werden im Folgenden die Inhalte der Aufnahmeprüfungen ohne Vergleich zu einem Lehrplan dargestellt und diskutiert.

Deutsch

Die Deutschprüfung umfasst zwei Teile: *Textverständnis und Sprachbetrachtung* und *Verfassen eines Textes*. Für beide Bereiche werden je 50 Punkte vergeben. Im Bereich *Sprachbetrachtung* werden in erster Linie die Kompetenzen in der Grammatik erfasst. Daneben enthält dieser Prüfungsteil auch noch wenige Aufgaben zur Rechtschreibung und zum Wortschatz. Im Gegensatz zu den anderen Deutschprüfungen (BM I, BM II Typ «andere») ist für das Verfassen des Textes ein Wörterbuch als Hilfsmittel erlaubt. Diese Deutschprüfung enthält keine Aufgaben zum Leseverstehen. In Tabelle 15 sind die Prüfungsinhalte der Jahre 2017/18 und 2019 aufgeführt. Die geprüften Themen unterscheiden sich kaum zwischen den Jahren.

Tabelle 15: Prüfungsinhalte BM II Typ «Wirtschaft» Deutsch

	Punkte	
Prüfungsinhalt	2017/2018	2019
Kommas setzen	10	10
Synonyme	3	3
Rechtschreib- und Fallfehler	8	8
Ausdrücke in den korrekten Fall setzen	2	3
Verb in richtige Zeitform setzen (Synonym?)	2	2
Konjunktionen	3	
Präpositionen	3	3
Wortfamilie	4	4
Aktiv/Passiv	4	4
Einfacher Satz in Satzgefüge umwandeln	4	4
Indirekte Rede	2	3
Stammformen (Verb als Partizip 2 etc.)/verlangte Verbform	5	6
Aufsatz (Hilfsmittel: Rechtschreibewörterbuch)	50	50
Total	100	100

Englisch

In Tabelle 16 sind die Prüfungsinhalte der Englischprüfungen aufgeführt. Die Prüfung umfasst *Leseverstehen*, *Grammatik* und *Schreiben*. Es werden, verteilt auf diese Kompetenzbereiche, jeweils 70 Punkte vergeben.

Die geschriebenen Texte werden nach folgenden vier Kriterien beurteilt (5 Punkte pro Kriterium): *Content*, *Organisation*, *Grammar*, *Vocabulary*.

Tabelle 16: Prüfungsinhalte BM II Typ «Wirtschaft» Englisch

Prüfungsinhalt	Punkte
Reading Comprehension	20
Grammar (Diverses)	30
Writing	20
Total	70

Französisch

Bei der Französischprüfung werden jeweils die Kompetenzbereiche *Leseverstehen* und *Textproduktion* geprüft. Die Prüfungsinhalte und Punkte sind in Tabelle 17 aufgeführt.

Die Beurteilung der geschriebenen Texte erfolgt nach folgenden fünf Kriterien: *Erfüllung der Aufgabenstellung* (4 Punkte), *Kohärenz und Lexique* (5 Punkte), *Sprachkorrektheit* (5 Punkte), *Rituel épistolaire* (2 Punkte), *Genereller Eindruck* (2 Punkte).

Tabelle 17: Prüfungsinhalte BM II Typ «Wirtschaft» Französisch

	Punkte	
Prüfungsinhalt	2017/2018	2019
Leseverstehen 1 (vrai/faux + justification)	16	12
Leseverstehen 2 (Namen)	6	6
Schreiben (offizieller Brief/Geschäftsbrief)	18	18
Total	40	36

Finanz- und Rechnungswesen

In Tabelle 18 sind die Prüfungsinhalte und Punkte der Prüfung zum Finanz- und Rechnungswesen enthalten. Die geprüften Themen sind in den erfassten Jahren identisch. Es werden insgesamt 100 Punkte vergeben.

Tabelle 18: Prüfungsinhalte BM II Typ «Wirtschaft» Finanz- und Rechnungswesen

	Punkte	
Prüfungsinhalt	2017/2018	2019
Warenkonten	10	10
Mehrwertsteuer	10	10
Rechnungsabgrenzungen (Rückstellungen)	10	10
Einzelunternehmen, Privatkonto	10	10
Definitive Verluste aus Forderungen	5	5
Kennzahlen	5	5
Abschreibungen	5	5
Nutzschwelle	5	5
Lohnabrechnung	5	5
Stille Reserven	5	5
Zinsrechnen	3	5
Fremde Währungen	7	5
Aktiengesellschaft	10	10
Kalkulation im Handelsbetrieb	10	10
Total	100	100

Fazit

Die Inhalte der Prüfungen passen bis auf wenige Ausnahmen gut zu den Lehrplaninhalten. In Deutsch und in den Fremdsprachen lassen sich jedoch Abweichungen feststellen. So sind insbesondere für die Aufgaben zum Wortschatz im Lehrplan der Sekundarschule keine Referenzen zu finden. Im künftig massgebenden Lehrplan 21 sind hingegen auch zu den Wortschatzaufgaben Entsprechungen zu finden. Die entsprechenden Prüfungsaufgaben passen allerdings nur im Ansatz zu den diesbezüglichen Kompetenzbeschreibungen im Lehrplan 21. Auch die Inhalte der Checks

decken sich grösstenteils mit den Inhalten der Aufnahmeprüfungen, wobei in den Fremdsprachen mit den Checks auch der Kompetenzbereich *Hören* erfasst wird. Im Gegensatz zu den Aufnahmeprüfungen wird mit den Checks zudem der Kompetenzbereich *Lesen* ausführlich getestet, und zwar sowohl in Deutsch als auch in den Fremdsprachen.

Unabhängig von der Passung stellt sich die Frage, ob der Wortschatz in einer Aufnahmeprüfung geprüft werden soll oder nicht. Der Wortschatz ist stark herkunftsabhängig und entspricht deshalb einer Kompetenz, die von der Schule weniger stark gefördert werden kann als beispielsweise das Verfassen eines Textes oder Grammatikkenntnisse.

Zudem stellt sich die Frage, ob das Leseverständnis ausreichend geprüft wird. Zum einen wird das Leseverständnis in Deutsch nur innerhalb des Prüfungsteils *Textverständnis und Sprachbetrachtung* erfasst. Zum anderen fällt die Gewichtung aufgrund der vergebenen Punktzahl im Verhältnis von 1 zu 3 zuungunsten des Textverständnisses aus. Die marginale Bedeutung der Lesekompetenzen für das erfolgreiche Absolvieren eines Maturitätslehrgangs ist in Anbetracht der vergleichsweise geringen Lesekompetenzen von Schweizer Jugendlichen problematisch (Konsortium PISA.ch, 2019). Ein Grund für die geringe Präsenz der Lesekompetenzen in den Deutschprüfungen könnte darin liegen, dass die Entwicklung von geeigneten Aufgaben zur Erfassung von Lesekompetenzen wesentlich aufwendiger und schwieriger ist als die Entwicklung von Aufgaben zu Wortschatz und zu Grammatik.

Eine ähnliche Tendenz lässt sich auch für die Inhalte der Aufnahmeprüfungen in den Fremdsprachen feststellen. Wortschatz und Grammatik werden stärker gewichtet als das Verstehen eines Textes. Auf das Hörverständnis wird vollständig verzichtet, obwohl die rezeptiven Kompetenzen *Lesen* und *Hören* (gleich wie die produktiven Kompetenzen *Schreiben* und *Sprechen*) zu einer vollständigen Beurteilung der Sprachkompetenzen im Rahmen einer Zertifizierung von Sprachkompetenzen gehören.

Während die inhaltliche Passung zwischen Aufnahmeprüfungen und Lehrplänen als gegeben beurteilt werden kann, ist die Gewichtung der einzelnen Prüfungsinhalte mit den Zielen des Lehrplans und des Sprachunterrichts nicht kompatibel. Dies zeigt sich auch an der geringen Bedeutung, welche beispielsweise die Vorkenntnisse in den Fremdsprachen für den Prüfungsentscheid haben (vgl. Absatz 6.1).

Für eine bessere Abstimmung der Prüfungsinhalte auf die Lehrpläne der Volksschule müsste in den Sprachen das Leseverständnis ausführlich getestet werden. Zudem wäre es besser, das Leseverständnis mit Multiple-Choice-Aufgaben anstatt mit Binary-Choice-Aufgaben (richtig/falsch) zu prüfen. Bei Multiple-Choice-Aufgaben ist üblicherweise eine von vier oder fünf Wahlantworten eindeutig die richtige oder die beste. Falls Binary-Choice-Aufgaben eingesetzt werden, sollten jeweils mehrere Teilaufgaben zu einer Aufgabe zusammengefasst werden. Punkte werden dabei für das korrekte Beantworten von allen oder mehreren Teilfragen vergeben. Bei Binary-Choice-Aufgaben müssen die Aussagen immer eindeutig richtig oder falsch sein und jede Aussage muss unabhängig von allen anderen Aussagen sein. In den Fremdsprachen könnte eine bessere Passung zwischen Prüfungs- und Lehrplaninhalten durch den Einbezug des Hörverständnisses und der Sprechkompetenzen erreicht werden. Die Entwicklung von geeigneten Aufgaben im Hörverständnis ist anspruchsvoll und erfordert für alle Kandidatinnen und Kandidaten eine qualitativ gute und vergleichbare elektronische Präsentation des Hörtextes. Die zuverlässige Erfassung der Sprechkompe-

tenzen ist nur mit einem erheblichen Aufwand zu meistern.[5] Schliesslich würde die Passung zwischen Prüfungs- und Lehrplaninhalten auch dadurch optimiert, wenn der Anteil der einbezogenen Kompetenzbereiche gleich gewichtet würde.

6.3 Prüfungsteile und Prüfungsentscheid

Die Beurteilung der einzelnen Prüfungsteile für den Prüfungsentscheid hat zum Ziel, die Effizienz der Prüfung zu steigern. Ziel einer Prüfung ist es, dass sämtliche Prüfungsteile bzw. sämtliche geprüften Fächer und Kompetenzbereiche gleichermassen für den Prüfungsentscheid von Bedeutung sind. Ob dieses Ziel erreicht wird, hängt von der Qualität der Prüfungsteile und von den Beurteilungsmassstäben ab.

Für die Beurteilung der Prüfungsteile wurden die Aufgaben verschiedenen Kompetenzbereichen zugeordnet und die erreichten Punktzahlen für die einzelnen Kompetenzbereiche neu summiert. Die Zuordnung der einzelnen Aufgaben zu den Kompetenzbereichen kann Anhang B entnommen werden. Tabelle 19 enthält die mit den Prüfungsteilen erfassten Kompetenzbereiche der Prüfungen von 2019 für die drei Ausrichtungen.

Tabelle 19: Aufnahmeprüfungen 2019: Kompetenzbereiche nach Ausbildungsweg

	Ausrichtungen		
	BM I	BM II – Typ «andere»	BM II – Typ «Wirtschaft»
Deutsch	Lesen Sprache im Fokus Schreiben	Schreiben	Sprache im Fokus Schreiben
Englisch	Lesen Sprache im Fokus Schreiben	Lesen Sprache im Fokus	Lesen Sprache im Fokus Schreiben
Französisch	Lesen Sprache im Fokus Schreiben	Lesen Sprache im Fokus Schreiben	Lesen Schreiben
Mathematik	Zahl und Variable Form und Raum Grössen, Funktionen, Daten und Zufall	Zahl und Variable Form und Raum Grössen, Funktionen, Daten und Zufall	Finanz- und Rechnungswesen

Die Beurteilung der Bedeutung einzelner Kompetenzbereiche – bspw. *Lesen* oder *Sprache im Fokus* – erfolgt anhand von logistischen Regressionsanalysen. Mit diesen Analysen wird überprüft, inwieweit der Prüfungsentscheid (bestanden/nicht bestanden) mit den Noten in den einzelnen Prüfungsteilen zusammenhängt oder mit anderen Worten, welche Bedeutung jeder Prüfungsteil für den Prüfungsentscheid hat.

[5] Aufgrund des grossen Aufwands, der mit einer zuverlässigen Erfassung der Sprechkompetenzen einhergeht, sind die Sprechkompetenzen nicht Bestandteil des Check S3.

BM I

Werden alle Prüfungsteile zur Vorhersage des Prüfungsentscheids (bestanden/nicht bestanden) in die Regressionsanalyse einbezogen, dann kann der Prüfungsentscheid nahezu perfekt vorausgesagt werden.

Damit die Bedeutung einzelner Prüfungsteile beurteilt werden kann, werden in mehreren Regressionsanalysen systematisch einzelne Prüfungsteile ausgeschlossen. Die Ergebnisse der verschiedenen Analysen werden in einem nächsten Schritt miteinander verglichen. Mit diesem Vorgehen kann festgestellt werden, ob der Prüfungsentscheid mit weniger Prüfungsteilen ebenso gut vorhergesagt werden kann wie mit allen Prüfungsteilen. Dargestellt wird jeweils das kürzeste Modell mit jenen Prüfungsteilen, die für die Vorhersage des Prüfungsentscheids statistisch signifikant sind und den Prüfungsentscheid ebenso gut vorhersagen wie das vollständige Modell, mit dem der Prüfungsentscheid nahezu perfekt vorhergesagt werden kann.

Tabelle 20 enthält die Ergebnisse von drei Regressionsanalysen. Bei Modell 1 handelt es sich um das vollständige Modell, das alle geprüften Kompetenzbereiche einbezieht. *Deutsch Schreiben*, *Französisch Sprache im Fokus* und *Mathematik* hängen statistisch signifikant mit dem Prüfungsentscheid zusammen. Die erklärte Varianz liegt bei R^2 = .64. Modell 2 zeigt das kürzeste Modell. Es enthält die Kompetenzbereiche *Deutsch Lesen*, *Deutsch Schreiben*, *Französisch Sprache im Fokus* und *Mathematik.* Die Regressionskoeffizienten sind für alle vier Kompetenzbereiche statistisch signifikant. Die erklärte Varianz liegt bei R^2 = .67. Das bedeutet, dass der Prüfungsentscheid mit nur vier Kompetenzbereichen ebenso gut vorhergesagt werden kann wie mit allen Kompetenzbereichen. Die Prüfungsaufgaben der Kompetenzbereiche *Deutsch Schreiben*, *Französisch Sprache im Fokus* und *Mathematik* haben sowohl in Modell 1 als auch in Modell 2 eine statistisch signifikante Bedeutung für den Prüfungsentscheid. Die Bedeutung von *Deutsch Lesen* zeigt sich in Modell 2, nicht aber in Modell 1.

Zur Einordnung der Bedeutung der einzelnen Kompetenzbereiche enthält Modell 3 jene Kompetenzbereiche, die in Modell 2 nicht einbezogen wurden. Wird der Prüfungsentscheid ausschliesslich mit den in Modell 1 statistisch nicht signifikanten Kompetenzbereichen *Deutsch Sprache im Fokus*, *Englisch Lesen*, *Englisch Sprache im Fokus*, *Englisch Schreiben*, *Französisch Lesen* und *Französisch Schreiben* vorhergesagt, dann liegt der Anteil der erklärten Varianz mit R^2 = .20 deutlich tiefer. Dementsprechend passt Modell 3 deutlich schlechter zu den vorliegenden Daten als die Modelle 1 und 2. Die in Modell 3 einbezogenen Kompetenzbereiche hängen nur schwach mit dem Prüfungsentscheid zusammen.

Ein Vergleich der Koeffizienten der einzelnen Fächer in den drei Modellen lässt bezüglich ihrer Bedeutung für den Prüfungsentscheid folgenden Schluss zu: Für den Prüfungsentscheid sind die Kompetenzbereiche *Mathematik* und *Französisch Sprache im Fokus* von sehr grosser Bedeutung. Auch die erreichten Punkte im Bereich *Deutsch Schreiben* hängen statistisch signifikant mit dem Prüfungsentscheid zusammen. Das Abschneiden in allen drei Kompetenzbereichen *Englisch* sowie in den Kompetenzbereichen *Französisch Lesen* und *Französisch Schreiben* spielen für den Prüfungsentscheid hingegen eine untergeordnete Rolle. Die je nach Modell unterschiedlichen Ergebnisse für *Deutsch Lesen* und *Deutsch Sprache im Fokus* legen die Vermutung nahe, dass sich diese Kompetenzbereiche in Deutsch und in den Fremdsprachen überschneiden. Dies trifft besonders für Französisch zu. Werden in einer Prüfung *Englisch Lesen* und *Französisch Lesen* nicht geprüft, dann ist die Bedeutung von *Deutsch Lesen* für den Prüfungsentscheid signifikant (Modell 2). Wird in einer Prüfung *Französisch Sprache im Fokus* nicht geprüft, dann ist die Bedeutung von *Deutsch Sprache im Fokus* für den Prüfungsentscheid signifikant (Modell 3).

Tabelle 20: Zusammenhang zwischen dem Prüfungsentscheid (bestanden/nicht bestanden) und den Punktzahlen in den geprüften Kompetenzbereichen (BM I)

	Modell 1	Modell 2	Modell 3
Konstante	–23.200***	–20.955***	–7.334***
Deutsch			
Lesen	.243	.300*	
Sprache im Fokus	.026		.325**
Schreiben	.147*	.140*	
Englisch			
Lesen	.084		.131
Sprache im Fokus	–.094		.100
Schreiben	.194		.030
Französisch			
Lesen	.157		.101*
Sprache im Fokus	.180**	.217***	
Schreiben	–.112		.122
Mathematik[1]	.903***	.837***	
McFaddens korrigiertes R^2	.64	.67	.20

[1] Aufgrund der geringen Anzahl an Aufgaben in Mathematik pro Kompetenzbereich werden hier alle Kompetenzbereiche zusammengenommen.

* $p < .05$, ** $p < .01$, *** $p < .001$

Sensitivität und Spezifität

Tabelle 21 zeigt die nahezu perfekte Vorhersage des Prüfungsentscheids mit allen Prüfungsteilen (Modell 1). Wenn die erreichten Punktzahlen der 153 Kandidatinnen und Kandidaten für alle geprüften Kompetenzbereiche vorliegen, können 73 Kandidatinnen und Kandidaten richtigerweise der Kategorie «bestanden» (Sensitivität = .94) und 78 Kandidatinnen und Kandidaten richtigerweise der Kategorie «nicht bestanden» (Spezifität = .95) zugeordnet werden. Die Gesamttrefferquote liegt bei 94 Prozent.

Tabelle 21: Vierfeldertafel Modell 1 mit sämtlichen Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	73	5	Sensitivität = .94
nicht bestanden	4	71	Spezifität = .95

Der Grund, weshalb nicht alle Kandidatinnen und Kandidaten perfekt zugeordnet werden, liegt darin, dass in der Regressionsgleichung ausschliesslich lineare Zusammenhänge modelliert werden. In der Realität liegt perfekte Linearität aber nicht vor. Durch die zur Berechnung der Endnote ver-

wendete Formel und die Regel, dass höchstens eine der Fachnoten in Deutsch, Mathematik und Fremdsprachen unter 4.0 liegen darf, wird die perfekte Linearität verletzt.

Tabelle 22 enthält die Vorhersage des Prüfungsentscheids mit nur vier Prüfungsteilen, dem kürzesten Modell 2. Die Gesamttrefferquote liegt bei 93 Prozent und somit nur unwesentlich tiefer als bei Modell 1 mit sämtlichen Prüfungsteilen (94 Prozent). Von den 153 Kandidatinnen und Kandidaten wird für 11 eine von der Wirklichkeit abweichende Vorhersage getroffen. Mit einer verkürzten Prüfung, die nur die Kompetenzbereiche *Deutsch Lesen*, *Deutsch Schreiben*, *Französisch Sprache im Fokus* und *Mathematik* einbeziehen würde, könnte der Prüfungsentscheid aber immer noch für 93 Prozent der Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, und für 92 Prozent, die die Prüfung nicht bestanden haben, korrekt vorhergesagt werden.

Tabelle 22: Vierfeldertafel Modell 2 mit vier Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	72	6	Sensitivität = .93
nicht bestanden	5	70	Spezifität = .92

Die Vorhersagen in Tabelle 22 wurden unter der Annahme getroffen, dass falsch positive Fälle gleich gewichtet werden wie falsch negative Fälle. Das heisst, dass es als gleich bedeutend erachtet wird, wenn eine Kandidatin, die die Aufnahmeprüfung nicht bestanden hat, einen positiven Bescheid erhält, wie wenn ein Kandidat, der die Aufnahmeprüfung bestanden hat, einen negativen Bescheid erhält. Aus einer *bildungspolitischen Perspektive* könnte auch das Ziel verfolgt werden, falsch negative Fälle möglichst zu vermeiden, dass also Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, auch mit einer kürzeren Prüfung zu 99 Prozent der Kategorie «bestanden» zugeordnet würden. Dazu müsste die Sensitivität mindestens .99 betragen. Dies lässt sich durch die Variation des Schwellenwertes erreichen, allerdings mit der Folge, dass die Sensitivität auf Kosten der Spezifität erhöht wird. Wenn möglichst alle Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, mit den Ergebnissen in der verkürzten Prüfung der Kategorie «bestanden» zugeordnet werden sollen, dann werden auch mehr Kandidatinnen und Kandidaten dieser Kategorie zugeordnet, die die Prüfung nicht bestanden haben. Die Ergebnisse des Modells 2 mit einer Sensitivität von .99 sind in Tabelle 23 dargestellt.

Tabelle 23: Vierfeldertafel: Modell 2 mit vier Kompetenzbereichen und einer Sensitivität von .99

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	74	1	Sensitivität = .99
nicht bestanden	13	65	Spezifität = .83

Wenn die Sensitivität .99 betragen soll, dann sinkt die Spezifität auf .83. 13 Kandidatinnen und Kandidaten, die die Prüfung nicht bestanden haben, würden dadurch fälschlicherweise der Kategorie «bestanden» zugeordnet. *Ökonomisch* gesehen führen Kandidatinnen und Kandidaten, die fälschlicherweise einen positiven Prüfungsentscheid bekommen, zu einer Fehlallokation von Ausbildungsressourcen. Deswegen könnte gegenteilig mit der Knappheit an Ausbildungsplätzen ar-

gumentiert werden, dass falsch positive Fälle möglichst vermieden werden sollten. In diesem Fall sollte die Spezifität mindestens .99 betragen. Die Ergebnisse des Modells 2 mit einer Spezifität von .99 sind in Tabelle 24 dargestellt. Wenn die Spezifität .99 betragen soll, dann sinkt die Sensitivität auf .75. 19 Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, würden folglich fälschlicherweise der Kategorie «nicht bestanden» zugeordnet.

Tabelle 24: Vierfeldertafel Modell 2 mit vier Kompetenzbereichen und einer Spezifität von .99

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	56	19	Sensitivität = .75
nicht bestanden	1	77	Spezifität = .99

In Tabelle 25 sind schliesslich die Ergebnisse der Vorhersage des Prüfungsentscheids mit jenen Kompetenzbereichen dargestellt, die in Modell 2 nicht einbezogen wurden, also die Ergebnisse von Modell 3. Die Gesamttrefferquote liegt bei 73 Prozent. 20 von 75 Kandidatinnen und Kandidaten, die die Prüfungen bestanden haben, hätten fälschlicherweise einen negativen Prüfungsentscheid erhalten (Sensitivität = .73). Von den 78 Kandidatinnen und Kandidaten, die die Aufnahmeprüfung nicht bestanden haben, wurden 57 korrekterweise der Kategorie «nicht bestanden» zugeordnet (Spezifität = .73). Dies zeigt noch einmal deutlich, dass die Bedeutung der Kompetenzbereiche *Deutsch Sprache im Fokus*, *Französisch Lesen* und *Französisch Schreiben* sowie der gesamten Englischprüfung im Vergleich zu den Kompetenzbereichen *Deutsch Lesen*, *Deutsch Schreiben*, *Französisch Sprache im Fokus* und *Mathematik* sehr klein ist. Die Fehlallokation steigt von 9 (Modell 1) auf 11 (Modell 2) und schliesslich auf 41 Fälle (Modell 3).

Tabelle 25: Vierfeldertafel Modell 3 mit nur unbedeutenden Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	55	20	Sensitivität = .73
nicht bestanden	21	57	Spezifität = .73

BM II Typ «andere»

Tabelle 26 enthält die Ergebnisse von drei Regressionsanalysen anhand der Ergebnisse in der Prüfung für die BM II. Bei Modell 1 handelt es sich um das vollständige Modell, das alle geprüften Kompetenzbereiche einbezieht. Deutsch Schreiben, Englisch Sprache im Fokus, Französisch Sprache im Fokus und Mathematik hängen statistisch signifikant mit dem Prüfungsentscheid zusammen. Die erklärte Varianz liegt bei R^2 = .76. Modell 2 zeigt das kürzeste Modell. Es enthält die Kompetenzbereiche Deutsch Schreiben, Englisch Sprache im Fokus, Französisch Sprache im Fokus und Mathematik. Die Regressionskoeffizienten dieser vier Kompetenzbereiche sind statistisch signifikant. Die erklärte Varianz liegt bei R^2 = .77. Das bedeutet, dass der Prüfungsentscheid mit nur vier Kompetenzbereichen ebenso gut vorhergesagt werden kann wie mit allen Kompetenzbereichen. Die von den Kandidatinnen und Kandidaten erreichte Punktzahl in Deutsch Schreiben, Englisch Sprache im Fokus, Französisch Sprache im Fokus und Mathematik hat sowohl in Modell 1 als auch in Modell 2 eine statistisch signifikante Bedeutung für den Prüfungsentscheid.

Zur Einordnung der Bedeutung der einzelnen Kompetenzbereiche enthält Modell 3 jene Kompetenzbereiche, die in Modell 2 nicht einbezogen sind. Wird der Prüfungsentscheid ausschliesslich mit den in Modell 1 statistisch nicht signifikanten Kompetenzbereichen *Englisch Lesen*, *Französisch Lesen* und *Französisch Schreiben* vorhergesagt, dann liegt der Anteil der erklärten Varianz mit R^2 = .09 deutlich tiefer. Dementsprechend passt Modell 3 deutlich schlechter zu den vorliegenden Daten als die Modelle 1 und 2. Die in Modell 3 einbezogenen Kompetenzbereiche hängen nur schwach mit dem Prüfungsentscheid zusammen.

Ein Vergleich der Koeffizienten der einzelnen Kompetenzbereiche in den drei Modellen lässt bezüglich ihrer Bedeutung für den Prüfungsentscheid folgenden Schluss zu: Für den Prüfungsentscheid ist die Bedeutung der Kompetenzbereiche *Deutsch Schreiben*, *Englisch Sprache im Fokus*, *Französisch Sprache im Fokus* und *Mathematik* sehr gross. Die zusätzlichen Informationen, die aus den Kompetenzbereichen *Englisch Lesen*, *Französisch Lesen* und *Französisch Schreiben* gewonnen werden können, spielen für den Prüfungsentscheid hingegen nahezu keine Rolle.

Tabelle 26: Zusammenhang zwischen dem Prüfungsentscheid (bestanden/nicht bestanden) und den Punktzahlen in den geprüften Kompetenzbereichen (BM II)

	Modell 1	Modell 2	Modell 3
Konstante	–42.837***	–35.032***	–4.266***
Deutsch			
Schreiben	1.269***	1.119***	
Englisch			
Lesen	.377		.420***
Sprache im Fokus	.498**	.535***	
Französisch			
Lesen	.207		.159**
Sprache im Fokus	.185**	.192***	
Schreiben	.126		.073
Mathematik[1]	1.471***	1.305***	
McFaddens korrigiertes R^2	.76	.77	.09

[1] Aufgrund der geringen Anzahl an Aufgaben in Mathematik pro Kompetenzbereich werden hier alle Kompetenzbereiche zusammengenommen.

* $p < .05$, ** $p < .01$, *** $p < .001$

Sensitivität und Spezifität

Tabelle 27 zeigt die nahezu perfekte Vorhersage des Prüfungsentscheids mit allen Prüfungsteilen (Modell 1). Wenn die erreichten Punktzahlen der 201 Kandidatinnen und Kandidaten für alle geprüften Kompetenzbereiche vorliegen, können 121 Kandidatinnen und Kandidaten richtigerweise der Kategorie «bestanden» (Sensitivität = .96) und 69 Kandidatinnen und Kandidaten richtigerweise der Kategorie «nicht bestanden» (Spezifität = .93) zugeordnet werden. Die Gesamttrefferquote liegt bei 95 Prozent. Der Grund, weshalb nicht alle Kandidatinnen und Kandidaten perfekt zugeordnet werden, liegt darin, dass in der Regressionsgleichung ausschliesslich lineare Zusammenhänge modelliert werden. In der Realität liegt perfekte Linearität aber nicht vor. Durch die zur Berechnung der Endnote verwendete Formel und die Regel, dass höchstens eine der Fachnoten in

Deutsch, Mathematik und Fremdsprachen unter 4.0 liegen darf, wird die perfekte Linearität verletzt.

Tabelle 27: Vierfeldertafel Modell 1 mit sämtlichen Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	123	5	Sensitivität = .96
nicht bestanden	5	68	Spezifität = .93

Tabelle 28 zeigt die Vorhersage des Prüfungsentscheids mit nur vier Prüfungsteilen, dem kürzesten Modell 2. Die Gesamttrefferquote liegt bei 95 Prozent und ist somit gleich hoch wie bei der bestmöglichen Vorhersage durch das vollständige Modell 1. Von den 201 Kandidatinnen und Kandidaten wird für 11 eine vom tatsächlichen Prüfungsentscheid abweichende Vorhersage getroffen. Mit einer verkürzten Prüfung, die nur die Kompetenzbereiche *Deutsch Schreiben*, *Englisch Sprache im Fokus, Französisch Sprache im Fokus* und *Mathematik* einbeziehen würde, könnte für 95 Prozent der Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, und für 95 Prozent, die die Prüfung nicht bestanden haben, der Prüfungsentscheid korrekt vorhergesagt werden.

Tabelle 28: Vierfeldertafel Modell 2 mit vier Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	121	7	Sensitivität = .95
nicht bestanden	4	69	Spezifität = .95

Die Vorhersagen in Tabelle 28 wurden unter der Annahme getroffen, dass falsch positive Fälle gleich gewichtet werden wie falsch negative Fälle. Das heisst, dass es als gleich bedeutend erachtet wird, wenn eine Kandidatin, die die Aufnahmeprüfung nicht bestanden hat, einen positiven Bescheid erhält, wie wenn ein Kandidat, der die Aufnahmeprüfung bestanden hat, einen negativen Bescheid erhält.

Tabelle 29 enthält wiederum ein Szenario, in dem falsch negative Fälle möglichst vermieden werden und die Sensitivität der verkürzten Prüfung bei mindestens .99 liegen soll. Wenn 99 Prozent der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung bestanden haben, auch unter verkürzten Bedingungen der Kategorie «bestanden» zugeordnet werden sollen, dann sinkt die Spezifität auf .84. 12 Kandidatinnen und Kandidaten würden fälschlicherweise der Kategorie «bestanden» zugeordnet.

Tabelle 29: Modell 2 mit vier Kompetenzbereichen und einer Sensitivität von .99

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	127	1	Sensitivität = .99
nicht bestanden	12	61	Spezifität = .83

Das Szenario, in dem mindestens 99 Prozent der Kandidatinnen und Kandidaten, die die Aufnahmeprüfung in die BM II nicht bestanden haben, auch unter verkürzten Bedingungen korrekt der Kategorie «nicht bestanden» zugeordnet würden, ist in Tabelle 30 dargestellt. Wenn 99 Prozent der Kandidatinnen und Kandidaten, die die Prüfung nicht bestanden haben, anhand der verkürzten Prüfung der Kategorie «nicht bestanden» zugeordnet werden sollen, dann werden 14 Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, fälschlicherweise der Kategorie «nicht bestanden» zugeordnet.

Tabelle 30: Modell 2 mit vier Kompetenzbereichen und einer Spezifität von .99

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	114	14	Sensitivität = .89
nicht bestanden	1	72	Spezifität = .99

In Tabelle 31 sind schliesslich die Ergebnisse der Vorhersage des Prüfungsentscheids mit jenen Kompetenzbereichen dargestellt, die in Modell 2 nicht einbezogen wurden, also die Ergebnisse von Modell 3. Die Gesamttrefferquote liegt bei 67 Prozent. 44 von 128 erfolgreichen Kandidatinnen und Kandidaten, die die Prüfungen bestanden haben, hätten fälschlicherweise einen negativen Prüfungsentscheid erhalten (Sensitivität = .66). Von den 73 Kandidatinnen und Kandidaten, die die Aufnahmeprüfung nicht bestanden haben, werden 50 korrekterweise der Gruppe «nicht bestanden» zugeordnet (Spezifität = .68). Dies zeigt noch einmal deutlich, dass die Bedeutung der Kompetenzbereiche *Englisch Lesen*, *Französisch Lesen* und *Französisch Schreiben* im Vergleich zu den Kompetenzbereichen *Deutsch Schreiben*, *Englisch Sprache im Fokus*, *Französisch Sprache im Fokus* und *Mathematik* sehr klein ist. Die Fehlallokation steigt von 10 (Modell 1) auf 11 (Modell 2) und schliesslich auf 67 Fälle (Modell 3).

Tabelle 31: Vierfeldertafel Modell 3 mit nur unbedeutenden Kompetenzbereichen

	Vorhersage		
Prüfung	bestanden	nicht bestanden	
bestanden	84	44	Sensitivität = .66
nicht bestanden	23	50	Spezifität = .68

Fazit

Für 93 Prozent der Kandidatinnen und Kandidaten für die BM I im Jahr 2019 lässt sich der Prüfungsentscheid anhand einer verkürzten Prüfung mit nur vier Kompetenzbereichen (*Deutsch Lesen*, *Deutsch Schreiben*, *Französisch Sprache im Fokus* und *Mathematik*) korrekt vorhersagen. Dieses Ergebnis zeigt, dass viele der in den Prüfungen erfassten Kompetenzbereiche für den Prüfungsentscheid insofern kaum von Bedeutung sind, als sie keine zusätzlichen Informationen für den Prüfungsentscheid liefern. Wird die verkürzte Prüfung durch die Kompetenzbereiche *Deutsch Sprache im Fokus*, *Englisch Lesen*, *Englisch Sprache im Fokus*, *Englisch Schreiben*, *Französisch Lesen* und *Französisch Schreiben* ergänzt, dann erhöht sich die Trefferquote nur unwesentlich.[6] Dies kann unter anderem dadurch zustande kommen, dass mit den verschiedenen Prüfungsteilen zum Teil ähnliche Kompetenzen, beispielsweise Schreibkompetenzen, gemessen werden und sich dadurch Redundanzen ergeben.

Dieses Ergebnis lässt sich anhand der Aufnahmeprüfung für die BM II im Jahr 2019 bestätigen. Die Unterscheidung zwischen Kandidatinnen und Kandidaten, die die Prüfung bestanden haben, und solchen, die sie nicht bestanden haben, hätte genauso gut getroffen werden können, wenn die Kompetenzbereiche *Englisch Lesen*, *Französisch Lesen* und *Französisch Schreiben* weggelassen worden wären. Für den Prüfungsentscheid von Bedeutung waren die Kompetenzbereiche *Deutsch Schreiben*, *Englisch Sprache im Fokus*, *Französisch Sprache im Fokus* und *Mathematik*.

Eine Schlussfolgerung davon könnte sein, die aktuellen Prüfungen deutlich schlanker zu gestalten, was den Prüfungsentscheid kaum verändern würde. Besser, als die bestehenden Prüfungen zu kürzen, wäre es jedoch, die Prüfungen so zu überarbeiten, dass sämtliche Kompetenzbereiche für den Prüfungsentscheid relevant werden.

Die Relevanz eines Prüfungs- bzw. Kompetenzbereichs könnte auf zwei Arten gesteigert werden. Zum einen könnte der jeweilige Kompetenzbereich ausführlicher geprüft werden. Durch eine Erhöhung der Anzahl Aufgaben bzw. eine ausführlichere Beurteilung der Schreibkompetenzen – indem die Mindestanzahl an Wörtern sowie die Anzahl an Beurteilungskriterien erhöht wird – wird die Varianz der Leistungen der Kandidatinnen und Kandidaten in einem Kompetenzbereich erhöht. Dadurch steigt in der Regel auch die Bedeutung dieses Bereichs. Zum anderen könnte die Bedeutung eines Kompetenzbereichs gesteigert werden, indem die darin enthaltenen Aufgaben so überarbeitet werden, dass sie besser zwischen schwächeren und stärkeren Kandidatinnen und Kandidaten unterscheiden (dazu mehr in Absatz 6.6).

Insbesondere die Lese- und Schreibkompetenzen in den Fremdsprachen werden nicht optimal erfasst bzw. sind sie für den Prüfungsentscheid nicht informationsbringend. Idealerweise werden die entsprechenden Kompetenzbereiche grundsätzlich überarbeitet und ergänzt. Sollte dies aus Ressourcengründen nicht möglich sein, wird es als sinnvoller erachtet, einen Kompetenzbereich ausführlich und valide zu prüfen und den anderen wegzulassen, anstatt zwei Kompetenzbereiche unzureichend zu prüfen.

Die insgesamt eher tiefe Bedeutung der Fachbereiche Englisch und Französisch, die sich besonders für die BM I zeigt, lässt sich teilweise auch mit der Formel zur Berechnung der Gesamtnote erklären. Die Noten in Englisch und Französisch werden erstens zu einer Fachnote Fremdsprachen zu-

---

[6] Das bedeutet nicht, dass diese Kompetenzbereiche einzeln betrachtet keine Bedeutung für den Prüfungsentscheid haben. Als Bestandteil der gesamten Prüfung sinkt ihre Bedeutung allerdings drastisch, da sie keine zusätzlichen Informationen liefern.

sammengefasst und zweitens wird diese auf halbe Noten gerundet, bevor sie in die Gesamtnote einfliesst (vgl. Kapitel 2). Das Gewicht von Englisch und Französisch auf die Endnote ist also zum einen nur halb so gross wie das Gewicht von Deutsch und Mathematik und zum anderen ist die Fachnote Fremdsprachen ungenauer als die Fachnoten Deutsch und Mathematik. Eine Anpassung der Formel zur Berechnung der Gesamtnote dürfte ebenfalls zu einer Verschiebung der Bedeutung einzelner Kompetenzbereiche für den Prüfungsentscheid führen.

6.4 Prüfungsteile und Beurteilungsmassstab

Die bisherigen Ergebnisse haben gezeigt, dass die einzelnen Prüfungsbestandteile für das Bestehen der Prüfung unterschiedlich wichtig sind. Kandidatinnen und Kandidaten mit ausreichenden Vorkenntnissen in Deutsch und Mathematik bzw. einer guten Vorbereitung auf diese Prüfungsteile erhöhen die Wahrscheinlichkeit, die Aufnahmeprüfung zu bestehen. Darüber hinaus sind Englisch und Französisch für das Bestehen der Prüfung kaum von Bedeutung.

Der Prüfungsentscheid kann aber auch davon abhängen, wie streng die Massstäbe bei den einzelnen Prüfungsteilen angesetzt werden. Die Erfahrungen zeigen, dass die Notenmassstäbe in den einzelnen Prüfungsteilen unterschiedlich sind, was die Frage aufwirft, ob diese Unterschiede gerechtfertigt sind oder eher auf einem Fachverständnis beruhen.

Zur Beurteilung der unterschiedlichen Beurteilungsmassstäbe werden die Ergebnisse im Check S3 beigezogen. Die Leistungstests führen zu einer unabhängigen Einschätzung der Fähigkeiten von Schülerinnen und Schülern und sind somit nicht durch eine Fremdbeurteilung verzerrt. Die Verteilung der Check-Ergebnisse entspricht dadurch einer Abbildung des tatsächlichen Leistungsstands der Schülerschaft am Ende der obligatorischen Schulzeit.

Der Vergleich der Verteilung der Ergebnisse im Check S3 mit der Verteilung der Prüfungsnoten ermöglicht deshalb eine Beurteilung der fachspezifischen Beurteilungsmassstäbe anhand eines externen Kriteriums. Für den Vergleich liegen genügend Fälle von Kandidatinnen und Kandidaten für die BM I und für die BM II Typ «andere» vor ($N = 147$ bzw. $N = 42$). Die Prüfungsnoten der Kandidatinnen und Kandidaten für die BM II Typ «Wirtschaft» konnten nicht mit den Check-Ergebnissen verknüpft werden.

BM I

Abbildung 3 zeigt die Notenverteilungen in den Aufnahmeprüfungen für die BM I nach Jahr und Fachbereich. Oben links ist der Notendurchschnitt vermerkt, die Prozentzahl oben rechts gibt an, wie hoch der Anteil der Noten 4.0 oder höher ist. Es zeigt sich, dass die Notenverteilungen in den einzelnen Fächern über die Jahre hinweg relativ konstant sind (die Notendurchschnitte bewegen sich innerhalb einer halben Note). Zwischen den Fächern sind zum Teil jedoch klare Unterschiede auszumachen. Die Differenz zwischen den Notendurchschnitten in Deutsch und Englisch beträgt in einem Jahr höchstens eine halbe Note, und die grosse Mehrheit der Noten liegt über 4.0, ist somit genügend. Die Differenz zwischen den Notendurchschnitten in Französisch und Mathematik beträgt höchstens eine halbe Note. Die Mehrheit der Noten liegt jedoch unter 4.0 im ungenügenden Bereich. In Französisch beträgt der Anteil an Noten unter 4.0 zwischen 16 und 39 Prozent, in Mathematik zwischen 28 und 46 Prozent.

Auch die Verteilungen der Noten in Deutsch und Englisch einerseits sowie Französisch und Mathematik andererseits unterscheiden sich. Die Noten in Deutsch und Englisch streuen enger um den Mittelwert als die Noten in Französisch und Mathematik. Insbesondere in Mathematik wird die gesamte Notenskala ausgenutzt. Über alle Verteilungen fällt die Mathematikprüfung von 2018

besonders auf, weil sie am unteren Ende der Skala kaum flacher wird. Die Note 1.0 wurde in diesem Jahr vergleichsweise oft vergeben.

Abbildung 3: Notenverteilung nach Fach und Jahr (BM I)

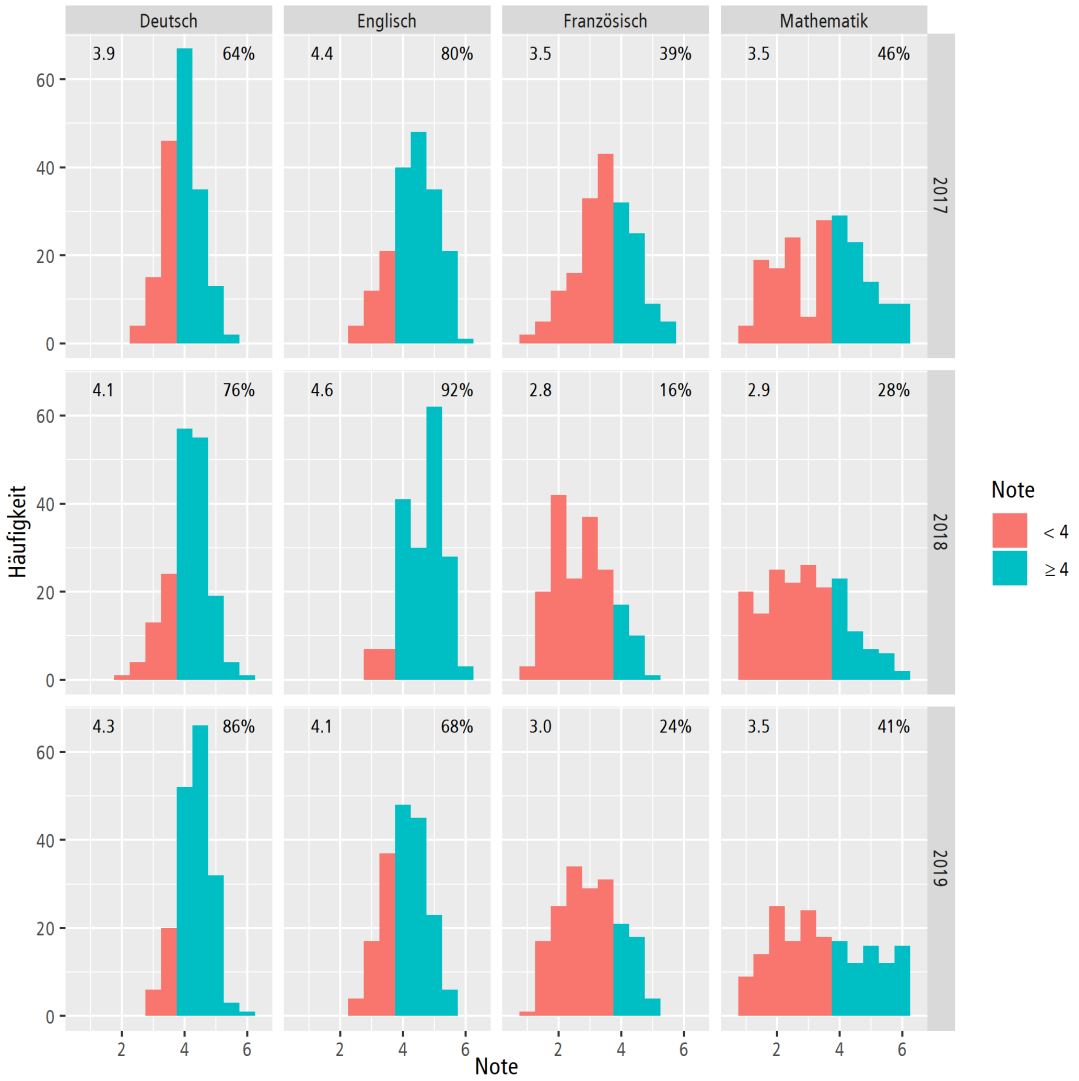
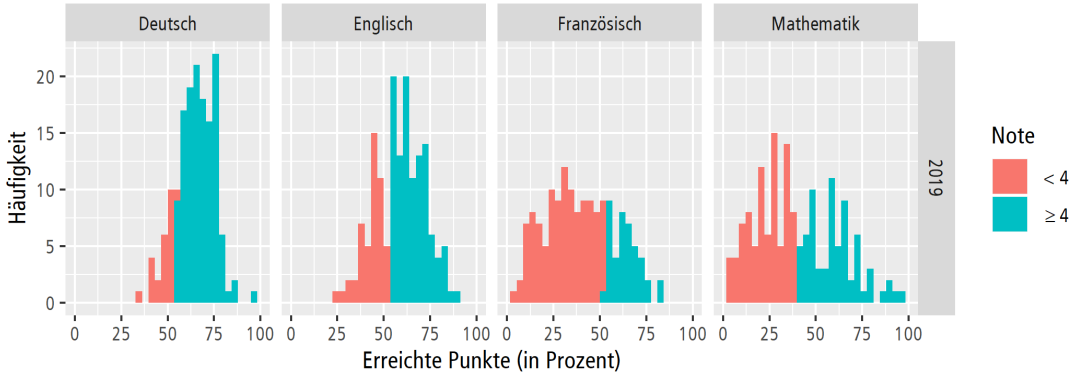


Abbildung 4 zeigt die Verteilungen der erreichten Punktzahlen in Prozent nach Fach in der Aufnahmeprüfung 2019 für die BM I. Für eine Note von 4.0 oder höher mussten in den sprachlichen Fächern etwas mehr als 50 Prozent der möglichen Punkte erreicht werden (55 Prozent in Deutsch, 54 Prozent in Englisch und Französisch). In Mathematik reichten dafür 43 Prozent der maximal erreichbaren Punktzahl aus. Auch bei der Verteilung der Punktzahlen zeigt sich, dass in Französisch und Mathematik die Streuung grösser ist als in Deutsch und Englisch.

Abbildung 4: Verteilung der Punktzahlen nach Fach im Jahr 2019 (BM I)



Um die Notenverteilungen der Aufnahmeprüfung für die BM I mit den Verteilungen der Check-Ergebnisse vergleichen zu können und so eine externe Einschätzung der Strenge bzw. Milde des Beurteilungsmassstabs zu erhalten, wurden die Prüfungsnoten und die Check-Punktzahlen in die z-Skala transformiert (z-standardisiert). Die z-Standardisierung ermöglicht den Vergleich zweier Verteilungen, die aufgrund von unterschiedlichen Skalen zustande gekommen sind. Die unterschiedlichen Rohwerte (Noten versus Punkte auf einer Skala von 0 bis 1200) werden durch die z-Transformation in standardisierte z-Werte gebracht, wodurch sie vergleichbar werden.

Tabelle 32 enthält die Mittelwerte und die Standardabweichungen der Prüfungsnoten und der Check-Ergebnisse in Form von z-Werten. In Abbildung 5 sind die Verteilungen der z-Werte grafisch dargestellt.

Tabelle 32: Prüfungsnoten und Check-Ergebnisse in Form von z-Werten: Mittelwerte und Standardabweichungen (BM I)

	2017		2018		2019	
	M	*SD*	*M*	*SD*	*M*	*SD*
BM I						
Deutsch	.18	.52	.35	.61	.54	.47
Englisch	.55	.70	.81	.58	.32	.63
Französisch	–.23	.85	–.91	.81	–.63	.87
Mathematik	–.15	1.19	–.62	1.16	–.22	1.36
Check S3						
Deutsch	–.16	.77	–.13	.75	–.09	.73
Englisch	.63	.89	.86	.80	.85	.99
Französisch	–.80	.91	–.82	.74	–.79	.85
Mathematik	.20	.82	.08	.79	.08	.83

M = Mittelwert, *SD* = Standardabweichung

Aus Tabelle 32 geht hervor, dass die Kandidatinnen und Kandidaten im Fach Französisch sowohl bei der Aufnahmeprüfung als auch im Check S3 im Vergleich zu den anderen Fachbereichen am

schlechtesten abschneiden.[7] Der Mittelwert im Check S3 bewegt sich zwischen $M = -.79$ und $M = -.82$ und ist folglich deutlich tiefer als der Gesamtmittelwert von $M = 0$. Der Mittelwert in den Aufnahmeprüfungen bewegt sich zwischen $M = -.23$ und $M = -.91$ und ist ebenfalls deutlich tiefer als der Gesamtmittelwert von $M = 0$. Ein im Vergleich zu den anderen Fachbereichen tieferer Notendurchschnitt in Französisch scheint damit gerechtfertigt zu sein.

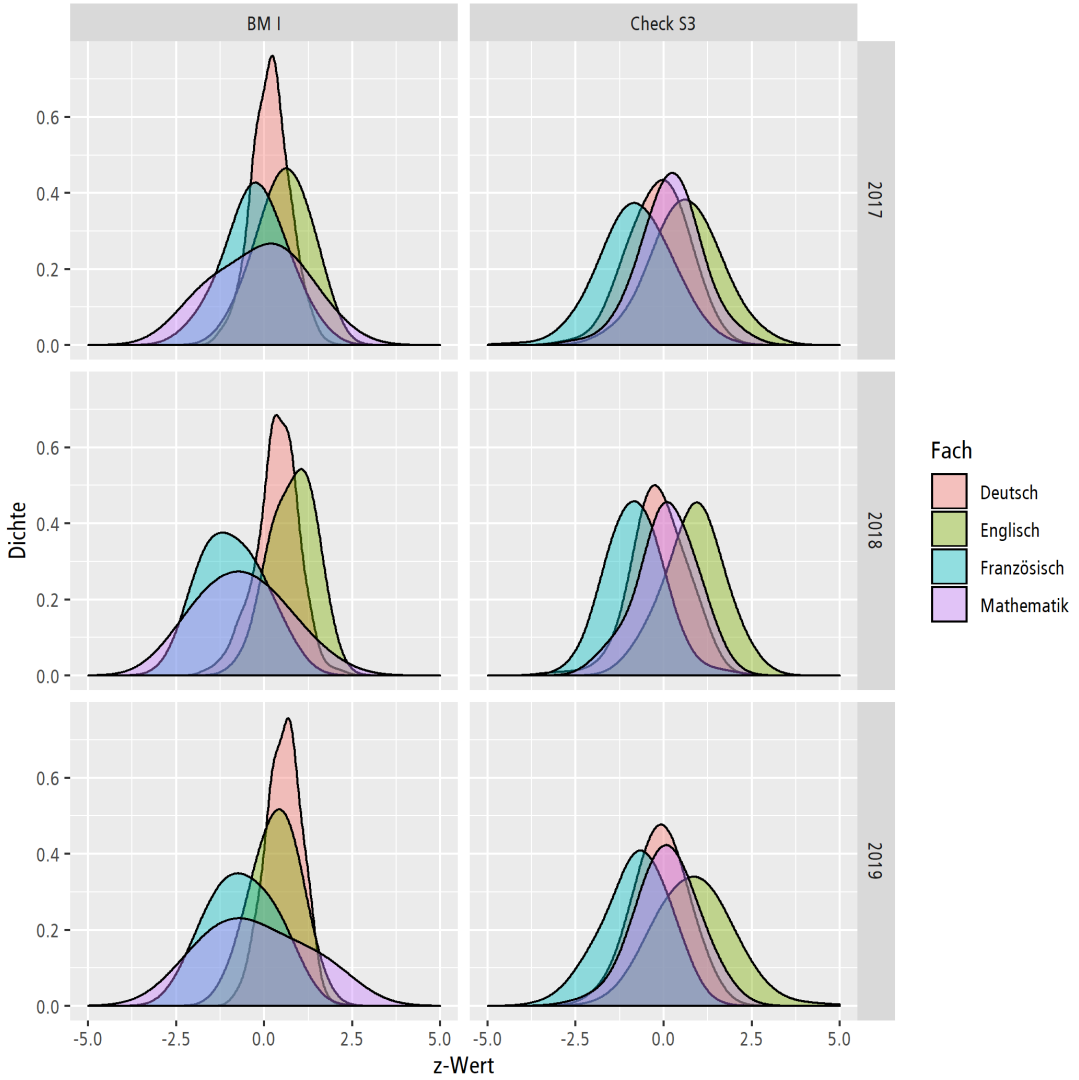
Am besten schneiden die Kandidatinnen und Kandidaten sowohl in den Aufnahmeprüfungen als auch im Check S3 in Englisch ab (mit Ausnahme der Aufnahmeprüfung 2019). Der Mittelwert im Check S3 bewegt sich zwischen $M = .63$ und $M = .86$ und ist folglich deutlich höher als der Gesamtmittelwert von $M = 0$. Der Mittelwert in den Aufnahmeprüfungen bewegt sich zwischen $M = .32$ und $M = .81$ und ist ebenfalls deutlich höher als der Gesamtmittelwert von $M = 0$. Ein vergleichsweise hoher Notendurchschnitt in Englisch scheint somit ebenfalls gerechtfertigt zu sein.

Die Mittelwerte in Deutsch liegen im Check S3 zwischen $M = -.09$ und $M = -.16$, also knapp unter dem Gesamtmittelwert von $M = 0$. Die Mittelwerte in den Aufnahmeprüfungen Deutsch liegen zwischen $M = .18$ und $M = .54$, also über dem Gesamtmittelwert von $M = 0$. Diese Differenzen sind ein Hinweis dafür, dass der Beurteilungsmassstab bei der Aufnahmeprüfung Deutsch eher etwas zu milde ausfällt. Allerdings trifft dies vor allem für die Aufnahmeprüfung 2019 zu.

Die Mittelwerte in Mathematik liegen im Check S3 zwischen $M = .08$ und $M = .20$, also knapp über dem Gesamtmittelwert von $M = 0$. Die Mittelwerte in den Aufnahmeprüfungen Mathematik liegen zwischen $M = -.15$ und $M = -.62$, also unter dem Gesamtmittelwert von $M = 0$. Diese Differenzen sind ein Hinweis dafür, dass der Beurteilungsmassstab bei der Aufnahmeprüfung Mathematik eher etwas zu streng ausfällt. Allerdings trifft dies in Mathematik vor allem für die Aufnahmeprüfung 2018 zu.

[7] Die Vergleiche der Mittelwerte der z-standardisierten Noten in Tabelle 32 können von den Vergleichen der Notendurchschnitte in Abbildung 3 abweichen, da sich die zugrundeliegende Population unterscheidet. Für die Berechnungen in Tabelle 32 wurden lediglich Fälle eingeschlossen, für die auch die Check-S3-Ergebnisse vorliegen. Für die Berechnungen in Abbildung 3 wurde hingegen die Gesamtheit der Prüfungsteilnehmenden berücksichtigt.

Abbildung 5: Verteilung der Prüfungsnoten und der Check-Ergebnisse in Form von z-Werten nach Fach und Jahr (BM I)



BM II Typ «andere»

Die Verteilungen der Noten der Aufnahmeprüfungen für die BM II Typ «andere» nach Fachbereich und Jahr sind in Abbildung 6 dargestellt. Auch hier zeigt sich innerhalb der Fächer über die Jahre hinweg eine Konstanz im Notendurchschnitt und im Anteil an Noten von mindestens 4.0. Wie in der BM I liegen auch hier die Notendurchschnitte in Deutsch und Englisch im genügenden, in Französisch und Mathematik im ungenügenden Bereich (mit den Ausnahmen von Englisch im Jahr 2017 und Mathematik im Jahr 2019). Am höchsten ist der Anteil genügender Noten in Deutsch, gefolgt von Englisch, Mathematik und schliesslich Französisch.

Auffallend ist der Vergleich der Mathematik-Prüfungen von 2018 und 2019. Während die Noten 1.0 und 1.5 2018 relativ häufig vergeben wurden, tauchen sie 2019 kaum auf. Dafür wurde die Note 6.0 in Mathematik 2019 deutlich häufiger erreicht als 2018. Diese Unterschiede können entweder durch unterschiedliche Beurteilungsmassstäbe oder aber durch die Population von Kandidatinnen und Kandidaten erklärt werden. Aufgrund der vorliegenden Daten lässt sich nicht nachwei-

sen, ob es sich bei den jeweiligen Kandidatinnen und Kandidaten bezüglich des Vorwissens um vergleichbare Populationen handelt.

Abbildung 6: Notenverteilung nach Fach und Jahr (BM II Typ «andere»)

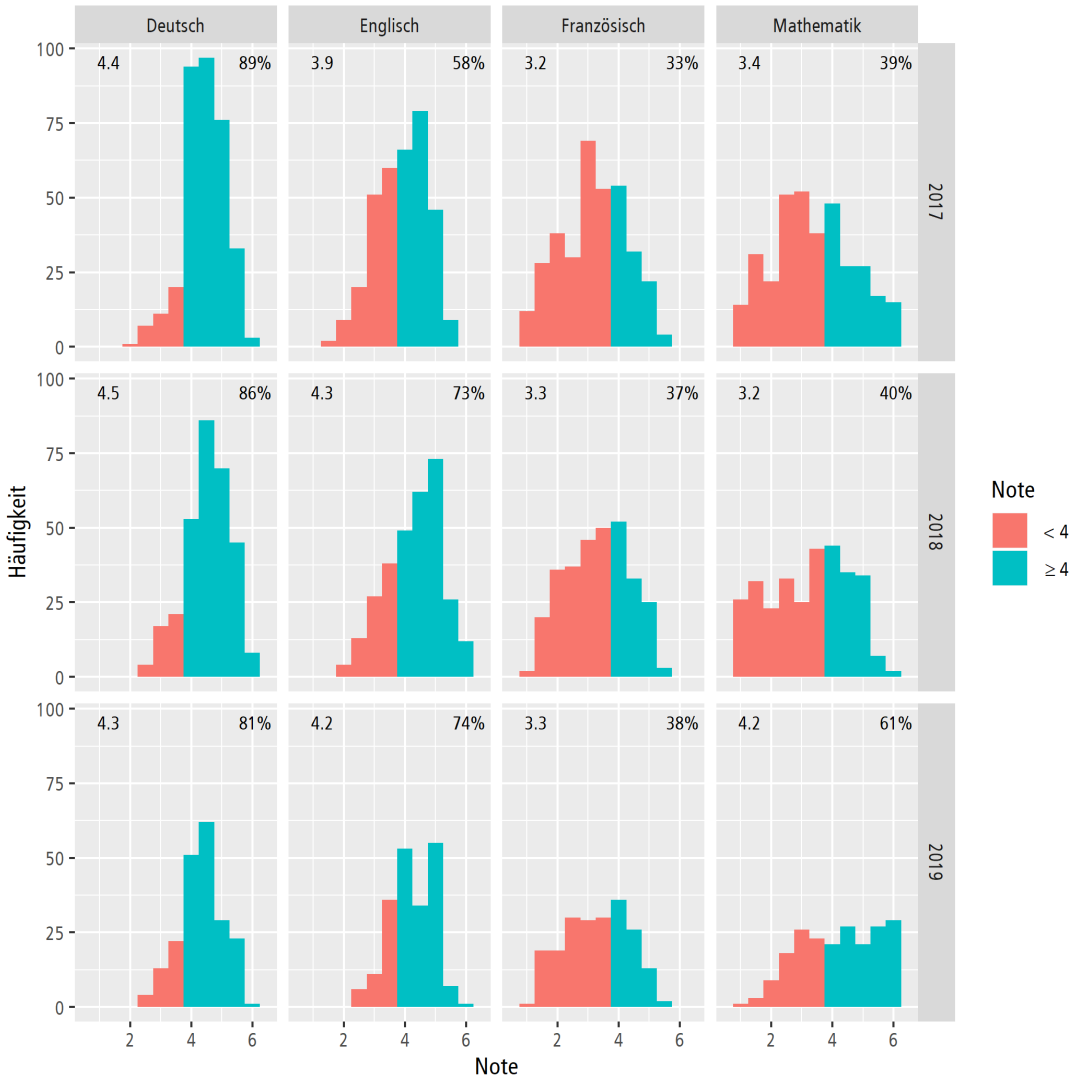
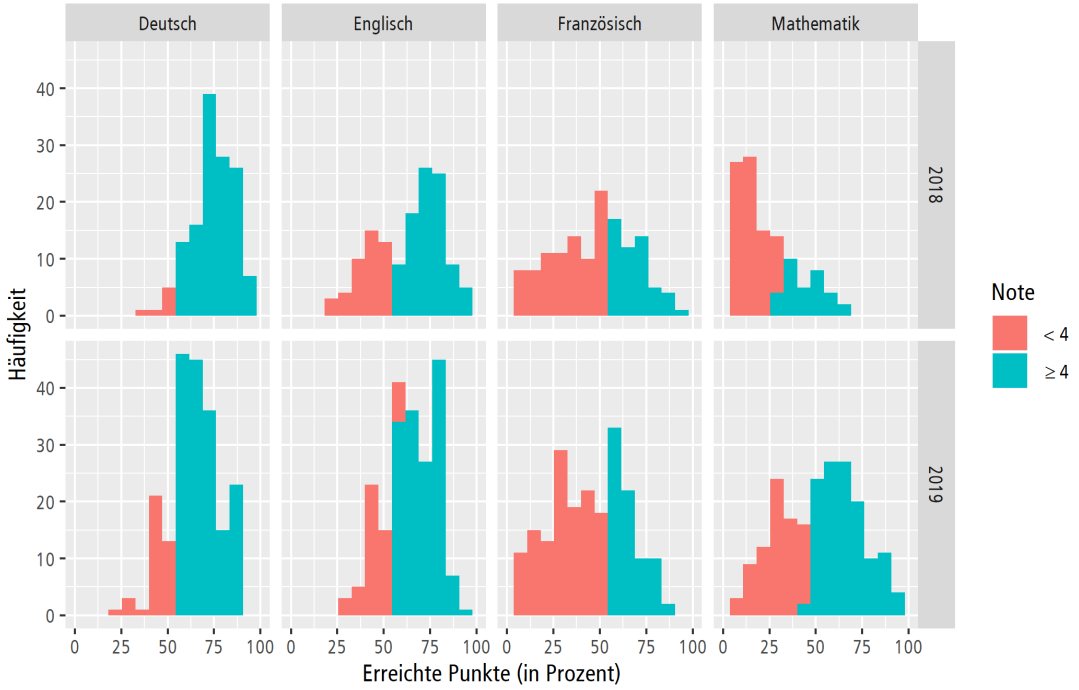


Abbildung 7 zeigt die Verteilungen der erreichten Punktzahlen in Prozent nach Fach und Jahr. In den sprachlichen Fächern ist wiederum etwas mehr als die Hälfte (rund 55 Prozent) der maximalen Punktzahl notwendig, um die Note 4.0 zu erreichen. In Mathematik sind gemäss den Musterlösungen der Jahre 2018 und 2019 für die Note 4.0 45 Prozent der Maximalpunktzahl notwendig. Im Jahr 2018 scheint es jedoch zu einer Anpassung der Skala gekommen zu sein. Die Note 4.0 wurde bereits mit 33 Prozent der maximalen Punktzahl erreicht (6.5 von 20 möglichen Punkten).

Abbildung 7: Verteilung der Punktzahlen nach Fach und Jahr (BM II Typ «andere»)



Um die Notenverteilungen der Aufnahmeprüfung für die BM II mit den Verteilungen der Check-Ergebnisse vergleichen zu können und so eine externe Einschätzung der Strenge bzw. Milde des Beurteilungsmassstabs zu erhalten, wurden die Prüfungsnoten und die Check-Punktzahlen in die z-Skala transformiert (z-standardisiert).

Tabelle 33 enthält die Mittelwerte und die Standardabweichungen der Prüfungsnoten und der Check-Ergebnisse in Form von z-Werten. In Abbildung 8 sind die Verteilungen der z-Werte grafisch dargestellt.

Tabelle 33: Prüfungsnoten und Check-Ergebnisse in Form von z-Werten: Mittelwerte und Standardabweichungen (BM II)

	BM II 2019		Check S3 2016	
	M	*SD*	*M*	*SD*
Deutsch	.60	.58	–.06	.76
Englisch	.10	.72	.62	.92
Französisch	–.66	.93	–.76	1.01
Mathematik	–.03	1.24	.08	.86

M = Mittelwert, *SD* = Standardabweichung

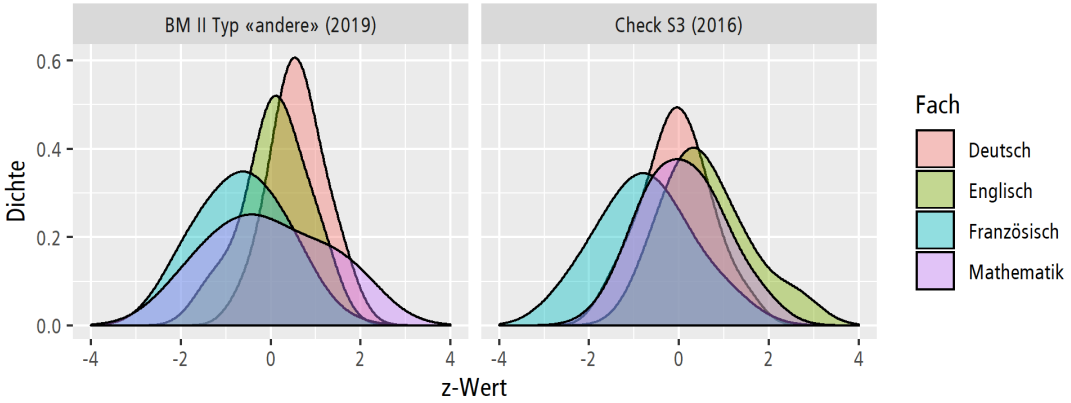
Auch für die BM II zeigt sich, dass die Kandidatinnen und Kandidaten im Fach Französisch sowohl bei der Aufnahmeprüfung als auch im Check S3 im Vergleich zu den anderen Fächern am schlech-

testen abschneiden.[8] Der Mittelwert in der Aufnahmeprüfung Französisch liegt bei $M = -.66$, jener im Check S3 bei $M = -.76$. Ein im Vergleich zu den anderen Fachbereichen tieferer Notendurchschnitt in Französisch scheint gerechtfertigt zu sein.

Wie für die BM I zeigt sich auch für die BM II, dass die Kandidatinnen und Kandidaten in der Aufnahmeprüfung in Deutsch deutlich besser abschneiden als in Mathematik. Für die Check-Ergebnisse gilt dies nicht. Hier schneiden die Kandidatinnen und Kandidaten in Mathematik leicht besser ab als in Deutsch. Der Mittelwert in der Aufnahmeprüfung Deutsch liegt bei $M = .60$, jener im Check S3 bei $M = -.06$, der Mittelwert in der Aufnahmeprüfung Mathematik liegt bei $M = -.03$, jener im Check S3 bei $M = -.08$. Ausgehend von den Ergebnissen im Check S3 als externem Massstab ist der höhere Notendurchschnitt in Deutsch im Vergleich zur Mathematik nicht gerechtfertigt. Der Notendurchschnitt in Mathematik ist demgegenüber aufgrund der Check-Ergebnisse gerechtfertigt. Der eher strenge Beurteilungsmassstab in Englisch lässt sich aufgrund der Check-Ergebnisse eher nicht rechtfertigen. Der Mittelwert in der Aufnahmeprüfung Englisch liegt bei $M = .10$, jener im Check S3 bei $M = .62$.

In Abbildung 8 sind die Verteilungen der Prüfungsnoten und der Check-Ergebnisse dargestellt. Während die Prüfungsnoten Mathematik sehr breit um den Mittelwert streuen, liegen die Noten in Deutsch vergleichsweise nahe beim Mittelwert. Die Notenskala wird vor allem in der Mathematik ausgenutzt, was auf eine gute Diskrimination der Prüfung hinweist.

Abbildung 8: Verteilung der Prüfungsnoten 2019 und der Check-Ergebnisse 2016 in Form von z-Werten nach Fach (BM II, Typ «andere»)



BM II Typ «Wirtschaft»

Abbildung 9 zeigt die Notenverteilungen der Aufnahmeprüfungen für die BM II Typ «Wirtschaft» nach Fach und Jahr. Die Notendurchschnitte in Deutsch, Französisch sowie Finanz- und Rechnungswesen sind über die Jahre hinweg einigermassen konstant. In Englisch schwankt der Notendurchschnitt über die Jahre hinweg etwas stärker. In allen drei Jahren ist der Notendurchschnitt im Fach Finanz- und Rechnungswesen am tiefsten und in Englisch mit Ausnahme von 2017 am höchsten. Die Streuung der Noten um den Mittelwert ist im Fach Deutsch über alle drei Jahre hin-

[8] Die Vergleiche der Mittelwerte der z-standardisierten Noten in Tabelle 33 können von den Vergleichen der Notendurchschnitte in Abbildung 6 abweichen, da sich die zugrundeliegende Population unterscheidet. Für die Berechnungen in Tabelle 33 wurden lediglich Fälle eingeschlossen, für die auch die Check-S3-Ergebnisse vorliegen. Für die Berechnungen in Abbildung 6 wurde hingegen die Gesamtheit der Prüfungsteilnehmenden berücksichtigt.

weg am geringsten. Sehr tiefe (< 2.5) und sehr hohe Noten (> 5.0) werden in Deutsch nicht vergeben. In Englisch werden sehr tiefe Noten (< 2.5) kaum vergeben. Die Notenbandbreite wird in Französisch sowie im Fach Finanz- und Rechnungswesen stärker genutzt. Die Note 6.0 wurde über alle drei Jahre in keinem Fachbereich vergeben.

Abbildung 9: Notenverteilung nach Fach und Jahr (BM II Typ «Wirtschaft)

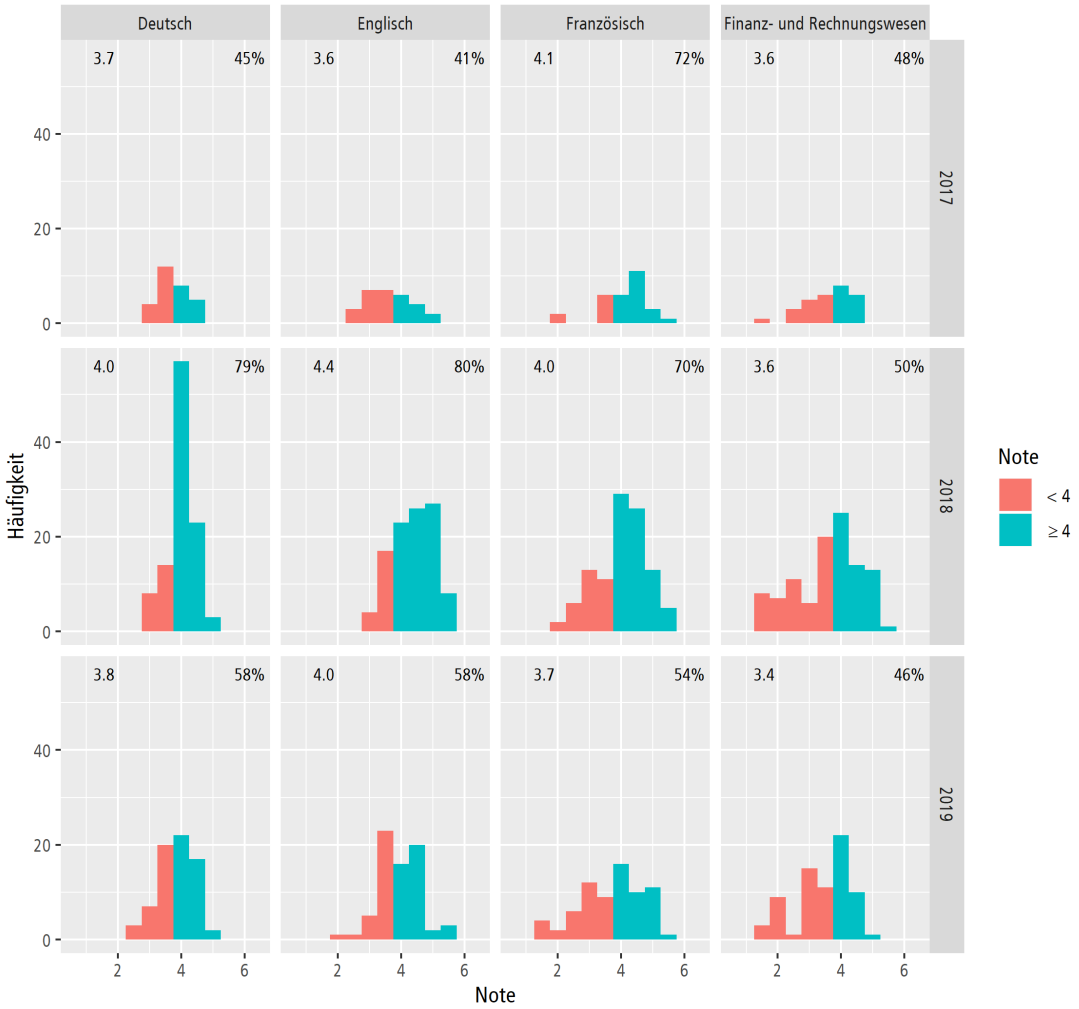
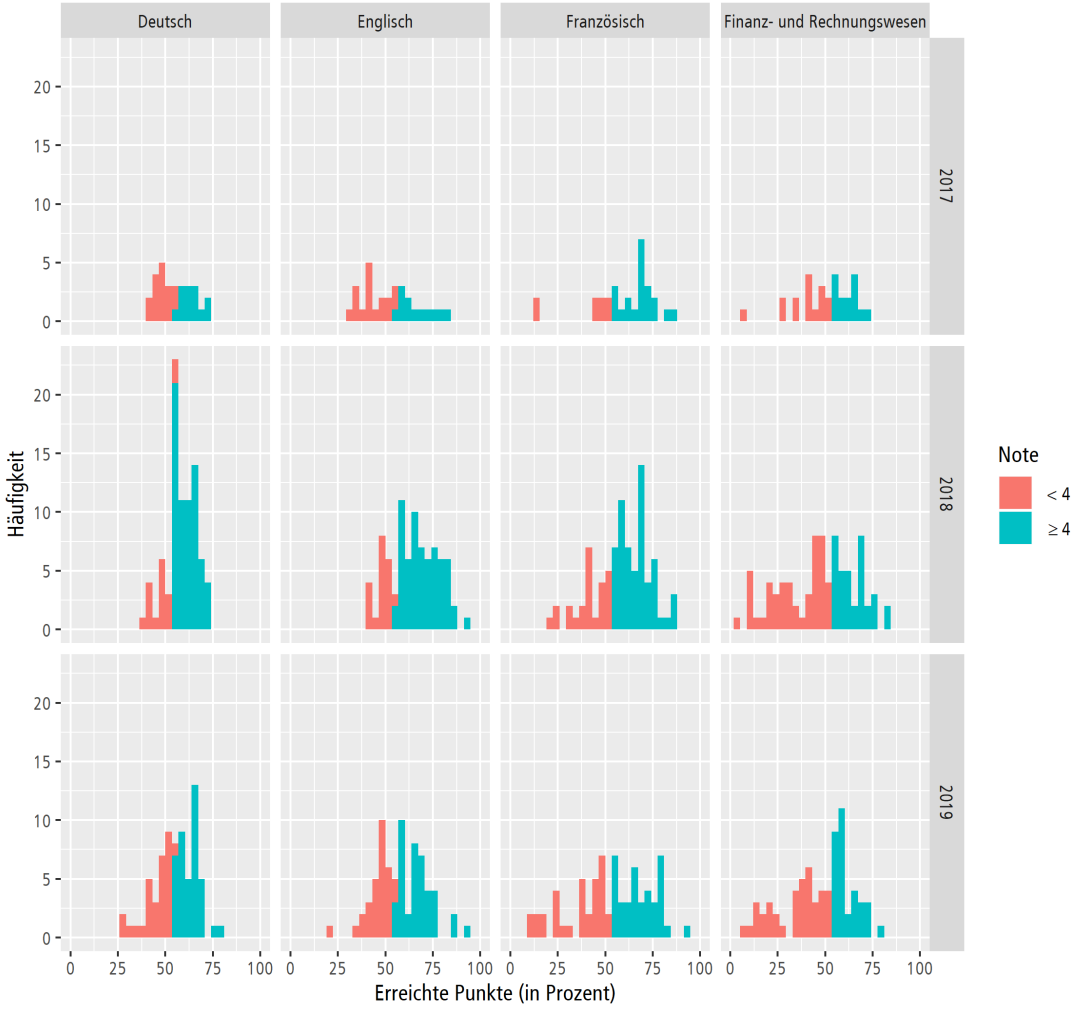


Abbildung 10 zeigt die Verteilungen der erreichten Punkte nach Fach und Jahr. Die Zuordnung von Punktzahlen zu Noten erfolgt in allen Fächern ähnlich. Für die Note 4.0 müssen rund 50 Prozent der maximal erreichbaren Punkte erreicht werden.

Abbildung 10: Verteilung der Punktzahlen nach Fach und Jahr (BM II Typ «Wirtschaft»)



Fazit

Beurteilungsmassstäbe basieren auf dem Fachverständnis bzw. sind sie das Ergebnis von fachspezifischen Anforderungen, die von den Kandidatinnen und Kandidaten erwartet werden. Aufgrund der Ergebnisse im Check S3, der zu einer unabhängigen Abbildung des tatsächlichen Leistungsstands der Schülerschaft am Ende der obligatorischen Schulzeit führt, kann überprüft werden, inwiefern die unterschiedlichen Beurteilungsmassstäbe aufgrund der Vorkenntnisse in der Schülerschaft gerechtfertigt – und somit nicht fachspezifisch – sind.

Die unterschiedlichen Beurteilungsmassstäbe scheinen zum grossen Teil gerechtfertigt zu sein. In der Tendenz decken sich die Verteilungen der Prüfungsnoten mit den Verteilungen der tatsächlichen Leistungen. Die vergleichsweise tiefen Notendurchschnitte in Französisch und die vergleichsweise hohen Notendurchschnitte in Englisch in den Prüfungen für die BM I werden aufgrund der Ergebnisse im Check S3 bestätigt. Etwas grösser sind die Abweichungen zwischen den Ergebnissen im Check S3 und den Prüfungsnoten in Deutsch und Mathematik. Allerdings kann die eher zu milde bzw. zu strenge Beurteilung in diesen Fächern jeweils nur für die Prüfung in einem Jahr nachgewiesen werden. Die Befunde für die BM I zeigen sich in ähnlicher Weise auch für die BM II Typ «andere». Insgesamt lassen sich die unterschiedlichen Beurteilungsmassstäbe aber durch die Ergebnisse im Check S3 relativ gut rechtfertigen. Insbesondere der tiefe Notendurch-

schnitt in Französisch ist aufgrund der tatsächlichen Leistungen in Französisch gerechtfertigt, ebenso der eher tiefe Notendurchschnitt in Mathematik. Eher etwas zu milde fällt die Beurteilung in Deutsch aus, eher etwas zu streng in Englisch.

Da sich die Beurteilungsmassstäbe im Grossen und Ganzen als gerechtfertigt erweisen, ist bei der Notengebung aufgrund dieser Analyse kein Handlungsbedarf angezeigt.

6.5 Gütekriterien der Aufnahmeprüfungen

Für die Überprüfung der wichtigsten Testgütekriterien Objektivität, Reliabilität und Validität wurden zum einen die Prüfungsaufgaben und Musterlösungen genutzt. Diese lagen von den Prüfungen der Jahre 2017 bis 2019 für alle Ausrichtungen ausser für BM II Typ «Wirtschaft» vor, für den die Prüfungsaufgaben und Musterlösungen der Jahre 2018 und 2019 genutzt werden konnten. Zum anderen wurden die von den Kandidatinnen und Kandidaten bearbeiteten Prüfungsaufgaben genutzt und die Bewertungen der einzelnen Aufgaben (richtig, falsch) elektronisch erfasst. Von 157 Kandidatinnen und Kandidaten konnten die Prüfungen für die BM I 2019 und von 205 Kandidatinnen und Kandidaten konnte die Prüfung für die BM II Typ «andere» genutzt werden. Für die BM II Typ «Wirtschaft» lagen keine bearbeiteten Prüfungen vor. Die so erfassten Rohdaten konnten mit den Check-Ergebnissen 2019 verknüpft werden. Insgesamt konnten aus dem Jahr 2019 Rohdaten und Check-Ergebnisse von 147 Kandidatinnen und Kandidaten für die BM I und von 42 Kandidatinnen und Kandidaten für die BM II Typ «andere» genutzt werden.

Objektivität

Tabelle 34 bietet einen Überblick über die Lösungsschlüssel der unterschiedlichen BM-Typen nach Fach und Jahr. Jede Aufnahmeprüfung wurde anhand folgender Kriterien beurteilt:

(1) Ist eine Musterantwort gegeben?
(2) Wird für jede Aufgabe aufgeschlüsselt, wie viele Punkte bei einzelnen Teilaufgaben erreicht werden können?
(3) Werden bei offenen Fragen Alternativen aufgelistet?
(4) Werden häufige Fehler, und wie mit ihnen umzugehen ist, beschrieben?
(5) Gibt es Hilfestellungen zur Bewertung der Schreibkompetenzen?
(6) Gibt es ein Kriterienraster zur Bewertung der Schreibkompetenzen?
(7) Ist die Transformation von der Punkte- auf die Notenskala eindeutig?

Ein Häkchen bedeutet, dass das entsprechende Kriterium erfüllt ist. Teilweise erfüllte Kriterien werden mit einer geschwungenen Linie gekennzeichnet. Kriterien, die als nicht erfüllt eingestuft werden, bleiben leer. Die grauen Flächen bedeuten, dass die entsprechenden Kriterien auf die jeweiligen Prüfungen nicht angewendet werden können.

Über alle BM-Typen hinweg lässt sich festhalten, dass sich die Formulierungen der Lösungsschlüssel zwischen den Fächern unterscheiden, innerhalb der Fächer über die drei Jahre hinweg jedoch relativ konstant sind. Besonders ausführlich und hinsichtlich der überprüften Kriterien grösstenteils vollständig sind die Lösungsschlüssel der Mathematikprüfungen.

Die Minimalanforderung, dass für jede Aufgabe – mit Ausnahme der Schreibaufträge – eine Musterlösung gegeben ist, wird von allen Prüfungen erfüllt.

In den meisten Fällen wird für jede Teilaufgabe angegeben, wie viele Punkte erreicht werden können. Ausnahmen bilden die Englischprüfungen für die BM II (Typ «andere» und Typ «Wirtschaft») sowie die Deutschprüfung für die BM II Typ «Wirtschaft». In diesen Fällen müssen die Korrektorin-

nen und Korrektoren die für jede Teilaufgabe zu vergebende Punktzahl von der Punktzahl, die für die gesamte Aufgabe vergeben wird, ableiten. Besteht eine Aufgabe beispielsweise aus vier Teilaufgaben und werden für die gesamte Aufgabe 4 Punkte vergeben, so muss entsprechend daraus geschlossen werden, dass für jede Teilaufgabe 1 Punkt zu vergeben ist. Aus dem Lösungsschlüssel ist dies allerdings nicht klar ersichtlich. In der Englischprüfung für die BM II Typ «Wirtschaft» ist zudem unklar, wie sich die 20 Punkte, die im Teil *Reading Comprehension* erreicht werden können, auf die 16 zugehörigen Aufgaben aufteilen.

Tabelle 34: Ausführlichkeit der Lösungsschlüssel nach Ausbildungsweg, Fachbereich und Jahr

		Ausbildungsweg											
		BM I				BM II Typ «andere»				BM II Typ «Wirtschaft»			
	Jahr	D	E	F	M	D	E	F	M	D	E	F	FRW
Musterantwort	2017	✓	✓	✓	✓		✓	✓	✓				
	2018	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
	2019	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Punktzahl	2017	✓	✓		✓	✓		✓	✓				
	2018	✓	✓	✓	✓	✓		✓	✓				✓
	2019		✓	~	✓	✓		✓	~			✓	~
Lösungsalternativen	2017	✓	✓		✓		✓	✓	✓				
	2018	✓	✓	✓	✓		✓	✓	✓	✓	✓		
	2019	✓	✓	✓	✓		✓	✓	✓	✓	✓		
Häufige Fehler	2017		✓				✓	✓					
	2018		✓	✓			✓	✓					
	2019	✓	✓				✓	✓					
Hilfestellungen	2017		✓					~					
	2018		✓					~					
	2019		✓					~					
Kriterienraster	2017												
	2018												
	2019												
Transformation	2017	✓	✓		✓		✓		✓				
	2018	✓	✓	✓	✓		✓		✓				
	2019		✓	✓	✓		✓		✓				

Bei Teilaufgaben, bei denen mehrere Antwortalternativen als richtig gewertet werden, werden diese in den meisten Fällen ebenfalls aufgelistet. Aus den Lösungsschlüsseln geht allerdings nicht immer klar hervor, ob die Liste der Antwortalternativen abschliessend ist. Bei der Durchsicht der bearbeiteten und bewerteten Aufnahmeprüfungen sind Einzelfälle aufgefallen, bei denen Punkte für Antwortalternativen vergeben wurden, die von anderen Korrektorinnen oder Korrektoren mit 0 Punkten bewertet wurden. Daraus wird geschlossen, dass es nicht zu einem berufsschulübergreifenden Austausch zwischen allen Examinatorinnen und Examinatoren kommt.

Häufig gemachte Fehler und die Anweisungen, wie damit umzugehen ist, werden in allen Lösungsschlüsseln der Prüfungen für die BM II Typ «andere» – bei den Aufgaben, in denen dieses

Kriterium relevant ist – aufgelistet. In den Lösungsschlüsseln der Prüfungen für die BM I werden sie in Deutsch und Französisch teilweise und in Englisch durchgehend aufgelistet. In den Lösungsschlüsseln der Prüfungen für die BM II Typ «Wirtschaft» werden häufige Fehler (und wie damit umzugehen ist) nicht erwähnt.

Für die Beurteilung der Schreibaufträge gibt es einzig für die Englischprüfung für die BM I eine Hilfestellung zur Bewertung. Dort wird eine mögliche Musterlösung gezeigt und erklärt, wie die Punktzahl auf der Notenskala von 1 bis 6 einzuordnen ist. Was bei dieser Hilfestellung allerdings fehlt, ist, wie mit zu kurzen Texten umzugehen ist. Eine Mindestlänge des Textes wird hier sowie auch bei allen anderen Schreibaufträgen für die Kandidatinnen und Kandidaten zwar sichtbar angegeben. Wie mit zu kurzen Texten umzugehen ist, wird allerdings nur in der Französischprüfung für die BM II Typ «andere» spezifiziert.

Die Kriterien, anhand derer ein Text zu bewerten ist, sind bei allen Prüfungen aufgelistet. Ein Kriterienraster, das für die zu bewertenden Kriterien ein ordinales Bewertungsschema definiert, fehlt jedoch bei allen Lösungsschlüsseln. Ein Beispiel eines Kriterienrasters (aus dem Check S3) kann in Anhang A eingesehen werden.

Als letzter Punkt wurden die Lösungsschlüssel bezüglich ihrer Eindeutigkeit bei der Transformation der Punktzahl auf die Notenskala beurteilt. Für die Transformation wird in einer Tabelle angegeben, welche Note mit welcher Punktzahl erreicht wird. In der Regel wird jeder Note ein Intervall an Punkten zugeordnet. In Einzelfällen – Deutschprüfung 2019 für die BM I, Französischprüfung 2017 für die BM I sowie die Französischprüfungen von 2017 bis 2019 für die BM II Typ «andere» – ist die Umrechnung nicht eindeutig. Und zwar wird jeweils jeder Note ein ganzzahliges Punktintervall zugeordnet, obwohl halbe Punkte vergeben werden können. Wird für 36 bis 42 Punkte beispielsweise die Note 4.0 und für 43 bis 48 Punkte die Note 4.5 vergeben, ist unklar, welche Note mit 42.5 Punkten vergeben wird. Bei der Deutschprüfung für die BM II Typ «andere» und bei den Prüfungen für die BM II Typ «Wirtschaft» ist in den Lösungsschlüsseln keine Umrechnung der Punkte- auf die Notenskala angegeben. Für diese Prüfungen können diesbezüglich also keine Angaben gemacht werden.

Die Deutschprüfung für die BM I 2019 enthält zwei kleine Fehler in der Musterlösung. Diese wurden bei der Korrektur von einer Mehrheit der Korrektorinnen und Korrektoren bemerkt und entsprechend korrigiert. Von einem Korrektoren-Team wurden die Fehler nicht bemerkt und die Aufgaben dementsprechend falsch korrigiert.

Bei der Durchsicht der bearbeiteten Prüfungen ist das Nichtbefolgen des Lösungsschlüssels bei einigen Aufgaben durch bestimmte Korrektorinnen und Korrektoren in Einzelfällen, aber auch in systematischer Weise aufgefallen. Folgendes wurde beobachtet:

- Multiple-Choice-Aufgaben, bei denen die falsche Antwortoption als korrekt bewertet wurde.
- Teilpunkte, die gemäss Musterlösung für unvollständige Zwischenschritte vergeben wurden.
- Offene Aufgaben, für die halbe Punkte vergeben wurden, obwohl im Lösungsschlüssel explizit vermerkt wurde, dass keine halben Punkte vergeben werden dürfen.

In seltenen Fällen gibt es Hinweise, dass das Vier-Augen-Prinzip allenfalls nicht eingehalten und Prüfungen nicht von zwei Examinatoren begutachtet wurden. In 3 Fällen fehlt die Signatur des ersten und in 26 Fällen die Signatur des zweiten Examinators. Auch scheint das Vier-Augen-Prinzip falschen Korrekturen nur teilweise entgegengewirkt zu haben, was zum einen die Beispiele oben und zum anderen der Anteil falsch zusammengezählter Punkte (vgl. Tabelle 35) zeigen.

In Tabelle 35 wird dargestellt, in wie viel Prozent der Fälle die total erreichte Punktzahl falsch zusammengezählt wurde und in wie viel Prozent der Fälle dies zu einer falschen Benotung geführt hat.

Bei der Deutschprüfung 2019 für die BM I zeigt sich mit 45 Prozent ein sehr hoher Anteil falsch zusammengezählter Punkte. Hierbei handelt es sich in über der Hälfte der Fälle allerdings um eine minimale Abweichung von einem halben Punkt, wobei davon ausgegangen wird, dass einige Korrektorinnen und Korrektoren halbe Punkte standardmässig aufgerundet haben. Die nicht eindeutige Transformation von der Punkte- auf die Notenskala unterstützt diese Annahme. Durch das Aufrunden können Grenzfälle von vornherein umgangen werden. In den Lösungsschlüsseln von 2017 und 2018 wird zudem festgehalten, dass halbe Punkte aufzurunden sind. Es ist anzunehmen, dass Korrektorinnen und Korrektoren diese Anweisung auch im Jahr 2019 befolgt haben, obwohl dort dazu nichts explizit geschrieben steht. Nichtsdestotrotz hat das falsche Zusammenzählen der Punkte für zehn Prozent der Kandidatinnen und Kandidaten zu einer falschen Note geführt.

Bei der Englischprüfung 2019 wurden die Punkte bei 13 Prozent der Prüfungen falsch zusammengezählt. In vier Prozent der Fälle hat dies zu einer inkorrekten Note geführt, auch wenn sich die Abweichungen lediglich zwischen einem halben und zwei Punkten bewegen.

Auch bei den Prüfungen der Fächer Französisch und Mathematik bewegen sich die Abweichungen auf geringem Niveau und die inkorrekte Gesamtpunktzahl hat nur selten zu einer falschen Notengebung geführt. Für keine Kandidatin und für keinen Kandidaten hat eine falsch zusammengezählte Punktzahl zu einem Fehlentscheid bezüglich der Aufnahme in die Berufsmaturitätsschule geführt.

Tabelle 35: Anteil falsch zusammengezählter Punkte und inkorrekter Benotung in der Aufnahmeprüfung von 2019 nach BM-Typ und Fach

	Korrekturfehler bei Punkteverteilung			
	BM I		BM II Typ «andere»	
	falsch zusammengezählt	inkorrekte Benotung	falsch zusammengezählt	inkorrekte Benotung
Deutsch	45%	10%	0%	0%
Englisch	13%[1]	4%	4%	1%
Französisch	13%	5%	21%	18%
Mathematik	6%	6%	2%	0%

[1] Bei 8 Prozentpunkten handelt es sich um falsch bewertete MC-Aufgaben und somit nicht im engen Sinne um falsch zusammengezählte Punkte.

Bei den Prüfungen für die BM II Typ «andere» wurden die Punkte im Fach Deutsch durchwegs richtig zusammengezählt. Hier wird angemerkt, dass die Prüfung ausschliesslich aus einem Aufsatz besteht und die zusammenzurechnenden Punkte alle auf derselben Seite aufgeführt werden.

Bei den Prüfungen in Englisch und Mathematik von 2019 ist es nur in Einzelfällen zu nicht korrekt zusammengezählten Punkten gekommen. Bei der Französischprüfung 2019 hingegen hat das inkorrekte Aufsummieren der Punkte meistens auch zu einer inkorrekten Benotung geführt. Hier hat vor allem eine bestimmte Aufgabe zu Problemen geführt. Dabei handelt es sich um eine Verbentabelle mit 24 Lücken, für die je ein halber Punkt vergeben wird. In acht Zeilen müssen jeweils drei

Felder korrigiert werden, wobei sich diese drei Felder zufällig auf vier Spalten verteilen. Es überrascht also nicht, dass es hier während des Zusammenzählens zu Fehlern kommt.

Auch für die BM II Typ «andere» hat sich die falsche Benotung für keinen Kandidaten und keine Kandidatin auf den Prüfungsentscheid ausgewirkt.

Allerdings wurde festgestellt, dass die auf den Papierheften festgehaltene Note nicht immer identisch ist mit der Note in der konsolidierten Notenübersicht. Über alle Fächer und beide BM-Typen hinweg ist das bei 17 Einträgen der Fall. In drei Fällen hat dies dazu geführt, dass der Prüfungsentscheid falsch gefällt wurde. Eine Kandidatin hat einen positiven Bescheid bekommen, obwohl der Gesamtschnitt, berechnet auf Basis der (korrekten) Noten auf den Papierheften, ungenügend ist. Auf der anderen Seite haben zwei Kandidatinnen/Kandidaten einen negativen Bescheid erhalten, obwohl der Gesamtschnitt, berechnet auf Basis der (korrekten) Noten auf den Papierheften, genügend ist und jeweils nur eine Note unter 4.0 liegt.

Reliabilität

In Tabelle 36 sind die Kennzahlen zur Reliabilität enthalten. Das Cronbach Alpha ist ein Mass für die interne Konsistenz der Prüfung (Mittelwert der Interkorrelationen aller Teilaufgaben). Die Split-Half-Reliabilität gibt die Korrelation zwischen der in zwei Teile halbierten Prüfung an (Testhalbierungsreliabilität). Weil die Reliabilität von der Länge der Prüfung abhängt, sind in Klammern die Reliabilitätskennwerte enthalten, die sich bei einer Prüfungslänge von 50 Aufgaben ergeben würden. Mit der Aufnahmeprüfung Deutsch für die BM II werden nur die Schreibkompetenzen geprüft. Die Reliabilität lässt sich deshalb nicht aufgrund richtig oder falsch gelöster Aufgaben, sondern nur aufgrund der Bewertung der Texte berechnen.

Über alle Fächer hinweg und für beide BM-Typen kann die Reliabilität als mindestens ausreichend ($r > .70$) eingestuft werden. Die höchste Reliabilität erreichen die Französischprüfungen für die beiden BM-Typen. Die Reliabilitätskoeffizienten von $r > .90$ können als hervorragend eingestuft werden. Als gut ($r > .80$) ist die Reliabilität beider Englischprüfungen sowie der Mathematikprüfung der BM I und der Deutschprüfung der BM II einzustufen. Ausreichend ($r > .70$) ist die Reliabilität der Deutschprüfung für die BM I und der Mathematikprüfung für die BM II.

Tabelle 36: Reliabilität nach klassischer Testtheorie der Aufnahmeprüfung von 2019 nach BM-Typ und Fach

	Reliabilität			
	BM I		BM II Typ «andere»	
	Cronbach's α	Split-Half	Cronbach's α	Split-Half
Deutsch	.72 (.72)	.74 (.74)	.75 (.97)	.85 (.98)
Englisch	.81 (.81)	.88 (.88)	.81 (.82)	.82 (.82)
Französisch	.93 (.90)	.94 (.92)	.93 (.93)	.94 (.94)
Mathematik	.82 (.93)	.83 (.94)	.76 (.91)	.79 (.92)

Bemerkung: Bei den Koeffizienten in Klammern handelt es sich um die auf eine Testlänge von 50 Aufgaben hochgerechneten Reliabilitätskoeffizienten.

Es gilt nochmals zu betonen, dass die Reliabilität auch von der Prüfungslänge abhängt. Je mehr Aufgaben eine Prüfung enthält, desto höher ist tendenziell ihre Reliabilität. Angesichts dieser Tatsache lassen sich die Unterschiede zwischen den einzelnen Prüfungen teilweise aufgrund der Prü-

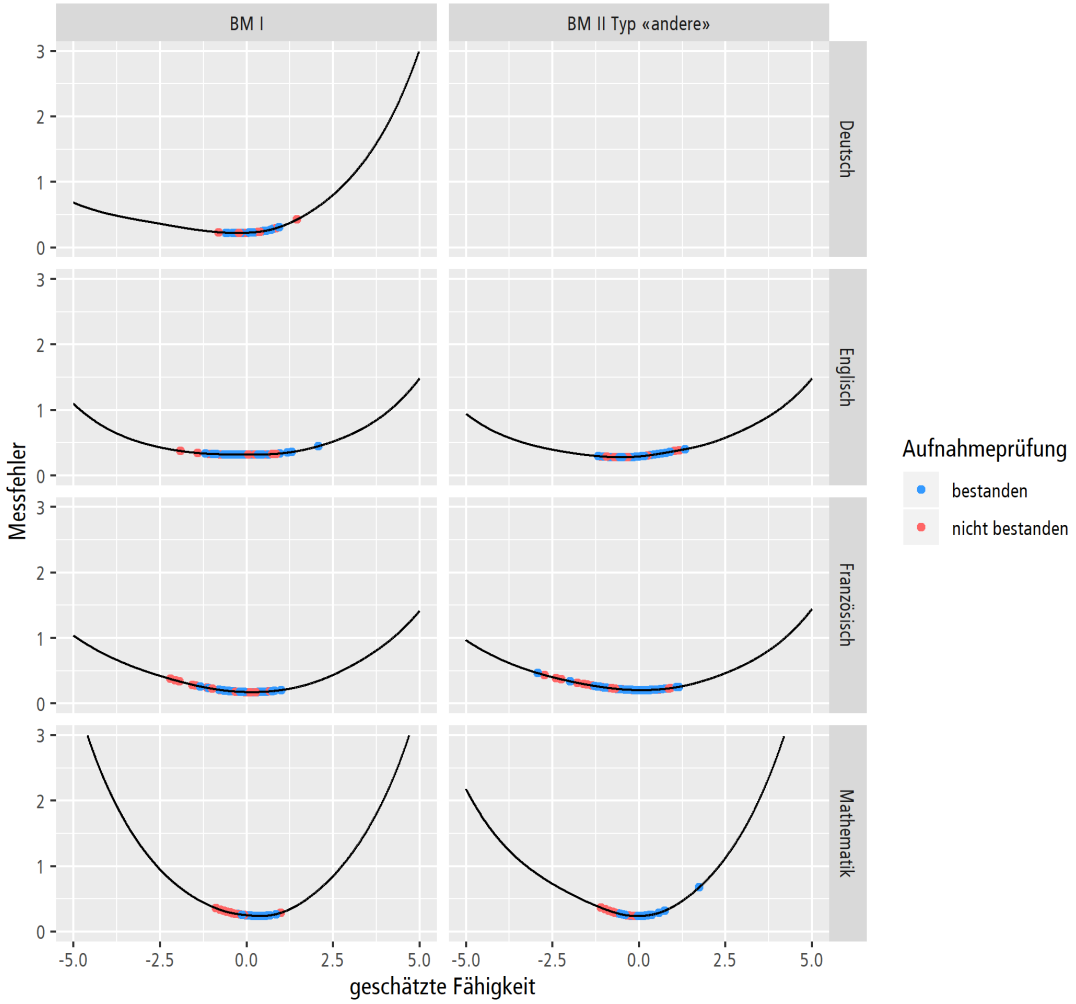
fungslänge erklären. Während die beiden Englischprüfungen, die Deutschprüfung für die BM I und die Französischprüfung für die BM II je rund 50 Aufgaben[9] enthalten, ist die Anzahl Aufgaben der Mathematikprüfung mit 16 Aufgaben und 21 vergebenen Punkten zu gering. Je mehr Chancen Kandidatinnen und Kandidaten haben, ihre Fähigkeiten zu zeigen, desto zuverlässiger können die Fähigkeiten erfasst werden. Dieser Grundsatz spiegelt sich auch in den Reliabilitätskennwerten der Französischprüfung, in der 71 Aufgaben eingesetzt wurden. Die Reliabilität kann entweder mit der Anzahl Aufgaben oder mit der Vergabe von Teilpunkten zu klar definierten Teilergebnissen einer Aufgabe erhöht werden. Wäre die Mathematikprüfung gleich umfassend wie die Sprachprüfungen (bzw. würden 50 Punkte vergeben), dann wären die Reliabilitätskennwerte $r > .90$, also als hervorragend zu bezeichnen (vgl. Tabelle 36, Reliabilitätskennwerte in Klammern).

Abbildung 11 zeigt die Messfehler entlang des Fähigkeitsspektrums der Kandidatinnen und Kandidaten in den einzelnen Fächern. Die geschätzte Personenfähigkeit ist auf der standardisierten Theta-Skala dargestellt, deren Mittelwert bei 0 und deren Standardabweichung bei 1 liegt. Der Theta-Wert ergibt sich durch die Anzahl Standardabweichungen vom Mittelwert der Fähigkeiten ($M = 0$). Dadurch bewegen sich 99 Prozent der geschätzten Fähigkeiten zwischen –3 und 3. Idealerweise sind die Messfehler dort am geringsten, wo die Unterscheidung zwischen «bestanden» und «nicht bestanden» gemacht wird. Der Schwellenwert für das Bestehen der Aufnahmeprüfung liegt bei 4.0. Dargestellt werden die Messfehler der Kandidatinnen und Kandidaten, deren Gesamtnote zwischen 3.5 und 4.5 liegt.

Aus Abbildung 11 geht hervor, dass der Messfehler in der Mitte des Fähigkeitsspektrums erwartungsgemäss am geringsten ist. Die u-förmigen Kurven haben hier ihren tiefsten Punkt. Die blauen und roten Punkte, welche die Kandidatinnen und Kandidaten mit Noten zwischen 3.5 und 4.5 darstellen, sammeln sich in allen Fächern um diese tiefsten Punkte. Das bedeutet, dass die Aufnahmeprüfungen tatsächlich beim Schwellenwert (bestanden/nicht bestanden) am genauesten messen. In Deutsch und Mathematik streuen die Kandidatinnen und Kandidaten im Schwellenbereich nicht so stark wie in den beiden Fremdsprachen. In den beiden Französischprüfungen ist eine leichte Tendenz nach links zu sehen. Die Aufnahmeprüfungen in Französisch messen also bei den Kandidatinnen und Kandidaten, die eine Gesamtnote von 4.0 bis 4.5 haben, etwas genauer als bei denjenigen, die eine Gesamtnote von 3.5 bis 4.0 haben.

[9] In diesem Zusammenhang wird unter «Aufgabe» jede für sich bewertete Antwort gezählt. Ein Lückentext mit zehn Lücken, der im Sinne des Prüfungsdesigns als eine Aufgabe dargestellt wird, bei dem die Punkte aber für jede Lücke einzeln vergeben werden, gilt also als zehn Aufgaben.

Abbildung 11: Verteilung der Messfehler über das gesamte Fähigkeitsspektrum nach BM-Typ und Fachbereich



Validität

Die Validität wird anhand der konvergenten und divergenten Validität beurteilt. Dazu werden die Korrelationen zwischen den Ergebnissen in den Prüfungen und im Check S3 berechnet. Die Korrelationen in gleichen Fächern (bspw. Deutsch) sollten möglichst hoch sein, damit von einer hohen konvergenten Validität gesprochen werden kann. Die Korrelationen zwischen unterschiedlichen Fächern (bspw. Deutsch BM I und Mathematik Check S3) sollten möglichst gering sein, damit von einer hohen divergenten Validität gesprochen werden kann.

Tabelle 37 enthält die Korrelationskoeffizienten der Ergebnisse in den Aufnahmeprüfungen für die BM I und im Check S3 2019 (Multitrait-Multimethod-Matrix). Die Korrelationskoeffizienten zur konvergenten Validität zwischen den Prüfungsteilen und den Check-Ergebnissen liegen zwischen $r = .57$ (Deutsch und Französisch) und $r = .74$ (Englisch) und sind allesamt statistisch signifikant. Die konvergente Validität ist somit gegeben.

Die Korrelationen zur divergenten Validität liegen zwischen $r = .13$ (Mathematik BM I und Englisch Check S3) und $r = .57$ (Englisch BM I und Deutsch Check S3). Insgesamt sind die Korrelationskoeffizienten zur divergenten Validität grösstenteils geringer als die Korrelationen zur konvergenten Validität, weshalb die Validität der Aufnahmeprüfung für die BM I insgesamt als genügend hoch beurteilt werden kann.

Tabelle 37: Konvergente und divergente Validität: Korrelationskoeffizienten (Check S3 und BM I)

		Check S3				BM I			
		D	E	F	M	D	E	F	M
Check S3	Deutsch (D)								
	Englisch (E)	.47*							
	Französisch (F)	.40*	.45*						
	Mathematik (M)	.39*	.23*	.22*					
BM I	Deutsch (D)	.57*	.20*	.32*	.29*	(.72)			
	Englisch (E)	.57*	.74*	.41*	.20*	.45*	(.81)		
	Französisch (F)	.44*	.21*	.57*	.34*	.36*	.42*	(.93)	
	Mathematik (M)	.30*	.13	.18*	.61*	.35*	.29*	.45*	(.82)

Bemerkung: Die mit einem Stern markierten Korrelationskoeffizienten sind statistisch signifikant ($p < .05$). In Klammern ist die Reliabilität (Cronbach's α) der Aufnahmeprüfungen angegeben. Für den Check S3 2019 wird die Reliabilität ausschliesslich auf Ebene der Kompetenzbereiche und nicht auf Ebene der Fachbereiche ausgegeben. Deswegen ist die Reliabilitätsdiagonale im linken oberen Quadrat leer.

Tabelle 38 enthält die Korrelationskoeffizienten der Ergebnisse in den Aufnahmeprüfungen für die BM II und im Check S3 2019 (Multitrait-Multimethod-Matrix). Im Vergleich zu den Ergebnissen der Prüfung für die BM I ergibt sich hier ein anderes Bild. Die Korrelationskoeffizienten zur konvergenten Validität zwischen den Prüfungsteilen und den Check-Ergebnissen liegen zwischen $r = .23$ (Deutsch) und $r = .61$ (Englisch). Nicht alle Korrelationskoeffizienten unterscheiden sich statistisch signifikant von null. Die konvergente Validität ist für die Deutsch- und die Mathematikprüfung nicht gegeben.

Tabelle 38: Konvergente und divergente Validität: Korrelationskoeffizienten (Check S3 und BM II)

		Check S3				BM II			
		D	E	F	M	D	E	F	M
Check S3	Deutsch (D)								
	Englisch (E)	.47*							
	Französisch (F)	.40*	.45*						
	Mathematik (M)	.39*	.23*	.22*					
BM II	Deutsch (D)	.23	.05	.20	– .14	(.75)			
	Englisch (E)	.05	.61*	.27	.21	.31*	(.81)		
	Französisch (F)	– .06	.15*	.59*	.19	.13*	.24*	(.93)	
	Mathematik (M)	– .01	.10	.42	.33	.24*	.31*	.26*	(.76)

Bemerkung: Die mit einem Stern markierten Korrelationskoeffizienten sind statistisch signifikant ($p < .05$). In Klammern ist die Reliabilität (Cronbach's α) der Aufnahmeprüfungen angegeben. Für den Check S3 2019 wird die Reliabilität ausschliesslich auf Ebene der Kompetenzbereiche und nicht auf Ebene der Fachbereiche ausgegeben. Deswegen ist die Reliabilitätsdiagonale im linken oberen Quadrat leer.

Die Korrelationskoeffizienten zur divergenten Validität liegen zwischen $r = -.06$ (Französisch BM II und Deutsch Check S3) und $r = .47$ (Englisch Check S3 und Deutsch Check S3). Da sich diese Korrelationen mit den Korrelationen zur konvergenten Validität überschneiden und da die konvergente Validität verletzt ist, wird die Validität der Aufnahmeprüfung in die BM II – beurteilt anhand des Check S3 – insgesamt als nicht gegeben beurteilt.

Fazit

In Bezug auf die Auswertungsobjektivität lässt sich zunächst feststellen, dass die Musterlösungen teilweise keine oder nur vage formulierte Korrekturanweisungen beinhalten. Das öffnet Spielraum, Aufgaben unterschiedlich zu bewerten, und ist angesichts der Tatsache, dass die Prüfungen dezentral korrigiert werden, problematisch. Zudem wird bei Schreibaufträgen, die ohne Kriterienraster bewertet werden, der Ermessensspielraum für die Korrektorinnen und Korrektoren als zu gross eingeschätzt. Kriterienraster, Korrektur- und Bewertungsanweisungen sowie Musterlösungen sind für eine zuverlässige Korrektur und Bewertung unerlässlich und somit die Grundlage für das Einhalten der Objektivität.

Das Beispiel mit der falschen Musterlösung in der Deutschprüfung für die BM I deutet darauf hin, dass sich die Korrektur-Teams der unterschiedlichen Berufsmaturitätsschulen untereinander nicht ausreichend austauschen bzw. dass allenfalls nicht alle Berufsmaturitätsschulen in den Austausch involviert sind. Um ein hohes Mass an Auswertungsobjektivität zu gewährleisten, sollten Fragen zur Bewertung einzelner Aufgaben besprochen und daraus resultierende Entscheide und ergänzende Korrekturanweisungen dem gesamten Korrekturteam mitgeteilt werden. Die Objektivität kann auch erhöht werden, wenn anstelle der Prüfungen einzelne Prüfungsaufgaben zur Korrektur und Beurteilung auf einzelne Personen aufgeteilt werden. Dass es trotz des Vier-Augen-Prinzips zu falschen Aufsummierungen der Punkte kommt, die sich darüber hinaus in manchen Fällen auf die Notengebung auswirken, weckt zudem den Verdacht, dass das Vier-Augen-Prinzip nicht durchgängig zum Tragen kommt.

Über die Reliabilität der Aufnahmeprüfungen lässt sich zusammenfassend festhalten, dass die Messgenauigkeit in allen Fachbereichen und über beide BM-Typen hinweg ausreichend bis hervorragend ist. Hervorragend ist die Reliabilität der Französischprüfungen, ausreichend ist die Reliabilität der Deutschprüfungen. Die Mathematikprüfung für die BM II schneidet bezüglich Reliabilität – vielleicht etwas überraschend – am zweitschlechtesten ab. Dies lässt sich allerdings aufgrund der geringen Anzahl Aufgaben bzw. aufgrund der geringen Anzahl Punkte, die vergeben werden, erklären. Reliabilitätskennzahlen hängen von der Testlänge ab. Je mehr Aufgaben eine Prüfung enthält – vorausgesetzt, dass diese Aufgaben gut zwischen stärkeren und schwächeren Kandidatinnen und Kandidaten unterscheiden bzw. trennscharf sind (vgl. Absatz 6.6) –, desto mehr Möglichkeiten haben die Kandidatinnen und Kandidaten, ihre Kompetenzen unter Beweis zu stellen, und desto geringer wird der Messfehler bzw. desto zuverlässiger fällt das Prüfungsergebnis aus.

Positiv zu werten ist darüber hinaus, dass die Messgenauigkeit in allen Fachbereichen und über beide BM-Typen hinweg dort am höchsten ist, wo es genau zu messen gilt: bei Kandidatinnen und Kandidaten, deren Gesamtnote sich um die Note 4.0 bewegt.

Die Beurteilung der Validität erfolgte anhand der unabhängigen Leistungstests des Check S3. Die Aufgaben dieser Leistungstests werden von Fachdidaktikerinnen und Fachdidaktikern der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz auf der Grundlage des Lehrplan 21 entwickelt und vom Institut für Bildungsevaluation auf psychometrische Gütekriterien geprüft. Aufgrund der Berechnung der konvergenten Validität (Korrelation zwischen gleichen Fächern) und der divergenten Validität (Korrelation zwischen unterschiedlichen Fächern) kann die Validität der Aufnahmeprüfung für die BM I grösstenteils als ausreichend bis gut beurteilt werden. Eine ab-

schliessende Beurteilung der Validität der Aufnahmeprüfung in die BM II ist aufgrund der Datenlage nicht möglich. Die Anzahl Kandidatinnen und Kandidaten, für die Prüfungsergebnisse und Ergebnisse im Check S3 vorliegen, ist mit 42 relativ klein, zudem können Verzerrungen durch Selektionseffekte nicht ausgeschlossen werden. Die vorliegenden Daten deuten allerdings darauf hin, dass die Validität der Aufnahmeprüfung in die BM II eher nicht gegeben ist. Dies könnte auch damit zusammenhängen, dass die Messzeitpunkte zwischen dem Check S3 und der Aufnahmeprüfung mindestens drei Jahre auseinanderliegen.

Handlungsbedarf wird bezüglich der Gütekriterien von Aufnahmeprüfungen vor allem in der Auswertungsobjektivität gesehen. Diese könnte erstens durch ausführlichere Lösungsschlüssel erhöht werden. Insbesondere die Prüfungen des Typs «Wirtschaft» weisen bezüglich der angewendeten Kriterien einige Lücken auf. Zumindest die Aufschlüsselung der Punktzahlen auf die einzelnen Teilaufgaben sollte gegeben sein. Zudem könnte für die Beurteilung der Schreibkompetenzen die Verwendung eines Kriterienrasters die Auswertungsobjektivität bei allen BM-Typen erhöhen. Zweitens sollte sichergestellt werden, dass die in den einzelnen Teilaufgaben erreichten Punkte korrekt aufsummiert werden. Dazu sollte auf Aufgabenstellungen verzichtet werden, die das Aufsummieren erschweren bzw. sollte das Layout so angepasst werden, dass genügend Platz für das (schriftliche) Aufsummieren in Teilschritten vorhanden ist. Drittens könnte die Objektivität erhöht werden, indem die Vergabe der Punkte vereinheitlicht oder standardisiert wird. Einzelne Korrektorinnen und Korrektoren arbeiten bspw. mit Minuspunkten, was von der zweiten Korrektorin bzw. dem zweiten Korrektor ein Umdenken erfordert. Viertens sollte sichergestellt werden, dass das Übertragen der Prüfungsnoten auf die Notenübersichtsliste korrekt erfolgt. Dabei könnte neben der Note auch die erreichte Gesamtpunktzahl in einem Fach übertragen und die Note im Excel basierend auf der Punktzahl und dem Notenschlüssel direkt berechnet werden. Durch den Abgleich der berechneten Note mit der eingetragenen Note könnte überprüft werden, ob die richtige Note vergeben und ob diese Note auch richtig übertragen wurde. Um Fehlentscheide zu vermeiden, wird fünftens vorgeschlagen, die Prüfungen von Kandidatinnen und Kandidaten, die knapp bestehen oder knapp nicht bestehen, noch einmal zu überprüfen.

6.6 Gütekriterien der Prüfungsaufgaben

Die Qualität der einzelnen Aufgaben wird anhand der Schwierigkeit, Diskrimination und der Fairness der Prüfungsaufgaben beurteilt. Prüfungsaufgaben sollten möglichst gut zwischen geeigneten und weniger geeigneten Kandidatinnen und Kandidaten differenzieren, sie sollten das Fähigkeitsspektrum der Kandidatinnen und Kandidaten abdecken und sie sollten für alle Kandidatinnen und Kandidaten fair sein, also das Gleiche messen. Die Fairness der Aufgaben wird im Hinblick auf das Geschlecht untersucht, also daraufhin, ob die Aufgaben für weibliche und männliche Kandidatinnen und Kandidaten gleich funktionieren. Ebenso wichtig wäre es grundsätzlich, die Fairness in Bezug auf die soziale Herkunft oder den Migrationshintergrund zu prüfen. Diese Angaben liegen allerdings nicht vor.

Die Qualitätsprüfung wurde anhand der Aufgaben der Prüfungen 2019 für die BM I und für die BM II Typ «andere» durchgeführt. Dazu wurden die gelösten und bewerteten Aufgaben von 157 Kandidatinnen und Kandidaten für die BM I und von 205 Kandidatinnen und Kandidaten für die BM II elektronisch erfasst und für die Berechnung der Kennwerte genutzt. Schreibkriterien werden in dieser Analyse nicht berücksichtigt. Da die Deutschprüfung für die BM II Typ «andere» lediglich aus einem Schreibauftrag besteht und keine Aufgaben enthält, die als richtig, teilweise richtig oder falsch beurteilt werden können, ist sie von dieser Betrachtung ausgenommen.

Aufgabenschwierigkeit

Die Aufgabenschwierigkeit entspricht dem Anteil Kandidatinnen und Kandidaten, die die jeweilige Aufgabe vollständig korrekt gelöst haben. Tabelle 39 enthält eine Zusammenfassung zu den Aufgabenschwierigkeiten der Prüfungsaufgaben für die BM I 2019. Aufgaben mit einer Schwierigkeit von 0 bis 25 Prozent gelten als schwierig, Aufgaben mit einer Schwierigkeit von 25 bis 50 Prozent als eher schwierig, Aufgaben mit einer Schwierigkeit von 50 bis 75 Prozent als eher einfach und Aufgaben mit einer Schwierigkeit von über 75 Prozent als einfach. Die Aufgabenschwierigkeit wird für jedes Fach und jeden Kompetenzbereich angegeben. Die Aufgabenschwierigkeiten der einzelnen Prüfungsaufgaben sind dem Anhang B zu entnehmen.

Tabelle 39: Anzahl Aufgaben nach Schwierigkeit und mittlere Schwierigkeit nach Fach und Kompetenzbereich (Aufnahmeprüfung für die BM I von 2019)

	Lösungswahrscheinlichkeit				
	Anzahl Aufgaben pro Kategorie				
	0–25%	25–50%	50–75%	75–100%	Mittelwert
Deutsch	6	5	14	24	67%
Lesen	5	3	6	10	60%
Sprache im Fokus	1	2	8	14	72%
Englisch	17	13	13	5	39%
Lesen	3	5	7	1	48%
Sprache im Fokus	14	8	6	4	35%
Französisch	13	24	23	9	47%
Lesen	1	3	13	2	23%
Sprache im Fokus	12	21	10	7	44%
Mathematik	8	6	2	0	33%
Zahl und Variable	3	3	0	0	29%
Form und Raum	3	0	2	0	36%
Grössen, Funktionen, Daten und Zufall	2	3	0	0	34%

Knapp die Hälfte aller Aufgaben der Deutschprüfung ist einfach (24 von 49 Aufgaben). Diese Aufgaben werden von über 75 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst. 14 von 49 Aufgaben wurden von 50 bis 75 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst. Nur gerade ein Achtel der Aufgaben wird von weniger als der Hälfte der Kandidatinnen und Kandidaten richtig gelöst. Die Aufgaben in *Deutsch Lesen* sind tendenziell schwieriger als die Aufgaben in *Deutsch Sprache im Fokus*. Die mittlere Aufgabenschwierigkeit liegt bei $p = 67$ Prozent. Die Deutschprüfung ist dementsprechend zu einfach.

Rund ein Drittel aller Aufgaben der Englischprüfung wird von weniger als 25 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst. Gut die Hälfte der Aufgaben befindet sich im mittleren Schwierigkeitsbereich (25–75 Prozent) und 5 von 48 Aufgaben sind sehr einfach. Die mittlere Aufgabenschwierigkeit liegt bei $p = 39$ Prozent, womit die Englischprüfung als eher etwas zu schwierig einzustufen ist.

50 von 69 Aufgaben der Französischprüfung befinden sich im mittleren Schwierigkeitsbereich. 9 Aufgaben werden von mehr als 75 Prozent und 13 Aufgaben von weniger als 25 Prozent der Kan-

didatinnen und Kandidaten vollständig richtig gelöst. Die Aufgaben in *Französisch Lesen* sind tendenziell etwas einfacher als die Aufgaben in *Französisch Sprache im Fokus*. Die mittlere Aufgabenschwierigkeit liegt bei $p = 47$ Prozent. Mit dieser mittleren Schwierigkeit liegt die Französischprüfung nahe bei der idealen mittleren Aufgabenschwierigkeit von $p = 50$ Prozent.

Die Mathematikprüfung enthält keine einfachen Aufgaben. Die Hälfte der Aufgaben wird von weniger als 25 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst und ist somit schwierig. 6 von 16 Aufgaben sind eher schwierig. Zudem werden lediglich zwei Aufgaben im Kompetenzbereich *Mathematik Form und Raum* von mehr als 50 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst. Die mittlere Aufgabenschwierigkeit liegt bei $p = 33$ Prozent, womit die Mathematikprüfung als zu schwierig beurteilt wird.

Tabelle 40 enthält die Zusammenfassung zu den Aufgabenschwierigkeiten der Prüfungsaufgaben für die BM II Typ «andere» aus dem Jahr 2019. Aufgaben mit einer Schwierigkeit von 0 bis 25 Prozent gelten als schwierig, Aufgaben mit einer Schwierigkeit von 25 bis 50 Prozent als eher schwierig, Aufgaben mit einer Schwierigkeit von 50 bis 75 Prozent als eher einfach und Aufgaben mit einer Schwierigkeit von über 75 Prozent als einfach. Die Aufgabenschwierigkeit wird für jedes Fach und jeden Kompetenzbereich angegeben. Die Aufgabenschwierigkeiten der einzelnen Prüfungsaufgaben sind dem Anhang B zu entnehmen.

Tabelle 40: Anzahl Aufgaben nach Schwierigkeit und mittlere Schwierigkeit nach Fach und Kompetenzbereich (Aufnahmeprüfung für die BM II von 2019)

	Lösungswahrscheinlichkeit				
	Anzahl Aufgaben pro Kategorie				
	0–25%	25–50%	50–75%	75–100%	Mittelwert
Englisch	14	8	14	13	48%
Lesen	4	3	5	3	49%
Sprache im Fokus	10	5	9	10	45%
Französisch	8	24	21	12	51%
Lesen	3	1	3	5	56%
Sprache im Fokus	5	23	18	7	50%
Mathematik	3	7	4	2	47%
Zahl und Variable	1	4	1	0	39%
Form und Raum	1	3	1	0	41%
Grössen, Funktionen, Daten und Zufall	1	0	2	2	61%

Von 49 Aufgaben der Englischprüfung sind 14 schwierig und 13 einfach. 45 Prozent der Aufgaben bewegen sich im mittleren Schwierigkeitsbereich. Die Aufgaben des Kompetenzbereichs *Englisch Lesen* verteilen sich relativ gleichmässig auf die vier Kategorien (schwierig, eher schwierig, eher einfach, einfach). Auch die Aufgaben in *Englisch Sprache im Fokus* verteilen sich in ähnlichem Ausmass auf die vier Kategorien, wobei die Kategorie «eher einfach» mit 5 Aufgaben weniger Aufgaben enthält als die anderen drei Kategorien (9–10 Aufgaben). Die mittlere Aufgabenschwierigkeit liegt bei $p = 48$ Prozent.

Die Mehrheit der Aufgaben der Französischprüfung bewegt sich im mittleren Schwierigkeitsbereich (45 von 65 Aufgaben). 12 Aufgaben werden von mehr als 75 Prozent und 8 Aufgaben von

weniger als 25 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst. Die Aufgaben in *Französisch Lesen* sind tendenziell etwas einfacher als die Aufgaben in *Französisch Sprache im Fokus*. Die mittlere Aufgabenschwierigkeit liegt bei $p = 51$ Prozent.

Die Mehrheit der Aufgaben der Mathematikprüfung bewegt sich im mittleren Schwierigkeitsbereich (11 von 16 Aufgaben). 2 von 16 Aufgaben werden von mehr als 75 Prozent der Kandidatinnen und Kandidaten vollständig richtig gelöst. 3 von 16 Aufgaben werden von weniger als 25 Prozent der Kandidatinnen und Kandidaten gelöst. Die Aufgaben des Kompetenzbereichs *Mathematik Grössen, Funktionen, Daten und Zufall* sind am einfachsten. Die mittlere Aufgabenschwierigkeit liegt bei $p = 47$ Prozent.

Die mittleren Aufgabenschwierigkeiten bewegen sich somit für die Englisch-, Französisch- und Mathematikprüfung der BM II im idealen Bereich.

Diskrimination

Für die Beurteilung der Diskrimination wurde der Trennschärfekoeffizient berechnet. Die Trennschärfe gibt an, wie gut das Gesamtergebnis mit dem Lösen der Aufgabe korreliert bzw. wie hoch die Korrelation der Aufgabe mit dem Gesamtergebnis ist. Es wird erwartet, dass Personen, die eine Aufgabe falsch lösen, insgesamt eine tiefere Gesamtpunktzahl erreichen als Personen, die die Aufgabe richtig lösen. Tabelle 41 enthält eine Zusammenfassung zu den Trennschärfekoeffizienten der Prüfungsaufgaben für die BM I. Dazu wurden die Trennschärfekoeffizienten der Aufgaben den drei Kategorien «negative Trennschärfe», «niedrige Trennschärfe» und «ausreichende bis hohe Trennschärfe» zugeordnet. Zudem wird für jeden Kompetenzbereich die durchschnittliche Trennschärfe (Mittelwert der jeweiligen Aufgaben) ausgewiesen. Die Gütekriterien zu den einzelnen Prüfungsaufgaben sind dem Anhang B zu entnehmen.

Tabelle 41: Aufnahmeprüfung 2019 für die BM I: Anzahl Aufgaben nach Kategorien der Trennschärfe sowie Mittelwert der Trennschärfekoeffizienten

	Kategorien			
	negativ	niedrig	ausreichend bis hoch	Trennschärfekoeffizient: Mittelwert
Deutsch				
Lesen	0	18	6	.25
Sprache im Fokus	0	16	9	.27
Englisch				
Lesen	3	6	7	.27
Sprache im Fokus	0	13	19	.36
Französisch				
Lesen	0	5	14	.34
Sprache im Fokus	1	13	36	.36
Mathematik				
Zahl und Variable	0	0	5	.61
Form und Raum	0	0	4	.48
Grössen, Funktionen, Daten und Zufall	0	0	6	.47

Die Aufnahmeprüfung Deutsch enthält keine Prüfungsaufgaben mit negativer Trennschärfe. Allerdings ist von lediglich 13 der 49 Aufgaben die Trennschärfe als ausreichend bis hoch einzustufen. Die Mehrheit der Aufgaben (36 von 49) ist nicht geeignet, um zwischen Kandidatinnen und Kandidaten mit hohen und tiefen Punktzahlen zu unterscheiden. Der Anteil an trennscharfen Aufgaben ist im Bereich *Sprache im Fokus* höher als im Bereich *Lesen*. Schaut man sich die einzelnen Aufgaben genauer an, fallen einzelne Aufgabentypen durch tiefere Trennschärfen auf. Bei 4 der 24 Aufgaben in *Lesen* handelt es sich beispielsweise um Binary-Choice-Aufgaben. Die Kandidatinnen und Kandidaten müssen beurteilen, ob eine Aussage zum Text richtig oder falsch ist, ohne ihre Antwort zu begründen. Diese Aufgaben erweisen sich insgesamt als wenig trennscharf, weil die Ratewahrscheinlichkeit mit 50 Prozent sehr hoch ist. In einer weiteren Aufgabe werden die Kandidatinnen und Kandidaten aufgefordert, insgesamt acht Wörtern je ein Synonym zuzuordnen, wobei 20 Begriffe zur Auswahl gegeben werden. Auch diese acht (Teil-)Aufgaben weisen allesamt eine geringe Trennschärfe auf. Bei den Binary-Choice-Aufgaben können statistisch gesehen die Hälfte der Punkte erreicht werden, ohne dass der Text dazu gelesen wird. Bei der Zuordnungsaufgabe werden Verben und Nomen gemischt. Durch Ausschliessen der jeweiligen Wortart können auch hier vergleichsweise einfach Punkte durch Raten erzielt werden. Als besonders trennscharf haben sich in Deutsch offene Fragen erwiesen, die in ganzen Sätzen beantwortet werden müssen und bei denen die Punkte nicht nur für den Inhalt, sondern auch für die sprachliche Richtigkeit vergeben werden.

Die Aufnahmeprüfung Englisch enthält drei Aufgaben mit negativer Trennschärfe. 19 von 48 Aufgaben verfügen über eine geringe Trennschärfe. Gut die Hälfte der Aufgaben verzeichnet eine ausreichende bis hohe Trennschärfe. Der Anteil an trennscharfen Aufgaben ist in *Sprache im Fokus* höher als in *Lesen*. Bei der Betrachtung der einzelnen Aufgaben der Englischprüfung fällt ebenfalls auf, dass Binary-Choice-Aufgaben zum Lesetext eine niedrige oder gar negative Trennschärfe aufweisen. Dieses Aufgabenformat ist nicht geeignet, um zwischen starken und schwachen Kandidatinnen und Kandidaten zu unterscheiden. Bei den übrigen Aufgabenformaten finden sich sowohl trennscharfe als auch wenig trennscharfe Aufgaben.

Die Mehrheit der Aufgaben der Aufnahmeprüfung Französisch weist eine hohe Trennschärfe auf (50 von 69 Aufgaben). Dies trifft für die Aufgaben beider Kompetenzbereiche *Lesen* und *Sprache im Fokus* zu. Es ist kein Aufgabentyp auszumachen, der sich als insgesamt wenig trennscharf herausstellt. Die Binary-Choice-Aufgaben zum Lesetext weisen grösstenteils eine hohe Trennschärfe auf. Anders als in Deutsch und Englisch muss die Entscheidung hier mit einer Textstelle begründet werden, was die Ratewahrscheinlichkeit deutlich verringert. Als besonders trennscharf haben sich Fragen zur Satzstruktur herausgestellt. Hierbei müssen Sätze zu einer Frage umgestellt und ein bestimmtes Satzglied durch ein Interrogativpronomen ersetzt werden.

Die Mathematikaufgaben in allen drei Kompetenzbereichen verfügen über eine hohe Trennschärfe. Alle Aufgaben in der Mathematikprüfung sind dementsprechend gut geeignet, um zwischen schwächeren und stärkeren Kandidatinnen und Kandidaten zu differenzieren.

Der Vergleich der Trennschärfen zwischen den einzelnen Fächern zeigt, dass die Trennschärfekoeffizienten der Mathematikaufgaben am höchsten sind. Die Trennschärfekoeffizienten der Französischaufgaben sind tendenziell etwas höher als jene der Englischaufgaben und jene der Englischaufgaben sind tendenziell etwas höher als die der Deutschaufgaben.

Tabelle 42 enthält eine Zusammenfassung zu den Trennschärfekoeffizienten der Prüfungsaufgaben für die BM II Typ «andere». Dazu wurden wiederum die Trennschärfekoeffizienten der Aufgaben den drei Kategorien «negative Trennschärfe», «niedrige Trennschärfe» und «ausreichend bis hohe Trennschärfe» zugeordnet. Zudem wird für jeden Kompetenzbereich die durchschnittliche

Trennschärfe (Mittelwert der jeweiligen Aufgaben) ausgewiesen. Die Gütekriterien zu den einzelnen Prüfungsaufgaben sind dem Anhang B zu entnehmen.

Tabelle 42: Aufnahmeprüfung 2019 für die BM II: Anzahl Aufgaben nach Kategorien der Trennschärfe sowie Mittelwert der Trennschärfekoeffizienten

	Kategorien			
	negativ	niedrig	ausreichend bis hoch	Trennschärfekoeffizient: Mittelwert
Englisch				
Lesen	0	8	7	.28
Sprache im Fokus	0	9	25	.36
Französisch				
Lesen	0	6	6	.28
Sprache im Fokus	0	8	45	.46
Mathematik				
Zahl und Variable	0	0	6	.48
Form und Raum	0	0	5	.49
Grössen, Funktionen, Daten und Zufall	0	1	4	.37

32 Aufgaben der Englischprüfung weisen eine ausreichende bis hohe und 17 Aufgaben eine niedrige Trennschärfe auf. Aufgaben mit negativer Trennschärfe kommen nicht vor. Es fällt wiederum auf, dass der Anteil an trennscharfen Aufgaben in *Sprache im Fokus* höher ist als in *Lesen*. Am wenigsten trennscharf sind die Binary-Choice-Aufgaben zum Lesetext. Die Trennschärfekoeffizienten dieser Aufgaben liegen zwischen r^{pt}= .00 und r^{pt}= .21. Am höchsten ist die Trennschärfe von Leseaufgaben, bei denen Sätze mit einem Wort aus dem Text vervollständigt werden müssen.

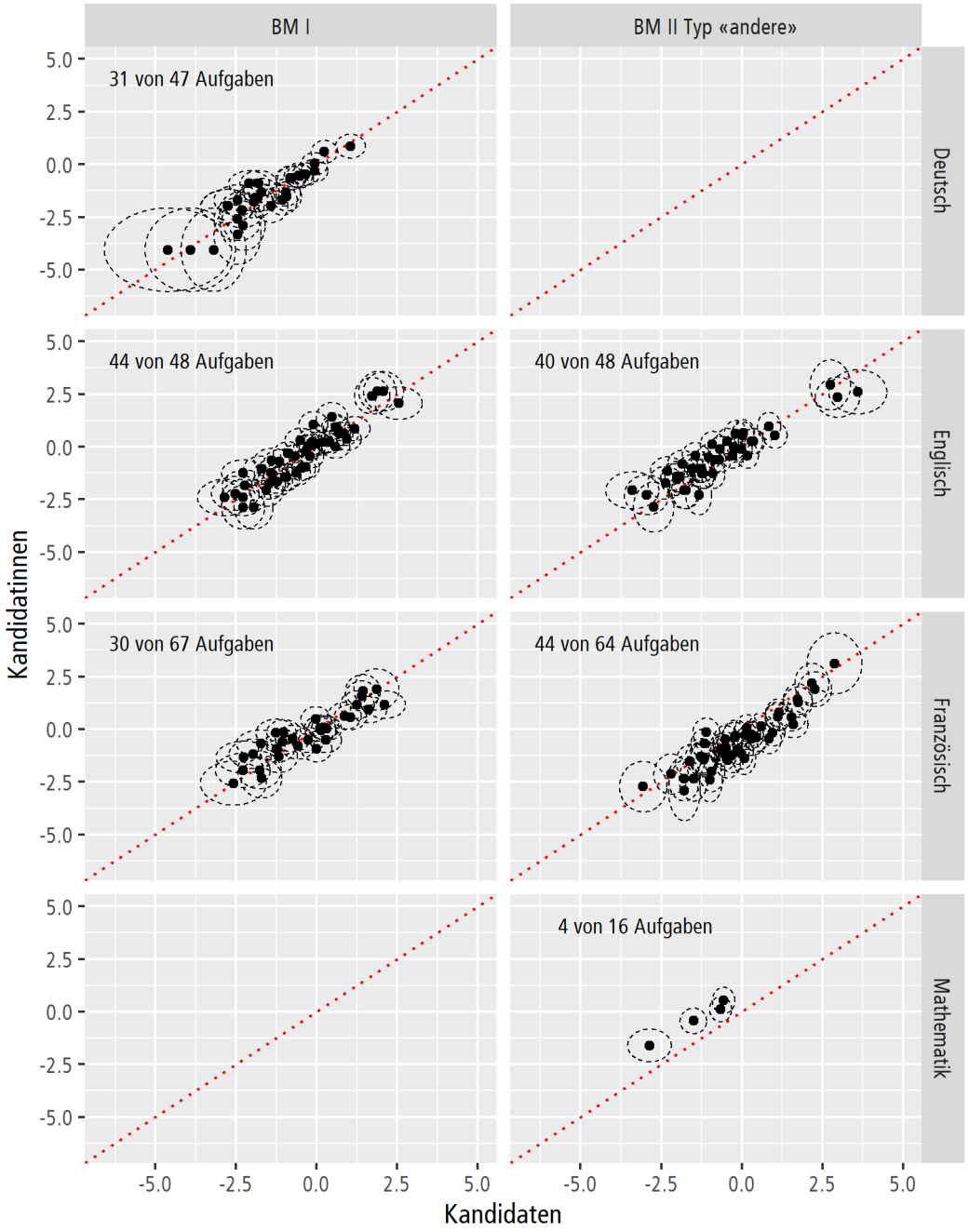
51 der 65 Aufgaben der Aufnahmeprüfung Französisch verfügen über eine hohe Trennschärfe. Die übrigen 14 Aufgaben haben niedrige Trennschärfekoeffizienten. Auch hier zeigt sich, dass Aufgaben im Bereich *Sprache im Fokus* häufiger eine hohe Trennschärfe aufweisen als Aufgaben im Bereich *Lesen*. In der Französischprüfung wird kein Aufgabenformat eingesetzt, das systematisch niedrige Trennschärfekoeffizienten aufweist. Im Vergleich zu den Englischaufgaben ist die Trennschärfe der Französischaufgaben tendenziell etwas höher.

Von den 16 Mathematikaufgaben zeichnen sich 15 durch eine ausreichende bis hohe Trennschärfe aus. Eine Aufgabe in *Grössen, Funktionen, Daten und Zufall* weist eine niedrige Trennschärfe auf.

Fairness

Um zu überprüfen, ob die Aufgaben in Bezug auf das Geschlecht als fair bezeichnet werden können, wurde die relative Schwierigkeit der Aufgaben für weibliche und männliche Kandidatinnen und Kandidaten berechnet. Dazu wird die Schwierigkeit der Aufgaben für beide Gruppen separat berechnet und überprüft, ob die Reihenfolge der Schwierigkeitsparameter jeweils gleich ist und die Aufgaben somit für beide Gruppen gleich funktionieren (bspw. gleich verstanden werden). Falls dies nicht der Fall ist, wird von *Differential Item Functioning (DIF)* gesprochen. Zur Bestimmung des DIF gibt es verschiedene Methoden, von denen hier der grafische Modelltest angewendet wird.

Abbildung 12: Grafischer Modelltest nach BM-Typ und Fachbereich



Der grafische Modelltest ist in Abbildung 12 abgebildet. Dargestellt ist die auf Basis des Rasch-Modells berechnete relative Aufgabenschwierigkeit für die männlichen Kandidierenden auf der X-Achse und für die weiblichen auf der Y-Achse. Die Aufgabenschwierigkeit wird – wie die geschätzte Personenfähigkeit (vgl. Abbildung 11 in Absatz 6.5) – auf der Theta-Skala abgetragen. Die Theta-Skala wird so gebildet, dass die Mittelwerte für weibliche und männliche Kandidierende bei *M* = 0 liegen. Aus diesem Grund kann auch von relativen Schwierigkeiten gesprochen werden. Die Schwierigkeiten werden nur für dichotome Aufgaben angegeben. Für jede Aufnahmeprüfung ist deshalb vermerkt, wie viele Aufgaben in der Grafik zu sehen sind. Aufgaben, die für beide Geschlechter gleich schwierig sind, befinden sich auf der rot gepunkteten Winkelhalbierenden. Geringe Abweichungen von der Winkelhalbierenden sind aufgrund der zufälligen Schwankungen innerhalb einer Stichprobe zulässig. Starke Abweichungen weisen jedoch auf einen systematischen

Unterschied der relativen Aufgabenschwierigkeit zwischen den beiden Gruppen hin. Um zu überprüfen, ob die Abweichungen von Bedeutung sind, sind in Abbildung 12 zusätzlich zu den Punkten, welche die Aufgabenschwierigkeit darstellen, die Konfidenzregionen eingezeichnet, innerhalb welcher die Abweichungen als nicht bedeutsam bezeichnet werden können. Solange die Konfidenzregionen die Winkelhalbierende schneiden, gilt die Abweichung als gering und ein Differential Item Functioning ist nicht angezeigt (vgl. Strobl, 2015, S. 40–41).

Anhand des grafischen Modelltests lässt sich festhalten, dass es vereinzelt Ellipsen gibt, welche die Winkelhalbierende nicht schneiden. Allerdings wird die Winkelhalbierende von diesen Konfidenzbereichen nahezu tangiert. Die Fairness der (dichotomen) Aufgabenstellungen hinsichtlich der Geschlechter wird deshalb als im Grossen und Ganzen gegeben angenommen.

Fazit

Werden die Aufgabentrennschärfen mit der in Absatz 6.3 besprochenen Bedeutung der geprüften Kompetenzbereiche für den Prüfungsentscheid verglichen, zeigt sich besonders für die BM II deutlich, dass eine geringe Bedeutung eines Kompetenzbereichs mit tendenziell tieferen Trennschärfekoeffizienten der entsprechenden Aufgaben einhergeht. Für den Prüfungsentscheid in die BM II Typ «andere» sind die Kompetenzbereiche *Deutsch Schreiben*, *Englisch Sprache im Fokus*, *Französisch Sprache im Fokus* und *Mathematik* von Bedeutung. In diesen Kompetenzbereichen – mit Ausnahme von *Deutsch Schreiben*, das in dieser Analyse nicht berücksichtigt wird – sind die Aufgaben grösstenteils gut geeignet, um zwischen stärkeren und schwächeren Kandidatinnen und Kandidaten zu differenzieren. In *Englisch Lesen* und *Französisch Lesen* verfügt nur rund die Hälfte der Aufgaben über eine ausreichende Trennschärfe. Das spiegelt sich in deren geringen Bedeutung für den Prüfungserfolg wider.

Für die Aufnahmeprüfung in die BM I lässt sich die hohe Bedeutung von *Mathematik* und *Französisch Sprache im Fokus* ebenfalls mit den grösstenteils hohen Trennschärfen der Aufgaben erklären. In Deutsch sind es andererseits die Kompetenzbereiche *Lesen* und *Schreiben*, die eine hohe Bedeutung für den Prüfungsentscheid haben, obwohl die Trennschärfen in *Sprache im Fokus* leicht höher sind als in *Lesen*. Allerdings bewegen sich die Trennschärfen in beiden Deutschbereichen auf eher tiefem Niveau. Die geringe Bedeutung der Englischprüfung lässt sich nicht mit der Diskriminationsfähigkeit der Aufgaben erklären.

Aufgrund der Diskrimination der Prüfungsaufgaben lässt sich zusammenfassend sagen, dass es in den sprachlichen Fächern einen nicht vernachlässigbaren Anteil an nicht trennscharfen Aufgaben gibt. Dies trifft vor allem für die Binary-Choice-Aufgaben zu, die zur Erfassung des Textverständnisses eingesetzt werden.

Die Fairness der (dichotomen) Aufgabenstellungen kann in Bezug auf das Geschlecht als im Grossen und Ganzen gegeben beurteilt werden. Es gibt nur wenige Aufgaben, die bei Kandidatinnen leicht unterschiedlich funktionieren als bei Kandidaten (nur wenige Aufgaben werden von den beiden Gruppen also bspw. unterschiedlich verstanden). Ebenso wichtig wäre es grundsätzlich, die Fairness in Bezug auf die soziale Herkunft oder den Migrationshintergrund zu prüfen. Diese Angaben liegen allerdings nicht vor.

Eine gute Prüfungsaufgabe zu entwickeln ist anspruchsvoll. Die Entwicklung von zuverlässigen (reliablen) und gültigen (validen) Aufgaben setzt theoretisches Wissen zur Aufgabenentwicklung sowie genügend Zeit voraus. Die Reliabilität bzw. die Trennschärfe lässt sich erhöhen, wenn die Aufgaben von der Schwierigkeit her für die Zielgruppe angemessen sind, sprachlich einfach und klar formuliert sind und wenn die korrekte Antwort einer Multiple-Choice-Aufgabe eindeutig bestimmbar ist. Die nicht korrekten Antworten (Distraktoren) müssen nicht allesamt völlig unplausibel sein bzw. können sie Personen, ohne das erforderliche Wissen, durchaus plausibel erscheinen,

die richtige Antwort muss aber eindeutig die beste sein. Die Antwortoptionen sollten alle gleich lang und differenziert sein und nur eine Aussage enthalten. Die Antwortoptionen sind hingegen ungeeignet, wenn sie auch ohne das verlangte Wissen als Distraktoren erkannt werden und von vornherein ausgeschlossen werden können (vgl. Haladyna, 2015; Downing & Haladyna, 2009; Scully, 2017).

Eine Gefahr bei Multiple-Choice-Aufgaben und insbesondere bei Binary-Choice-Aufgaben liegt darin, dass die richtige Lösung vollständig oder teilweise durch das Ausschlussverfahren erschlossen oder erraten werden kann. Rateeffekte haben einen negativen Einfluss auf die Reliabilität, Diskrimination und Validität. Durch attraktive Distraktoren, genügend Aufgaben und spezielle Scoring-Systeme lassen sich solche Effekte aber relativieren.

Dementsprechend werden Binary-Choice-Aufgaben mit Vorteil auch nicht einzeln bewertet, sondern in der Regel als Gruppen von vier Aussagen zu einer übergeordneten Aufgabenstellung, für die jeweils eindeutig bestimmt werden kann, ob sie wahr oder falsch sind. Es wird dabei nicht vorgegeben, wie viele der Aussagen richtig bzw. falsch sind. Um die Wahrscheinlichkeit auf Erfolg durch Raten zu verringern, richtet sich die Bewertung nach der Anzahl richtig angekreuzter Aussagen innerhalb der gesamten (Vierer-)Gruppe (bspw. ergeben vier richtige Antworten 2 Punkte, drei richtige Antworten 1 Punkt sowie weniger als drei richtige Antworten 0 Punkte; vgl. Lindner, Strobel & Köller, 2015).

Schliesslich sind es aber nicht nur technische Regeln, die eingehalten werden müssen, sondern vor allem auch inhaltliche Überlegungen, die vorab geklärt werden müssen. Es ist relativ einfach, reliable und diskriminierende Aufgaben zu Wortschatz und Grammatik zu entwickeln, sehr viel anspruchsvoller und zeitintensiver ist es jedoch, solche Aufgaben auch zum Leseverstehen und zum Hörverstehen zu entwickeln. Aufgaben tragen vor allem dann zur Validität bei, wenn sie für die Anforderungen des Maturitätslehrgangs relevant sind. Die Validität einer Prüfung ist erst dann gegeben, wenn damit auch wirklich das zuverlässig gemessen wird, was tatsächlich beurteilt werden soll – und nicht irgendetwas anderes.

7 Literaturverzeichnis

Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4. überarb. Aufl.). Heidelberg: Springer.

Barthel, A. (2019). Soziale Ungleichheit in der Wortschatzentwicklung von der ersten zur dritten Jahrgangsstufe. *Zeitschrift für Grundschulforschung*, 12, S. 213–228.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.

Bortz, J. & Schuster, J. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. überarb. Aufl.). Berlin: Springer.

Bundesamt für Statistik (2017). *Maturitätsquote nach Maturitätstyp, Kanton und Geschlecht*. Abgerufen am 20.07.2020 von https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/tabellen.assetdetail.2421478.html

Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences*. Hillsdale, Mahwah NJ: Lawrence Erlbaum Associates.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, S. 297–334.

Downing, S. M. & Haladyna, T. M. (2009). *Handbook of Test Development* (3. Aufl.). New York: Routledge.

Ender, S., Moser, U., Imlig, F. & Müller, S. (2017). *Bildungsbericht Nordwestschweiz 2017*. Zürich: Institut für Bildungsevaluation.

Haladyna, T. M. (2015). Developing and Validating Multiple-Choice Test Items (3. Aufl.). Mahwah NJ: Lawrence Erlbaum Associates.

Hirsig, R. (2001). *Statistische Methoden in den Sozialwissenschaften. Eine Einführung im Hinblick auf computergestützte Datenanalysen mit SPSS für Windows* (3. erw. Aufl.). Band I. Zürich: Seismo.

Konsortium PISA.ch (2019). *PISA 2018: Schülerinnen und Schüler der Schweiz im internationalen Vergleich*. Bern und Genf: SBFI/EDK und Konsortium PISA.ch.

Lindner, M. A., Strobel, B. & Köller, O. (2015). Multiple-Choice-Prüfungen an Hochschulen? – Ein Literaturüberblick und Plädoyer für mehr praxisorientierte Forschung. *Zeitschrift für Pädagogische Psychologie*, 29(3–4), S. 133–149.

Scully, D. (2017). Constructing Multiple Choice Items to Measure Higher-Order Thinking. *Practical Assessment, Research & Evaluation*, 22(4), S. 1–13.

Strobl, Carolin (2015). *Das Rasch-Modell* (3. Aufl.). München und Mering: Rainer Hampp.

Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.

8 Anhang A

Inhalt: Auftragserfüllung und Aussagekraft

1.1 Die Ausführungen passen …

- nicht oder nur teilweise zum gestellten Thema. ..... ☐ 1
- weitgehend zum gestellten Thema. ..... ☐ 2
- gut zum gestellten Thema. ..... ☐ 3

1.2 Die Fragen werden …

- nicht oder nur teilweise beantwortet. ..... ☐ 1
- weitgehend beantwortet. ..... ☐ 2
- ausführlich, detailliert beantwortet. ..... ☐ 3

1.3 Der Text ist …

- wenig gehaltvoll. ..... ☐ 1
- weitgehend gehaltvoll. ..... ☐ 2
- gehaltvoll. ..... ☐ 3

1.4 Inhaltliches Wagnis / Kreativität

- wagt wenig, wenig kreativ ..... ☐ 1
- wagt etwas, teilweise kreativ ..... ☐ 2
- wagt viel, kreativ ..... ☐ 3
- wagt sehr viel, sehr kreativ ..... ☐ 4

Textaufbau und Textzusammenhang

2.1 Der Textaufbau (äussere Gliederung – Abschnitte) ist …

- zufällig, unüberlegt, ungegliedert oder inkonsequent. ..... ☐ 1
- gegliedert (Einleitung, Hauptteil und Schluss sind ersichtlich). ..... ☐ 2

2.2 Logischer Zusammenhang (innere Gliederung – Kohärenz):
Der Text besteht aus …

- ansatzweise logisch ausgeführten Gedanken. ..... ☐ 1
- teilweise logisch ausgeführten Gedanken. ..... ☐ 2
- meistens logisch ausgeführten Gedanken. ..... ☐ 3
- logisch ausgeführten Gedanken. ..... ☐ 4

2.3 Sprachlicher Zusammenhang (innere Gliederung – Kohäsion):
Die Sätze und Abschnitte sind …

- ansatzweise sinnvoll verbunden. ..... ☐ 1
- teilweise sinnvoll verbunden. ..... ☐ 2
- meistens sinnvoll verbunden. ..... ☐ 3
- sinnvoll verbunden, auch bei komplexeren Verbindungen. ..... ☐ 4

Sprachrichtigkeit

3.1 Rechtschreibung

- ansatzweise beherrscht ..... ☐ 1
- teilweise beherrscht ..... ☐ 2
- weitgehend beherrscht ..... ☐ 3
- nahezu fehlerfrei ..... ☐ 4

3.2 Grammatik (Genus, Numerus, Kasus, Person, Tempus, Modus)

- ansatzweise beherrscht ..... ☐ 1
- teilweise beherrscht ..... ☐ 2
- weitgehend beherrscht ..... ☐ 3
- nahezu fehlerfrei ..... ☐ 4

3.3 Satzbau (Korrektheit, Vollständigkeit – inklusive Kohäsionsmittel)

- ansatzweise beherrscht ..... ☐ 1
- teilweise beherrscht ..... ☐ 2
- weitgehend beherrscht ..... ☐ 3
- nahezu fehlerfrei ..... ☐ 4

3.4 Satzzeichen (Satzschlusszeichen, Kommas, Anführungszeichen etc.)

- ansatzweise beherrscht ..... ☐ 1
- teilweise beherrscht ..... ☐ 2
- weitgehend beherrscht ..... ☐ 3
- nahezu fehlerfrei ..... ☐ 4

Sprachangemessenheit, Schreibstil und Ästhetik

4.1 Der Satzbau ist …

- wenig abwechslungsreich. ..... ☐ 1
- teilweise abwechslungsreich. ..... ☐ 2
- abwechslungsreich. ..... ☐ 3
- sehr abwechslungsreich, vielseitig. ..... ☐ 4

4.2 Die Wortwahl ist …

- wenig treffend, wenig angemessen. ..... ☐ 1
- teilweise treffend, teilweise angemessen. ..... ☐ 2
- treffend, angemessen. ..... ☐ 3
- sehr treffend, angemessen, auch bei komplexeren Begriffen. ..... ☐ 4

4.3 Der Schreibstil ist …

- sprachlich wenig gewandt. ..... ☐ 1
- sprachlich teilweise gewandt. ..... ☐ 2
- sprachlich gewandt. ..... ☐ 3
- sprachlich sehr gewandt, ausdrucksstark. ..... ☐ 4

4.4 Sprachliches Wagnis – Kreativität und Ästhetik

wagt wenig, wenig kreativ ..... ☐ 1

wagt etwas, teilweise kreativ ..... ☐ 2

wagt viel, kreativ ..... ☐ 3

wagt sehr viel, sehr kreativ ..... ☐ 4

9 Anhang B

Tabelle 43: Qualität der Teilaufgaben im Fachbereich Deutsch der Aufnahmeprüfung in die BM I von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
Teil 1: Textverständnis und Sprachbetrachtung						
1.	Fragen zum Text *(dles)*	6.0				
	a. Vier Beispiele für «Coolness wie weggeblasen»				0% (*n* = 0)	0.46
		0.0	6	4%		– 0.11
		0.5	26	17%		– 0.39
		1.0	18	12%		– 0.16
		1.5	28	18%		0.12
		2.0	78	50%		0.35
	b. Bedeutung «sein Letzigrund»				6% (*n* = 9)	0.28
		0.0	40	26%		– 0.26
		0.5	12	8%		– 0.02
		1.0	95	61%		0.26
	c. Erklärung «nun sammelte er seine Gegner ein»				0% (*n* = 0)	0.12
		0.0	11	7%		– 0.11
		0.5	5	3%		– 0.02
		1.0	140	90%		0.11
	d. Sprachliches Bild, um Schlusskampf zu beschreiben				4% (*n* = 6)	0.29
		0.0	96	62%		– 0.30
		0.5	2	1%		– 0.02
		1.0	52	33%		0.31
	e. Erklärung, warum er zum zweiten Mal die Fassung verlor und wie sich das zeigte				2% (*n* = 3)	0.10
		0.0	44	28%		– 0.11
		0.5	25	16%		– 0.02
		1.0	84	54%		0.11
2.	Richtig/Falsch ankreuzen *(dles)*	4.0				
	Der Start gelang Hussein viel besser als im Halbfinal.	0.0	3	2%	0% (*n* = 0)	0.13
		1.0	153	98%		
	Man wurde auf Kariem Hussein aufmerksam, …	0.0	17	11%	0% (*n* = 0)	0.11
		1.0	139	89%		
	Husseins Konkurrenten kamen im Finallauf …	0.0	24	15%	0% (*n* = 0)	0.24
		1.0	132	85%		
	Kariem Hussein konnte während seines …	0.0	12	8%	0% (*n* = 0)	0.09
		1.0	144	92%		
3.	Gegenteil benennen *(dles)*	4.0				
	a. Start				0% (*n* = 0)	0.24
		0.0	12	8%		– 0.26
		0.5	1	1%		0.00
		1.0	143	92%		0.25
	b. glücken				4% (*n* = 7)	0.38
		0.0	64	41%		– 0.36
		0.5	9	6%		0.03
		1.0	76	49%		0.34
	c. Sieger	0.0	13	8%	1% (*n* = 1)	0.20
		1.0	142	91%		
	d. stets				6% (*n* = 9)	0.49
		0.0	35	22%		– 0.47

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
		0.5	3	2%		– 0.04
		1.0	109	70%		0.47
4.	Synonym zuordnen *(dles)*	4.0				
	Physis	0.0	92	59%	0% (*n* = 0)	0.20
		0.5	64	41%		
	Talent	0.0	5	3%	0% (*n* = 0)	0.13
		0.5	151	97%		
	realisieren	0.0	34	22%	0% (*n* = 0)	0.23
		0.5	122	78%		
	tigern	0.0	15	10%	0% (*n* = 0)	0.26
		0.5	141	90%		
	Ökonomie	0.0	63	40%	0% (*n* = 0)	0.20
		0.5	93	60%		
	Moment	0.0	10	6%	0% (*n* = 0)	0.15
		0.5	146	94%		
	Konkurrent	0.0	2	1%	0% (*n* = 0)	0.19
		0.5	154	99%		
	Rhythmus	0.0	14	9%	0% (*n* = 0)	0.04
		0.5	142	91%		
5.	Im Text heisst es … *(dles)*	6.0				
	a. Erklären Sie die Bedeutung des				0% (*n* = 0)	0.44
	Wortes Fehltritt …	0.0	23	15%		– 0.26
		0.5	5	3%		– 0.17
		1.0	45	29%		– 0.18
		1.5	13	8%		– 0.07
		2.0	70	45%		0.45
	b. Welche Bedeutung hat das Wort				3% (*n* = 4)	0.45
	Fehltritt im übertragenen Sinn? …	0.0	87	56%		– 0.40
		0.5	8	5%		– 0.15
		1.0	27	17%		0.19
		1.5	5	3%		0.06
		2.0	25	16%		0.40
	c. Nennen Sie zwei deutsche Rede-				12% (*n* = 18)	0.50
	wendungen oder Sprichwörter …	0.0	54	35%		– 0.44
		0.5	1	1%		0.05
		1.0	55	35%		0.08
		1.5	0	0%		
		2.0	28	18%		0.42
6.	Kommas setzen *(dsif)*	2.0				
	Setzen Sie im folgenden Abschnitt die				0% (*n* = 0)	0.31
	fehlenden Kommas.	0.0	42	27%		– 0.22
		0.5	32	21%		– 0.03
		1.0	50	32%		– 0.02
		1.5	21	13%		0.23
		2.0	11	7%		0.15
7.	Wortarten bestimmen *(dsif)*	4.0				
	Jungen	0.0	2	1%	1% (*n* = 1)	0.11
		0.5	153	98%		
	vergrub	0.0	3	2%	0% (*n* = 0)	0.13
		0.5	153	98%		

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	ihn	0.0 0.5	10 146	6% 94%	0% (*n* = 0)	0.26
	jene	0.0 0.5	23 132	15% 85%	1% (*n* = 1)	0.21
	verletzlichen	0.0 0.5	5 151	3% 97%	0% (*n* = 0)	0.17
	zwischen	0.0 0.5	56 96	36% 62%	3% (*n* = 4)	0.18
	dass	0.0 0.5	40 116	26% 74%	0% (*n* = 0)	0.39
	doch	0.0 0.5	12 142	8% 91%	1% (*n* = 2)	0.18
8.	Wortfamilien *(dsif)*	3.0				
	«Grund»: Verb	0.0 0.5	28 120	18% 77%	5% (*n* = 8)	0.16
	«Grund»: Adjektiv	0.0 0.5	51 92	33% 59%	8% (*n* = 13)	0.29
	«sicher»: Nomen	0.0 0.5	20 131	13% 84%	3% (*n* = 5)	0.30
	«sicher»: Verb	0.0 0.5	27 127	17% 81%	1% (*n* = 2)	0.32
	«verzaubern»: Nomen	0.0 0.5	27 127	17% 81%	1% (*n* = 2)	0.19
	«verzaubern»: Adjektiv	0.0 0.5	109 44	70% 28%	2% (*n* = 3)	0.26
9.	Fälle bestimmen *(dsif)*	3.0				
	1. Coolness	0.0 0.5	38 114	24% 73%	3% (*n* = 4)	0.28
	2. Husseins	0.0 0.5	49 103	31% 66%	3% (*n* = 4)	0.53
	3. Augenspiel	0.0 0.5	70 83	45% 53%	2% (*n* = 3)	0.41
	4. 25-Jährige	0.0 0.5	26 127	17% 81%	2% (*n* = 3)	0.22
	5. ihn	0.0 0.5	74 78	47% 50%	3% (*n* = 4)	0.29
	6. Hose	0.0 0.5	48 104	31% 67%	3% (*n* = 4)	0.30
10.	Sätze in Zeitform setzen *(dsif)*	4.0				
	a. Futur I: «Der Start glückte viel besser als im Halbfinal.»	0.0 0.5 1.0	18 6 132	12% 4% 85%	0% (*n* = 0)	0.25 – 0.26 0.03 0.21
	b. Pefekt: «Nun leuchteten sie wieder, seine blaugrünen Augen.»	0.0 0.5 1.0	26 6 123	17% 4% 79%	1% (*n* = 1)	0.32 – 0.28 – 0.17 0.34
	c. Plusquamperfekt: «Die Gegner				1% (*n* = 2)	0.35

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	kamen zwar noch einmal näher.»	0.0	44	28%		– 0.34
		0.5	6	4%		0.01
		1.0	104	67%		0.33
	d. Präsens: «Husseins Trainer, Flavio Zberg, sollte recht behalten.»				0% (*n* = 0)	0.25
		0.0	79	51%		– 0.25
		0.5	1	1%		– 0.08
		1.0	76	49%		0.27

Bemerkung: dles = Deutsch Lesen, *dsif* = Deutsch Sprache(n) im Fokus

Tabelle 44: Qualität der Teilaufgaben im Fachbereich Englisch der Aufnahmeprüfung in die BM I von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
A 1.	Is the statement true or false? *(eles)*	8.0				
	a. Mobile phones have been banned in all schools in France.	0.0	55	35%	0% (*n* = 0)	0.23
		1.0	102	65%		
	b. The Claude Debussy School prepared its pupils by forbidding …	0.0	103	66%	1% (*n* = 1)	– 0.08
		1.0	53	34%		
	c. The response of the students of the Claude Debussy School was …	0.0	26	17%	0% (*n* = 0)	0.28
		1.0	131	83%		
	d. The teachers get the impression that the students are more …	0.0	29	18%	0% (*n* = 0)	0.14
		1.0	128	82%		
	e. Pupils cannot use their mobile phones in the classroom anymore.	0.0	104	66%	0% (*n* = 0)	– 0.03
		1.0	53	34%		
	f. According to the education minister, phone addiction receives …	0.0	104	66%	0% (*n* = 0)	– 0.03
		1.0	53	34%		
	g. The Parisian teacher believes pupils cannot concentrate anymore.	0.0	40	25%	0% (*n* = 0)	0.24
		1.0	117	75%		
	h. One of the students thinks that mobile phones are not as bad …	0.0	12	8%	0% (*n* = 0)	0.14
		1.0	145	92%		
A 2.	Vocabulary: Find a word in the text matching the definitions *(eles)*	4.0				
	a. various	0.0	58	37%	16% (*n* = 25)	0.48
		0.5	74	47%		
	b. reduce the amount of something harmful	0.0	77	49%	11% (*n* = 18)	0.36
		0.5	62	39%		
	c. pick up	0.0	18	11%	6% (*n* = 10)	0.15
		0.5	129	82%		
	d. understanding for other people	0.0	67	43%	8% (*n* = 12)	0.48
		0.5	78	50%		
	e. ground	0.0	73	46%	10% (*n* = 15)	0.54
		0.5	69	44%		
	f. reduction	0.0	83	53%	20% (*n* = 32)	0.39
		0.5	42	27%		
	g. protest	0.0	38	24%	10% (*n* = 15)	0.49
		0.5	104	66%		

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	h. pity	0.0 0.5	54 61	34% 39%	27% (*n* = 42)	0.49
B 1.	Choose the correct answer *(esif)*	6.0				
	a. Ostriches are birds but they …	0.0 0.5	17 139	11% 89%	1% (*n* = 1)	0.30
	b. When Michael was a boy, he often … fishing with his father.	0.0 0.5	25 132	16% 84%	0% (*n* = 0)	0.42
	c. Can you tell me … students participated in the fundraising event?	0.0 0.5	15 142	10% 90%	0% (*n* = 0)	0.23
	d. George … attend the meeting, but his car broke down on the way.	0.0 0.5	36 121	23% 77%	0% (*n* = 0)	0.30
	e. Brody plays the piano very …	0.0 0.5	34 123	22% 78%	0% (*n* = 0)	0.23
	f. I enjoy … bar of chocolate now and then.	0.0 0.5	69 88	44% 56%	0% (*n* = 0)	0.44
	g. Fortunately, our trip was … than we thought it would be.	0.0 0.5	59 98	38% 62%	0% (*n* = 0)	0.49
	h. Robert thinks he … in the lottery tonight.	0.0 0.5	85 72	54% 46%	0% (*n* = 0)	0.03
	i. While he … the attic, Jim came across some old photos.	0.0 0.5	34 122	22% 78%	1% (*n* = 1)	0.49
	j. By the time Jason arrived at the station, the train …	0.0 0.5	68 89	43% 57%	0% (*n* = 0)	0.37
	k. His new novel is going to … next month.	0.0 0.5	35 121	22% 77%	1% (*n* = 1)	0.45
	l. If you press that button, the machine …	0.0 0.5	15 140	10% 89%	1% (*n* = 2)	0.35
B 2.	Verb forms *(esif)*	8.0				
	wear	0.0 0.5	92 65	59% 41%	0% (*n* = 0)	0.31
	walk	0.0 0.5	42 115	27% 73%	0% (*n* = 0)	0.28
	enter	0.0 0.5	46 110	29% 70%	1% (*n* = 1)	0.45
	buy	0.0 0.5	138 19	88% 12%	0% (*n* = 0)	0.23
	come	0.0 0.5	105 52	67% 33%	0% (*n* = 0)	0.43
	scream	0.0 0.5	90 67	57% 43%	0% (*n* = 0)	0.34
	try	0.0 0.5	50 107	32% 68%	0% (*n* = 0)	0.41
	run	0.0 0.5	114 43	73% 27%	0% (*n* = 0)	0.07
	lose	0.0 0.5	135 22	86% 14%	0% (*n* = 0)	0.24
	not have	0.0 0.5	55 100	35% 64%	1% (*n* = 2)	0.49

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	shoot	0.0 0.5	76 80	48% 51%	1% (*n* = 1)	0.48
	get	0.0 0.5	102 55	65% 35%	0% (*n* = 0)	0.35
	give	0.0 0.5	60 97	38% 62%	0% (*n* = 0)	0.48
	arrive	0.0 0.5	16 141	10% 90%	0% (*n* = 0)	0.09
	disappear	0.0 0.5	141 16	90% 10%	0% (*n* = 0)	0.44
	travel	0.0 0.5	140 17	89% 11%	0% (*n* = 0)	0.04
B 3.	Make questions about the underlined word *(esif)*	2.0				
	a. The robber pointed the gun at the teller.	0.0 0.5 1.0	128 11 18	82% 7% 11%	0% (*n* = 0)	0.43 – 0.42 0.15 0.39
	b. He has had a lot of trouble lately.	0.0 0.5 1.0	112 4 41	71% 3% 26%	0% (*n* = 0)	0.30 – 0.26 – 0.23 0.35
B 4.	Negative sentences *(esif)*	2.0				
	Negative sentences 1	0.0 0.5 1.0	82 35 40	52% 22% 25%	0% (*n* = 0)	0.48 – 0.51 0.21 0.38
	Negative sentences 2	0.0 0.5 1.0	22 11 124	14% 7% 79%	0% (*n* = 0)	0.28 – 0.29 0.03 0.23

Bemerkung: eles = Englisch Lesen, *esif* = Englisch Sprache(n) im Fokus

Tabelle 45: Qualität der Teilaufgaben im Fachbereich Französisch der Aufnahmeprüfung in die BM I von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
1. Questions de compréhension						
1.1	Indiquez si l'affirmation est vraie ou fausse? *(fles)*	13.0				
	1. Thomas Adams a inventé le chewing-gum au Mexique.	0.0 1.0 2.0	69 36 50	44% 23% 32%	1% (*n* = 1)	0.10 – 0.15 0.06 0.1
	2. On peut vendre et acheter des chewing-Gums dans tous les pays du monde.	0.0 1.0	88 35	56% 22%	6% (*n* = 10)	0.38 – 0.35 0.00

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
		2.0	23	15%		0.43
	3. Avec des méthodes modernes, on peut facilement enlever les chewing-gums qui collent dans la rue.				6% (*n* = 9)	0.51
		0.0	62	40%		– 0.40
		1.0	32	21%		– 0.07
		2.0	53	34%		0.45
	4. Avec un chewing-gum on peut même provoquer des pannes …	0.0	37	24%	8% (*n* = 13)	0.32
		1.0	106	68%		
	5. Si on mâche un chewing-gum, on consomme plus de produits sucrés.				4% (*n* = 7)	0.45
		0.0	49	31%		– 0.30
		1.0	18	12%		– 0.21
		2.0	89	53%		0.41
	6. Les nazis ont donnné des chewing-gums aux enfants en Europe.				1% (*n* = 2)	0.38
		0.0	37	24%		– 0.33
		1.0	28	18%		– 0.12
		2.0	89	57%		0.35
	7. Les espagnols consomment moins de chewing-gums que les Britanniques.				2% (*n* = 3)	0.40
		0.0	31	20%		– 0.21
		1.0	24	15%		– 0.12
		2.0	98	63%		0.26
1.2	Répondes aux questions *(fles)*	7.0				
	1. Qu'est-ce qui se passe avec un chewing-gum qui colle dans la rue?				31% (*n* = 49)	0.32
		0.0	25	16%		– 0.47
		0.5	6	4%		0.09
		1.0	76	49%		0.40
	2. Quel est le prolème principal des chewing-gums?				29% (*n* = 46)	0.27
		0.0	46	29%		– 0.39
		0.5	9	6%		0.01
		1.0	55	35%		0.08
	3. Comment est-ce que Thomas Adams a fabriqué du chewing-gum?				24% (*n* = 38)	0.31
		0.0	51	33%		– 0.37
		0.5	3	2%		0.04
		1.0	64	41%		0.35
	4. Qu'est-ce que l'interdiction de fumer a changé?				48% (*n* = 75)	0.31
		0.0	35	22%		– 0.36
		0.5	4	3%		– 0.13
		1.0	42	27%		0.41
	5. Qu'est-ce qu'on peut faire avec un chewing-gum sans goût à Besançon?				49% (*n* = 77)	0.22
		0.0	22	14%		– 0.30
		0.5	7	4%		– 0.05
		1.0	50	32%		0.31
	6. Expliquez la signification du mot chewing-gum.				46% (*n* = 71)	0.44
		0.0	22	14%		– 0.54
		0.5	16	10%		– 0.02
		1.0	47	30%		0.48
	7. Qu'est-ce qui s'est passé un jour en 1992 à Paris?				19% (*n* = 29)	0.19
		0.0	20	13%		– 0.30
		0.5	12	8%		– 0.16
		1.0	95	61%		0.36
2. Vocabulaire en contexte						
2.1	Complétez les phrases par un mot du texte *(fles)*	5.0				

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
	1. Il y en a tellement dans la fôret, on ne peut pas les compter, les …	0.0 1.0	58 64	37% 41%	22% (*n* = 34)	0.42
	2. Ce que j'utilise pour manger, c'est la …	0.0 1.0	41 81	26% 52%	22% (*n* = 34)	0.32
	3. Les voitures roulent sur de …	0.0 1.0	22 102	14% 65%	21% (*n* = 32)	0.29
	4. Quand je traverse la rue, je dois faire …	0.0 1.0	17 99	11% 63%	26% (*n* = 40)	0.49
	5. Quand un pays est indépendant, ses habitants vivent en …	0.0 1.0	33 78	21% 50%	29% (*n* = 45)	0.45
2.2	Trouvez le contraire *(fsif)*	10.0				
	1. aujourd'hui	0.0 1.0	25 91	16% 58%	26% (*n* = 40)	0.38
	2. moins de	0.0 1.0	16 110	10% 71%	19% (*n* = 30)	0.27
	3. vend				46% (*n* = 72)	0.45
		0.0	29	19%		– 0.43
		0.5	8	5%		– 0.01
		1.0	47	30%		0.42
	4. jetés	0.0 1.0	46 3	29% 2%	69% (*n* =107)	0.09
	5. petite	0.0 1.0	1 146	1% 94%	6% (*n* = 9)	0.12
	6. sans	0.0 1.0	26 87	17% 56%	28% (*n* = 43)	0.49
	7. compliquées				32% (*n* = 50)	0.26
		0.0	25	16%		– 0.35
		0.5	9	6%		0.07
		1.0	72	46%		0.28
	8. arrivent				25% (*n* = 39)	0.45
		0.0	62	40%		
		0.5	2	1%		
		1.0	53	34%		
	9. blanches	0.0 1.0	8 116	5% 74%	21% (*n* = 32)	0.15
	10. longue	0.0 1.0	17 116	11% 74%	15% (*n* = 23)	0.38
2.3	Trouvez le nom de la même famille *(fsif)*	5.0				
	1. coller				21% (*n* = 32)	0.22
		0.0	90	58%		– 0.46
		0.5	8	5%		0.00
		1.0	26	17%		0.47
	2. permet				26% (*n* = 41)	0.40
		0.0	77	49%		– 0.35
		0.5	4	3%		0.00
		1.0	34	22%		0.36
	3. aiment	0.0 1.0	81 40	52% 26%	22% (*n* = 35)	0.42

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
	4. partir				31% (*n* = 49)	0.17
		0.0	93	60%		– 0.09
		0.5	5	3%		0.09
		1.0	9	6%		0.04
	5. est interdit	0.0	80	51%	38% (*n* = 60)	0.27
		1.0	16	10%		
3. Grammaire						
3.1	Les verbes *(fsif)*	10.0				
	1. partir à six heures, elle, passé composé				9% (*n* = 14)	0.43
		0.0	99	63%		– 0.45
		0.5	13	8%		0.11
		1.0	30	19%		0.42
	2. faire demo-tour, vous, présent	0.0	111	71%	8% (*n* = 12)	0.38
		1.0	33	21%		
	3. ne pas prendre de photos, nous, passé composé				9% (*n* = 14)	0.55
		0.0	95	61%		– 0.52
		0.5	6	4%		0.11
		1.0	41	26%		0.48
	4. être fatigué, tu, imparfait				11% (*n* = 17)	0.38
		0.0	83	53%		– 0.38
		0.5	11	7%		0.01
		1.0	45	29%		0.38
	5. avoir beaucoup de devoirs, ils, présent	0.0	42	27%	7% (*n* = 11)	0.35
		1.0	103	66%		
	6. ne pas jouer au foot, vous, futur composé				12% (*n* = 19)	0.43
		0.0	62	40%		– 0.41
		0.5	11	7%		– 0.02
		1.0	64	41%		0.41
	7. rester ici, tu, impératif				14% (*n* = 22)	0.36
		0.0	103	66%		– 0.38
		0.5	3	2%		0.15
		1.0	28	18%		0.34
	8. s'amuser, elles, imparfait				17% (*n* = 26)	0.37
		0.0	78	50%		– 0.46
		0.5	5	3%		0.07
		1.0	47	30%		0.43
	9. boire un café, tu, passé composé				10% (*n* = 15)	0.58
		0.0	77	49%		– 0.58
		0.5	7	4%		0.08
		1.0	57	37%		0.55
	10. ne pas dormir, vous, impératif	0.0	96	62%	17% (*n* = 26)	0.42
		1.0	34	22%		
3.2	Les pronoms *(fsif)*	5.0				
	1. Elle va rendre ce livre à Mireille.				8% (*n* = 12)	0.34
		0.0	91	58%		– 0.42
		0.5	22	14%		0.23
		1.0	31	20%		0.27
	2. Nous avons coupé le gâteau.				11% (*n* = 17)	0.50
		0.0	65	42%		– 0.44
		0.5	30	19%		– 0.09
		1.0	44	28%		0.53

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	3. Claire va souvent à Yverdon.				12% (*n* = 19)	0.42
		0.0	69	44%		– 0.46
		0.5	9	6%		– 0.03
		1.0	59	38%		0.47
	4. Je téléphonais souvent à mes copines.				15% (*n* = 24)	0.50
		0.0	77	49%		– 0.50
		0.5	5	3%		– 0.08
		1.0	50	32%		0.53
	5. Le magasin vend des fleurs.				15% (*n* = 23)	0.21
		0.0	110	71%		– 0.22
		0.5	3	2%		0.15
		1.0	20	13%		0.17
3.3	La structure de la phrase (fles)	5.0				
	1. Claudine et Marc vont au cinéma.				15% (*n* = 23)	0.65
		0.0	73	47%		– 0.65
		0.5	15	10%		0.16
		1.0	45	29%		0.56
	2. Le film commence à huit heures.				21% (*n* = 33)	0.57
		0.0	56	36%		– 0.54
		0.5	10	6%		– 0.03
		1.0	57	37%		0.54
	3. Ils regardent un film romantique.				20% (*n* = 31)	0.52
		0.0	65	42%		– 0.47
		0.5	22	14%		0.02
		1.0	38	24%		0.48
	4. Ils boivent un verre parce qu'ils ont soif.				30% (*n* = 47)	0.64
		0.0	56	36%		– 0.62
		0.5	11	7%		0.07
		1.0	42	27%		0.58
	5. Marc prend un coca.				15% (*n* = 24)	0.08
		0.0	80	51%		– 0.06
		0.5	5	3%		0.18
		1.0	47	30%		0.01
3.4	Les adjectifs *(fsif)*	5.0				
	1. Chaque jour, elle part à sept heures sur sa (beau) bicyclette.	0.0	44	28%	19% (*n* = 30)	0.46
		1.0	82	53%		
	2. Elle est très (sportif).	0.0	59	38%	13% (*n* = 20)	0.52
		1.0	77	49%		
	3. Aujourd'hui, elle porte des bottes (blanc).				14% (*n* = 22)	0.49
		0.0	51	33%		– 0.35
		0.5	6	4%		0.03
		1.0	77	49%		0.33
	4. Les filles sont très (content).	0.0	63	40%	17% (*n* = 26)	0.37
		1.0	67	43%		
	5. Jean aime ces tableaux (moderne).	0.0	53	34%	17% (*n* = 26)	0.43
		1.0	77	49%		
3.5	Les articles définis et indéfinis *(fsif)*	5.0				
	nous mangeons … pain	0.0	77	49%	7% (*n* = 11)	0.49
		0.5	68	44%		
	avec … confiture	0.0	119	76%	7% (*n* = 11)	0.32
		0.5	26	17%		

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
	nous mangeons … viande	0.0 0.5	110 32	71% 21%	9% (*n* = 14)	0.36
	avec … légumes	0.0 0.5	19 125	12% 80%	8% (*n* = 12)	0.16
	nous aimons beaucoup … légumes	0.0 0.5	115 28	74% 18%	8% (*n* = 13)	0.24
	nous ne buvons pas … vin	0.0 0.5	71 72	46% 46%	8% (*n* = 13)	– 0.05
	mais beaucoup … eau minérale	0.0 0.5	51 92	33% 59%	8% (*n* = 13)	0.35
	un paquet … biscuits	0.0 0.5	95 46	61% 29%	10% (*n* = 15)	0.32
	ou [un paquet] … pomme	0.0 0.5	74 67	47% 43%	10% (*n* = 15)	0.03
	[nous] buvons … café	0.0 0.5	81 62	52% 40%	8% (*n* = 13)	0.44

Bemerkung: fles = Französisch Lesen, *fsif* = Französisch Sprache(n) im Fokus

Tabelle 46: Qualität der Teilaufgaben im Fachbereich Mathematik der Aufnahmeprüfung in die BM I von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
1.	Terme vereinfachen *(mzuv)*	3.0				
	a. $c^2 + 10cd - c(2d + 3c) + 2c^2$				3% (*n* = 4)	0.52
		0.0	29	18%		– 0.43
		0.5	55	35%		– 0.16
		1.0	69	44%		0.47
	b. $3(2a - b) / (4a^2 - 4ab + b^2)$				6% (*n* = 9)	0.72
		0.0	79	50%		– 0.66
		0.5	17	11%		0.08
		1.0	52	33%		0.62
	c. $(7a + 2b) / 2a + (5a + 4b) / 3a$				3% (*n* = 5)	0.55
		0.0	71	45%		– 0.52
		0.5	48	31%		0.24
		1.0	33	21%		0.35
2.	Term in Faktoren zerlegen *(mzuv)*	1.0				
	$5m^3 + 25m^2 + 30m$				13% (*n* = 20)	0.53
		0.0	64	41%		– 0.44
		0.5	46	29%		0.12
		1.0	27	17%		0.39
3.	Gleichungen auflösen *(mzuv)*	3.0				
	a. $2x^2 - 3 - (x - 2)^2 = (x + 1)^2$				6% (*n* = 9)	0.68
		0.0	72	46%		– 0.60
		0.5	21	13%		– 0.02
		1.0	27	17%		0.27
		1.5	28	18%		0.49
	b. $(6x - 2) / 5 - (3x - 1) / 4 = 3$				8% (*n* = 13)	0.67

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
		0.0	37	24%		– 0.56
		0.5	36	23%		– 0.08
		1.0	25	16%		0.00
		1.5	46	29%		0.57
4.	Sachaufgabe mit Gleichung lösen *(mgfd)*	2.0				
	Priska, Maja und Stefan teilen sich				7% (*n* = 11)	0.64
	eine Wohnung. Da ihre Zimmer …	0.0	90	57%		– 0.58
		0.5	16	10%		0.03
		1.0	7	4%		0.08
		1.5	4	3%		0.14
		2.0	29	18%		0.57
5.	Sachaufgabe mit Gleichung lösen *(mgfd)*	2.0				
	Vor 9 Jahren war Sarah 5 mal so alt				15% (*n* = 23)	0.36
	wie Michael. Heute ist Sarah …	0.0	62	39%		– 0.31
		0.5	20	13%		– 0.05
		1.0	20	13%		0.03
		1.5	7	4%		0.05
		2.0	25	16%		0.39
6.	Sachaufgaben mit Gleichungen lösen *(mgfd)*	2.0				
	Herr Meier ist Kaffeehändler. Er will				6% (*n* = 10)	0.48
	aus zwei Kaffeesorten eine …	0.0	32	20%		– 0.37
		0.5	17	11%		– 0.21
		1.0	9	6%		– 0.08
		1.5	17	11%		0.00
		2.0	72	46%		0.47
7.	Grössen aufaddieren *(mgfd)*	2.0				
	a. Volumen:				8% (*n* = 12)	0.45
	3l + 6dl + 2000mm^3 + 0.07dm^3	0.0	89	57%		– 0.46
		0.5	2	1%		0.05
		1.0	54	34%		0.45
	b. Zeit:				4% (*n* = 6)	0.41
	2d + 62h + 500min + 2800s	0.0	55	35%		– 0.33
		0.5	27	17%		– 0.10
		1.0	69	44%		0.39
8.	Berechnungen in Körpern *(mfur)*	2.0				
	a. Ein Quader hat eine quadratische				6% (*n* = 9)	0.49
	Grundfläche mit der Seitenlänge …	0.0	27	17%		– 0.43
		0.5	20	13%		0.24
		1.0	101	64%		0.53
	b. Eine Pyramide besitzt eine recht-				15% (*n* = 23)	0.50
	eckige Grundfläche mit den …	0.0	81	52%		– 0.47
		0.5	20	13%		0.14
		1.0	33	21%		0.42
9.	Berechnungen im Trapez *(mfur)*	2.0				
	a. Bestimmen Sie die Winkel				21% (*n* = 33)	0.47
		0.0	42	27%		– 0.51
		0.5	18	11%		0.06
		1.0	64	41%		0.44

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
	b. Berechnen Sie die Fläche				17% (*n* = 26)	0.50
		0.0	66	42%		– 0.55
		0.5	34	22%		0.33
		1.0	31	20%		0.30
10.	Konstruktion eines Dreiecks *(mfur)*	2.0				
	Gegeben sind: Seite c, Höhe h_c und Umkreisradius r				13% (*n* = 21)	0.45
		0.0	67	43%		– 0.25
		0.5	34	22%		– 0.14
		1.0	13	8%		0.18
		1.5	6	4%		0.04
		2.0	16	10%		0.37

Bemerkung: mzuv = Mathematik Zahl und Variable, *mfur* = Mathematik Form und Raum, *mgfd* = Mathematik Grössen, Funktionen, Daten und Zufall

Tabelle 47: Qualität der Teilaufgaben im Fachbereich Englisch der Aufnahmeprüfung in die BM II von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
1.	Choose the correct answer *(esif)*	5.0				
	a. It was a really scary film and I can't stop … about it.	0.0	33	16%	0% (*n* = 1)	0.45
		0.5	170	83%		
	b. The government … Forget that there's record high unemployment …	0.0	88	43%	0% (*n* = 1)	0.48
		0.5	115	56%		
	c. If you … something hurtful to your friend in an argument, would you …	0.0	100	49%	0% (*n* = 1)	0.34
		0.5	103	50%		
	d. The death of his best friend hit him …	0.0	93	46%	0% (*n* = 1)	0.24
		0.5	110	54%		
	e. I hate crowds, so I like travelling to places that are … popular with …	0.0	14	7%	0% (*n* = 1)	0.31
		0.5	189	93%		
	f. The driver is responsible, not the person in … name the car is …	0.0	97	48%	1% (*n* = 2)	0.41
		0.5	105	51%		
	g. Most people eventually succeeded … solving the task.	0.0	139	68%	0% (*n* = 1)	0.11
		0.5	64	31%		
	h. I … you my camera if you don't look after it properly.	0.0	108	53%	0% (*n* = 1)	0.33
		0.5	95	47%		
	i. For his last birthday, he … a T-shirt he started hating the moment he …	0.0	69	34%	1% (*n* = 2)	0.57
		0.5	133	65%		
	j. … it might seem paradoxical at first sight, the latest development …	0.0	90	44%	0% (*n* = 1)	0.60
		0.5	113	55%		
2.	Ask about the underlined words *(esif)*	4.0				
	a. The car drove into the direction of the town.				0% (*n* = 1)	0.51
		0.0	78	38%		– 0.49
		0.5	15	7%		0.02
		1.0	110	54%		0.47
	b. The club will probably trade the player to China.				0% (*n* = 1)	0.06
		0.0	71	35%		– 0.06
		0.5	31	15%		0.06
		1.0	101	50%		0.01

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	c. Not everyone is good at playing fair.				0% (*n* = 1)	0.47
		0.0	64	31%		– 0.42
		0.5	49	24%		– 0.07
		1.0	90	44%		0.45
	d. The journey takes three and a half hours.				0% (*n* = 1)	0.65
		0.0	81	40%		– 0.62
		0.5	4	2%		0.00
		1.0	118	58%		0.61
3.	Make the following sentences negative of positive *(esif)*	4.0				
	a. After lunch he lay down on the sofa.				0% (*n* = 1)	0.21
		0.0	186	91%		– 0.24
		0.5	1	0%		– 0.03
		1.0	16	8%		0.25
	b. Do they usually meet at the gym?				1% (*n* = 2)	0.46
		0.0	56	27%		– 0.48
		0.5	19	9%		0.14
		1.0	127	62%		0.35
	c. She doesn't normally question her decisions.				0% (*n* = 1)	0.57
		0.0	44	22%		– 0.38
		0.5	82	40%		– 0.22
		1.0	77	38%		0.54
	d. The situation could have been worse.				0% (*n* = 1)	0.36
		0.0	32	16%		– 0.35
		0.5	12	6%		– 0.06
		1.0	159	78%		0.34
4.	Fill in the correct verb form *(esif)*	8.0				
	a. send	0.0	188	92%	0% (*n* = 1)	0.33
		0.5	15	7%		
	b. break	0.0	75	37%	0% (*n* = 1)	0.47
		0.5	128	63%		
	c. cross	0.0	42	21%	0% (*n* = 1)	0.35
		0.5	161	79%		
	d. achieve	0.0	44	22%	0% (*n* = 1)	0.27
		0.5	159	78%		
	e. say	0.0	31	15%	1% (*n* = 2)	0.32
		0.5	171	84%		
	f. be	0.0	65	32%	1% (*n* = 2)	0.38
		0.5	137	67%		
	g. feel	0.0	48	24%	1% (*n* = 2)	0.34
		0.5	154	75%		
	h. feel	0.0	50	25%	1% (*n* = 2)	0.36
		0.5	152	75%		
	i. begin	0.0	57	28%	1% (*n* = 2)	0.36
		0.5	145	71%		
	j. sleep	0.0	192	94%	1% (*n* = 2)	0.19
		0.5	10	5%		
	k. be	0.0	30	15%	1% (*n* = 2)	0.23
		0.5	172	84%		

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	l. tell	0.0 0.5	48 154	24% 75%	1% (*n* = 2)	0.37
	m. never/do	0.0 0.5	137 65	67% 32%	1% (*n* = 2)	0.33
	n. ever/manage	0.0 0.5	113 88	55% 43%	1% (*n* = 3)	0.42
	o. love	0.0 0.5	186 16	91% 8%	1% (*n* = 2)	0.21
	p. be	0.0 0.5	31 171	15% 84%	1% (*n* = 2)	0.17
5.	Reading comprehension I: Are the statements true or false? *(eles)*	7.0				
	a. The USA ranked in 13th place in the 2016 report.	0.0 1.0	30 170	15% 85%	0% (*n* = 1)	0.11
	b. People in Nordic countries don't mind paying high taxes because …	0.0 1.0	14 189	7% 93%	0% (*n* = 1)	0.21
	c. America's fall in happiness is explained by growing obesity, drug …	0.0 1.0	54 149	26% 73%	0% (*n* = 1)	0.00
	d. Togo in west Africa improved its score the most whereas Syria …	0.0 1.0	12 191	6% 94%	0% (*n* = 1)	0.19
	e. In general, people in Latin American countries are happier than …	0.0 1.0	33 169	16% 83%	1% (*n* = 2)	0.08
	f. The report suggests that wealth is less important than family bonds …	0.0 1.0	53 148	26% 73%	1% (*n* = 3)	0.11
	g. People in China who moved from the countryside to big cities are …	0.0 1.0	8 195	4% 96%	0% (*n* = 1)	0.00
6.	Reading comprehension II: Complete the sentence	4.0				
	a. If a question relates to a particular race of people, it is an … issue.	0.0 0.5	94 102	46% 50%	4% (*n* = 8)	0.50
	b. The quality of being kind, understanding, and not selfish and …	0.0 0.5	105 96	51% 47%	1% (*n* = 3)	0.55
	c. If something takes or has a top position in relation to others, …	0.0 0.5	85 116	42% 57%	1% (*n* = 3)	0.33
	d. If something does not normally change very much, it is …	0.0 0.5	39 163	19% 80%	1% (*n* = 2)	0.42
	e. … is the satisfactory state that someone should be in, involving …	0.0 0.5	113 87	55% 43%	2% (*n* = 4)	0.46
	f. If something becomes greater in size, amount, number, or …	0.0 0.5	60 142	29% 70%	1% (*n* = 2)	0.58
	g. The use of physical force so as to injure, abuse, damage, or …	0.0 0.5	70 130	34% 64%	2% (*n* = 4)	0.39
	h. People living outside urban areas live in the …	0.0 0.5	23 179	11% 88%	1% (*n* = 2)	0.24

Bemerkung: eles = Englisch Lesen, *esif* = Englisch Sprache(n) im Fokus

Tabelle 48: Qualität der Teilaufgaben im Fachbereich Französisch der Aufnahmeprüfung in die BM II von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
1. Compréhension de texte						
1.1	Cherchez dans le texte un mot qui exprime la même chose *(fles)*	4.0				
	a. les gens qui lisent	0.0 1.0	158 22	77% 11%	12% (*n* = 25)	0.05
	b. bête	0.0 1.0	65 64	32% 31%	37% (*n* = 76)	0.39
	c. ajouter ce qui manque, ce que n'est pas là	0.0 1.0	108 8	53% 4%	43% (*n* = 89)	0.28
	d. généralement	0.0 1.0	44 144	21% 70%	8% (*n* = 17)	0.12
1.2	Cherchez dans le texte l'expression française *(fles)*	2.0				
	a. Daraus folgt	0.0 1.0	55 141	27% 69%	4% (*n* = 9)	0.01
	b. die Mundpropaganda	0.0 1.0	39 155	19% 76%	5% (*n* = 11)	0.28
1.3	Cherchez dans le texte un passage qui exprime l'idée suivante *(fles)*	4.0				
	a. Les jeunes lisent des romans traditionnels seulement à l'école.	0.0 0.5 1.0	40 3 149	20% 1% 73%	6% (*n* = 13)	0.38 – 0.39 0.02 0.37
	b. Très souvent, les jeunes parlent d'un bon livre à des copains.	0.0 1.0	43 135	21% 66%	13% (*n* = 27)	0.29
	c. La personne qui écrit des romans n'est plus dans la même situation …	0.0 1.0	58 125	28% 61%	11% (*n* = 22)	0.31
	d. On peut combiner tout ça.	0.0 1.0	37 120	18% 59%	23% (*n* = 48)	0.33
1.4	Que savez-vous de la lecture des jeunes d'aujourd'hui? *(fles)*	3.0				
	Donnez trois exemples en allemand	0.0 1.0 2.0 3.0	55 71 21 34	27% 35% 10% 17%	12% (*n* = 24)	0.47 – 0.44 0.08 0.16 0.26
1.5	Ecrivez en allemand *(fles)*	1.0				
	Quel type de livres est-ce que les jeunes aiment moins ?	0.0 1.0	71 114	35% 56%	10% (*n* = 20)	0.39
2. Grammaire						
2.1	Complétez le tableau *(fsif)*	12.0				
	on complète: infinitif	0.0 0.5	159 38	78% 19%	4% (*n* = 8)	0.17
	on complète: trad. allemande	0.0 0.5	25 173	12% 84%	3% (*n* = 7)	0.30

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
	on complète: imparfait	0.0 0.5	170 26	83% 13%	4% (*n* = 9)	0.30
	écrire: trad. allemande	0.0 0.5	4 200	2% 98%	0% (*n* = 1)	0.19
	écrire: présent	0.0 0.5	43 160	21% 78%	1% (*n* = 2)	0.40
	écrire: imparfait	0.0 0.5	86 110	42% 54%	4% (*n* = 9)	0.63
	il voulait: infinitif	0.0 0.5	53 144	26% 70%	4% (*n* = 8)	0.45
	il voulait: trad. allemande	0.0 0.5	26 160	13% 78%	9% (*n* = 19)	0.32
	in voulait: présent	0.0 0.5	129 71	63% 35%	2% (*n* = 5)	0.60
	il faut: infinitif	0.0 0.5	98 82	48% 40%	12% (*n* = 25)	0.67
	il faut: trad. allemande	0.0 0.5	75 96	37% 47%	17% (*n* = 34)	0.61
	il faut: imparfait	0.0 0.5	125 54	61% 26%	13% (*n* = 26)	0.58
	gehen: infinitif	0.0 0.5	15 186	7% 91%	2% (*n* = 4)	0.31
	gehen: trad. Présent	0.0 0.5	72 124	35% 60%	4% (*n* = 9)	0.54
	gehen: imparfait	0.0 0.5	68 125	33% 61%	6% (*n* = 12)	0.57
	antworten: trad. allemande[1]	0.0 0.5	97 92	47% 45%	8% (*n* = 16)	0.32
	antworten: présent	0.0 0.5	126 52	61% 25%	13% (*n* = 27)	0.46
	antworten: imparfait	0.0 0.5	130 47	63% 23%	14% (*n* = 28)	0.49
	sehen: infinitif	0.0 0.5	48 142	23% 69%	7% (*n* = 15)	0.28
	sehen: présent	0.0 0.5	65 124	32% 60%	8% (*n* = 16)	0.56
	sehen: imparfait	0.0 0.5	108 73	53% 36%	12% (*n* = 24)	0.62
	il permet: infinitif	0.0 0.5	86 107	42% 52%	6% (*n* = 12)	0.59
	il permet: trad. allemande	0.0 0.5	65 91	32% 44%	24% (*n* = 49)	0.10
	il permet: imparfait	0.0 0.5	133 58	65% 28%	7% (*n* = 14)	0.57
2.2	Mettez les phrases au passé composé *(fsif)*	6.0				
	a. Voilà les livres de Steven King. - Oui, je les lis.	0.0	112	55%	7% (*n* = 15)	0.61 – 0.57

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
		1.0	44	21%		– 0.02
		2.0	34	17%		0.62
	b. Je n'ouvre jamais mes livres.	0.0	141	69%	8% (*n* = 17)	0.53
		1.0	47	23%		
	c. Mes livres tombent par terre.				7% (*n* = 15)	0.67
		0.0	94	46%		– 0.60
		0.5	10	5%		– 0.10
		1.0	86	42%		0.64
	d. Les livres de Harry Potter se vendent bien.				8% (*n* = 17)	0.52
		0.0	135	66%		– 0.48
		0.5	3	1%		0.04
		1.0	50	24%		0.47
	e. Mon frère met les livres sur mon bureau.	0.0	82	40%	9% (*n* = 18)	0.68
		1.0	105	51%		
2.3	Remplacez la partie soulignée par un pronom *(fsif)*	6.0				
	a. Est-ce que tu as rendu les livres à ta copine? [2]				6% (*n* = 12)	0.69
		0.0	77	38%		
		0.5	0	0%		
		1.0	31	15%		
		1.5	2	1%		
		2.0	46	22%		
		2.5	5	2%		
		3.0	32	16%		
	b. Je vais à la bibliothèque une fois par mois.				6% (*n* = 12)	0.64
		0.0	75	37%		– 0.61
		0.5	34	17%		0.12
		1.0	84	41%		0.49
	c. Nous aimons discuter les poésies avec nos copains.	0.0	103	50%	7% (*n* = 14)	0.63
		1.0	88	43%		
	d. Je ne sais pas lire cette revue anglaise.	0.0	152	74%	7% (*n* = 14)	0.52
		1.0	39	19%		
2.4	Formulez les questions en français standard *(fsif)*	4.0				
	a. Elle n'a pas terminé ce livre parce qu'il est ennuyeux.				6% (*n* = 12)	0.59
		0.0	96	47%		– 0.51
		0.5	14	7%		0.04
		1.0	83	40%		0.53
	b. Luc adore les livres de Harry Potter.				2% (*n* = 5)	0.57
		0.0	86	42%		– 0.49
		0.5	19	9%		– 0.07
		1.0	95	46%		0.52
	c. Luc adore les livres de Harry Potter.				4% (*n* = 9)	0.49
		0.0	84	41%		– 0.46
		0.5	4	2%		0.01
		1.0	108	53%		0.45
	d. J.K. Rowling a 53 ans.				6% (*n* = 12)	0.42
		0.0	146	71%		– 0.40
		0.5	23	11%		0.21
		1.0	24	12%		0.31
3. Vocabulaire						

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierig-keit	nicht bearbeitet	Trenn-schärfe
3.1	Complétez les phrases par un mot *(fsif)*	9.0				
	a. 3'000 francs pour un ordinateur! Mais c'est trop …				11% (*n* = 23)	0.48
		0.0	84	41%		– 0.41
		0.5	24	12%		0.05
		1.0	74	36%		0.38
	b. Hier, je suis allé au … seulement après minuit.				20% (*n* = 41)	0.38
		0.0	81	40%		– 0.36
		0.5	9	4%		0.15
		1.0	74	36%		0.29
	b. C'est pourquoi, aujourd'hui, je suis très …				20% (*n* = 41)	0.39
		0.0	124	60%		– 0.37
		0.5	3	1%		– 0.08
		1.0	37	18%		0.41
	b. Ce soir, je me couche plus …				25% (*n* = 51)	0.42
		0.0	110	54%		– 0.42
		0.5	2	1%		0.04
		1.0	42	20%		0.42
	c. Je suis très content de mon salaire. Je … plus de 5'000 francs par mois.				21% (*n* = 44)	0.46
		0.0	103	50%		– 0.45
		0.5	1	0%		– 0.06
		1.0	57	28%		0.46
	d. Cet après-midi, j'ai un rendez-vous chez le …				19% (*n* = 39)	0.29
		0.0	100	49%		– 0.27
		0.5	7	3%		0.04
		1.0	59	29%		0.26
	d. J'ai tellement mal à la tête et 39 degrés de …				19% (*n* = 39)	0.47
		0.0	88	43%		– 0.40
		0.5	26	13%		– 0.07
		1.0	52	25%		0.48
	e. La rose est ma … préférée.				9% (*n* = 18)	0.14
		0.0	76	37%		– 0.17
		0.5	4	2%		– 0.02
		1.0	107	52%		0.17
	f. Quand je vais aux États-Unis, je préfère voyager en …				13% (*n* = 27)	0.19
		0.0	74	36%		– 0.25
		0.5	5	2%		0.00
		1.0	99	48%		0.24
3.2	Quel est le contraire des mots *(fsif)*	7.0				
	a. chercher	0.0	28	14%	12% (*n* = 25)	0.47
		1.0	152	74%		
	b. sur	0.0	47	23%	7% (*n* = 14)	0.45
		1.0	144	70%		
	c. descendre	0.0	77	38%	19% (*n* = 38)	0.51
		1.0	90	44%		
	d. facile				11% (*n* = 23)	0.37
		0.0	51	25%		– 0.33
		0.5	8	4%		– 0.14
		1.0	123	60%		0.38
	e. avec	0.0	58	28%	20% (*n* = 40)	0.45
		1.0	107	52%		

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
	f. devant la maison				9% (*n* = 19)	0.55
		0.0	100	49%		– 0.48
		0.5	20	10%		0.01
		1.0	66	32%		0.48
	f. tout	0.0	63	31%	16% (*n* = 33)	0.37
		1.0	109	53%		

Bemerkung: fles = Französisch Lesen, *fsif* = Französisch Sprache(n) im Fokus
[1] In der Aufgabenstellung steht die Vorgabe «antworten» im falschen Feld (*infinitif* anstatt *trad. allemande*).
[2] Aufgrund der hohen Anzahl an Kategorien wurde diese Aufgabe für die Trennschärfeberechnung dichotomisiert.

Tabelle 49: Qualität der Teilaufgaben im Fachbereich Mathematik der Aufnahmeprüfung in die BM II von 2019

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
1.	Aufgabe zum Dreieck *(mfur)*	2.0				
	a. Konstruieren Sie ein Dreieck aus folgenden Grössen …				1% (*n* = 3)	0.44
		0.0	42	20%		– 0.41
		0.5	67	33%		– 0.01
		1.0	93	45%		0.34
	b. Berechnen Sie aus den gegebenen Grössen die Höhe des Dreiecks.				16% (*n* = 33)	0.54
		0.0	60	29%		– 0.49
		0.5	22	11%		– 0.06
		1.0	90	44%		0.50
2.	Grössen umwandeln *(mgfd)*	2.0				
	a. Länge	0.0	19	9%	0% (*n* = 0)	0.20
		0.5	186	91%		
	b. Zeit	0.0	48	23%	0% (*n* = 1)	0.37
		0.5	156	76%		
	c. Volumen	0.0	81	40%	0% (*n* = 1)	0.37
		0.5	123	60%		
	d. Fläche	0.0	85	41%	6% (*n* = 12)	0.33
		0.5	108	53%		
3.	Berechnungen in Körpern *(mfur)*	2.0				
	Berechnen Sie das Volumen und die Oberfläche des dargestellten Körpers.				4% (*n* = 8)	0.38
		0.0	25	12%		– 0.23
		0.5	14	7%		– 0.21
		1.0	79	39%		– 0.02
		1.5	39	19%		0.03
		2.0	40	20%		0.32
4.	Terme umformen *(mzuv)*	2.0				
	a. $(x - y)^3$				2% (*n* = 4)	0.43
		0.0	116	57%		– 0.38
		0.5	35	17%		0.02
		1.0	50	24%		0.41
	b. $(3a - 3b)/(a + b) : 5(a - b)^2/(a^2 - b^2)$				6% (*n* = 13)	0.51
		0.0	73	36%		– 0.46
		0.5	42	20%		0.04
		1.0	77	38%		0.42

Aufgabe	Teilaufgaben	Punkte	*N*	Schwierigkeit	nicht bearbeitet	Trennschärfe
5.	Gleichungen lösen *(mzuv)*	2.0				
	a. a(10 – (x – 2)) = (-5)(10x – 4)				2% (*n* = 5)	0.48
		0.0	68	33%		– 0.33
		0.5	57	28%		– 0.14
		1.0	75	37%		0.45
	b. $x(2x – 3) / 2 = (x – 5)^2$				8% (*n* = 16)	0.44
		0.0	68	33%		– 0.42
		0.5	29	14%		0.01
		1.0	92	45%		0.40
6.	Sachaufgaben mit Gleichungen lösen *(mgfd)*	2.0				
	Ein Kapital von 70'350 CHF wird in zwei unterschiedlich grossen …				13% (*n* = 27)	0.58
		0.0	97	47%		– 0.46
		0.5	25	12%		– 0.01
		1.0	13	6%		0.08
		1.5	1	0%		0.02
		2.0	42	20%		0.50
7.	Konstruktion im Koordinatensystem *(mfur)*	2.0				
	a. Zeichnen Sie die Ausgangssituation im Koordinatensystem ein. b. Konstruieren Sie die Spiegelachse *g* und das gespiegelte Dreieck.				4% (*n* = 8)	0.55
		0.0	14	7%		– 0.33
		0.5	65	32%		– 0.33
		1.0	16	8%		– 0.02
		1.5	6	3%		0.06
		2.0	96	47%		0.46
8.	Sachaufgabe mit Skizze *(mfur)*	2.0				
	Bei einem Quadrat werden zwei gegenüberliegende Seiten …				6% (*n* = 12)	0.55
		0.0	37	18%		– 0.34
		0.5	55	27%		– 0.23
		1.0	25	12%		– 0.02
		1.5	4	2%		0.02
		2.0	72	35%		0.50
9.	Gleichung aufstellen und lösen *(mzuv)*	2.0				
	Das arithmetische Mittel (der Durchschnitt) zweier Zahlen beträgt 40. …				12% (*n* = 24)	0.54
		0.0	48	23%		– 0.56
		0.5	16	8%		– 0.02
		1.0	7	3%		0.04
		1.5	4	2%		0.23
		2.0	106	52%		0.42
10.	ggT und kgV *(mzuv)*	2.0				
	Bestimmen Sie mit Hilfe der Primfaktorenzerlegung den ggT und …				6% (*n* = 12)	0.49
		0.0	55	27%		– 0.39
		0.5	10	5%		– 0.10
		1.0	66	32%		0.02
		1.5	10	5%		0.09
		2.0	52	25%		0.37

Bemerkung: mzuv = Mathematik Zahl und Variable, *mfur* = Mathematik Form und Raum, *mgfd* = Mathematik Grössen, Funktionen, Daten und Zufall