



universität
wien

Diplomarbeit

Titel der Arbeit

Evaluation des Eignungstests für das Medizinstudium
(EMS) in Österreich - Erklärungsvariablen für
Leistungsunterschiede

Verfasser:

Ing. Andreas Pfaffel

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im November 2010

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Ao. Univ.-Prof. Dipl.-Psych. Dr. Barbara Schober

Danksagung

Bedanken möchte ich mich bei meiner Betreuerin Prof. Dr. Barbara Schober für die wertvollen Hilfestellungen und Ratschläge bei schwierigeren Passagen beim Schreiben meiner Diplomarbeit.

Auch danke ich den Evaluatorinnen der Projektstudie Prof. Dr. Dr. Christiane Spiel, Prof. Dr. Barbara Schober und Dr. Magarete Litzenberger, die mir mit ihrer umfassenden Studie und dem riesengroßen Datensatz mein Diplomarbeitsthema erst ermöglichten.

Danke auch den beiden Projektmitarbeiterinnen Michaela Fasching und Eva-Maria Schiller.

Ein herzliches Dankeschön auch an Dr. Jürgen Falb, der mir bei kniffligen technischen Umsetzungen in \LaTeX half.

Auch dem gesamten Arbeitsbereich Bildungspsychologie und Evaluation der Fakultät für Psychologie danke ich für die Betreuung und Unterstützung, von der ich nicht nur bei meiner Diplomarbeit profitiert habe.

Ein liebes Danke auch an alle meine Freunde, die mich in der Abschlussphase meines Studiums unterstützt haben.

Inhaltsverzeichnis

1. Einleitung	7
2. Auswahlverfahren für den Hochschulzugang	9
2.1. Ziele von Auswahlverfahren	9
2.2. Anforderungen an Auswahlverfahren	11
2.2.1. Anforderungen unter allgemeiner, bildungswissenschaftlicher und psychologischer Sicht	11
2.2.2. Bedeutung eignungsdiagnostischer Gütekriterien	15
2.3. Auswahlverfahren an den Medizinischen Universitäten Wien und Inns- bruck	21
2.4. Der Eignungstest für das Medizinstudium (EMS)	22
3. Prädiktoren der Leistung	25
3.1. Schulnoten	26
3.1.1. Diagnostische Gütekriterien von Schulnoten	28
3.2. Schultypen	35
3.2.1. Schultypen und Lehrziele	37
3.2.2. Lehrpläne – AHS-Oberstufe und HTL	38
3.2.3. Leistungsunterschiede zwischen Schultypen – Empirische Eviden- zen	40
3.3. Regionale Herkunft - Stadt/Land	44
3.4. Zusammenfassung	48
4. Fragestellungen und Hypothesen	50
4.1. Schulische Aspekte	52
4.1.1. Schulnoten	52
4.1.2. Schultyp	54
4.2. Regionale Herkunft	55
4.3. Geschlecht und Nationalität	55
5. Methode	57
5.1. Untersuchungsdurchführung/Design	57
5.2. Erhebungsinstrumente	58
5.2.1. Eignungstest für das Medizinstudium - EMS	58

5.2.2. Fragebogen	59
5.2.3. Soziodemographische Daten im Rahmen des Auswahlverfahrens	62
5.3. Beschreibung der Stichprobe	62
5.3.1. Repräsentativität	63
5.3.2. Geschlecht und Nationalität	63
5.3.3. Regionale Herkunft	64
5.3.4. Schultyp	65
5.4. Vorgehen bei der Testverkürzung	66
5.5. Vorgehensweise bei der Auswertung und eingesetzte Verfahren	66
6. Ergebnisse	69
6.1. Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität	69
6.1.1. Testwert (Einzelfachnoten)	69
6.1.2. Untertests (Einzelfachnoten)	72
6.2. Leistungsunterschiede in Abhängigkeit von der Durchschnittsnote, regionaler Herkunft, Geschlecht und Nationalität	74
6.2.1. Testwert (Durchschnittsnote)	74
6.2.2. Untertests (Durchschnittsnote)	76
6.3. Einfluss des Schulnotenprofils	77
6.4. Leistungsunterschiede in Abhängigkeit von Schultyp, Durchschnittsnote und Geschlecht	78
6.4.1. Leistungsunterschiede in Abhängigkeit von Schultyp und Durchschnittsnote	78
6.4.2. Leistungsunterschiede in Abhängigkeit von Schultyp und Geschlecht	79
7. Zusammenfassung der Ergebnisse	81
8. Diskussion	85
9. Literatur	91
A. Anhang	102
A.1. Auswahlverfahren für den Hochschulzugang	102
A.1.1. Überblick über die Untertests des EMS nach Hänsgen und Spicher, 2007	104
A.2. Struktur des deutschen Bildungssystems	110

A.3. Fragebogen	111
A.4. Beschreibung der Stichprobe	113
A.4.1. Verteilung der Nationalitäten	113
A.4.2. Regionale Herkunft	114
A.4.3. Schultyp	115
A.5. Bildung des Notenprofils	116
A.6. Deskriptive Ergebnisse	117
A.7. Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität	118
A.7.1. Testwert (Einzelfachnoten)	118
A.7.2. Untertests (Einzelfachnoten)	120
A.8. Leistungsunterschiede in Abhängigkeit von der Durchschnittsnote, re- gionaler Herkunft, Geschlecht und Nationalität	123
A.8.1. Testwert (Durchschnittsnote)	123
A.8.2. Untertests (Durchschnittsnote)	124
A.9. Einfluss des Schulnotenprofils	126
A.10. Leistungsunterschiede in Abhängigkeit vom Schultyp, Durchschnittsnote und Geschlecht	127
B. Lebenslauf	131

1. Einleitung

In Folge des Urteils des Europäischen Gerichtshofs (EuGH) vom 7. Juli 2005 (Rs C-147/03, Kommission/Österreich¹), hat der österreichische Nationalrat am 8. Juli 2005 eine Novellierung des Universitätsgesetzes 2002 (BGBl. I Nr. 77/2005) beschlossen, welche die Beschränkung des Zugangs zu bestimmten Studienfächern an österreichischen Universitäten ermöglicht (§ 124b Universitätsgesetz 2002). Dies war eine Reaktion auf die erhöhte Nachfrage von Studienanfängerinnen und -anfängern aus der Bundesrepublik Deutschland für das Studium der Humanmedizin. Die Einführung der Zugangsbeschränkungen wurde auch in den Medien heftig diskutiert. Ab dem Studienjahr 2006/07 wurden an den drei Medizinischen Universitäten Wien, Innsbruck und Graz Auswahlverfahren durchgeführt. Die Medizinischen Universitäten Wien und Innsbruck orientierten sich dabei an einem gemeinsamen Testverfahren, dem Schweizer 'Eignungstest für das Medizinstudium' (EMS) der auf dem deutschen 'Test für medizinische Studiengänge' (TMS), basiert. An der Medizinischen Universität Graz wurde ein eigenes Testverfahren entwickelt und vorgegeben.

Bereits aus den ersten Ergebnissen der Auswahlverfahren war bekannt, dass Frauen im Test schlechter abschnitten als Männer und österreichische Kandidatinnen und Kandidaten schlechter als deutsche. Dies zeigte eine deutliche Auswirkung auf die Geschlechts- und Nationalitätenverteilung der Medizinstudierenden, die in Folge zu einer erneuten Novellierung des Universitätsgesetzes 2002 (BGBl. I Nr. 74/2006) führte, der 'Safeguard-Klausel', mit drei Kontingenten (gemäß § 124b Abs 5 Universitätsgesetz 2002) für Studienplätze: 1. Österreich-Kontingent (75%), 2. EU-Kontingent (20%), 3. Nicht-EU-Kontingent (5%). Um mögliche Erklärungen und Ursachen für die Leistungsunterschiede im Auswahlverfahren aufzudecken, evaluierten Spiel, Schober und Litzenberger (2008) im Auftrag des Bundesministeriums für Wissenschaft und Forschung (bm:wf) die Auswahlverfahren an den drei Medizinischen Universitäten. Die empirische Studie wurde mit allen Studienbewerberinnen und Studienbewerbern, die am Auswahlverfahren am 6. Juli 2007 teilnahmen durchgeführt.

Als Projektmitarbeiter an der Studie von Spiel und Mitarbeiterinnen (2008) war ich in alle wesentlichen Teile eingebunden. Neben der nach wie vor geführten Diskussion über Zugangsbeschränkungen und Auswahlverfahren sowie der bildungspolitischen Re-

¹Österreich hat gegen seine Verpflichtung aus dem Gemeinschaftsrecht (resultierend aus den Artikeln 12, 149 und 150 EGV) verstoßen. Es wurde nicht sichergestellt, dass alle Inhaberinnen und Inhaber von Sekundarschulabschlüssen der Mitgliedsstaaten die gleichen Voraussetzungen für den Zugang zum Hochschul- und Universitätsstudium in Österreich haben.

levanz ist die Auswahl des Themas meiner Diplomarbeit sicher auch auf einen persönlichen Bezug durch meine Projektstätigkeit zurückzuführen.

In meiner Diplomarbeit möchte ich Bereiche betrachten, die im damaligen Projekt-auftrag nicht enthalten waren und bis jetzt noch nicht oder nicht so detailliert analysiert wurden. Zentrale Aspekte der empirischen Studie von Spiel et al. (2008) waren die Analyse der Chancengleichheit hinsichtlich der Merkmale Geschlecht, Nationalität, soziale Herkunft und naturwissenschaftlicher versus nicht-naturwissenschaftlicher Schultyp in Bezug auf die Testleistung. Es wurden auch testtheoretische Analysen nach klassischer und nach probabilistischer Testtheorie der beiden Eignungstests durchgeführt.

Der Eignungstest an der Medizinischen Universität Graz wird aufgrund der Eigenständigkeit und Unterschiedlichkeit zum EMS sowie der relativ geringen Stichprobe nicht Inhalt meiner Diplomarbeit sein. Ebenso wegen der geringen Stichprobe der Bewerberinnen und Bewerber für die Studienrichtung Zahnmedizin werden sich die Analysen ausschließlich auf Bewerberinnen und Bewerber für das Studium der Humanmedizin beschränken.

Die vorliegende Arbeit gliedert sich in zwei große Abschnitte: Der erste, theoretische Teil wird sich mit wissenschaftlichen Erkenntnissen zu Zielen von und Anforderungen an Auswahlverfahren beschäftigen. Der Nutzen eignungsdiagnostischer Gütekriterien wird diskutiert. Unter diesen Aspekten wird zum Eignungstest für das Medizinstudium (EMS) übergeleitet und vorliegende Ergebnisse und testtheoretische Befunde zum EMS thematisiert. Ein weiteres Kapitel des theoretischen Teils widmet sich den Prädiktoren zur Leistungsvorhersage. Primär sind dies die Prädiktoren Schulnoten, Schultyp und regionale Herkunft die hinsichtlich ihrer Prädiktorqualität anhand von empirischen Befunden diskutiert werden. Auf Geschlecht und Nationalität wird nur kurz eingegangen, sie sind nicht zentrales Thema der hier vorliegenden empirischen Studie. Da Spiel et al. (2008) bereits hinreichend ihren prognostischen Wert auf die Testleistung zeigen konnten, können sie jedoch nicht aus den empirischen Analysen ausgeschlossen werden. Im Anschluss werden die Fragestellungen der hier vorliegenden Arbeit präsentiert.

Der zweite Teil der Arbeit informiert über den empirischen Teil der Studie. Er beinhaltet die Beschreibung des Untersuchungsdesigns und der Untersuchungsdurchführung, der Stichprobe sowie der eingesetzten Erhebungsinstrumente. Auch eine kurze Beschreibung der verwendeten Auswertungsverfahren wird gegeben. Sämtliche Analysen werden fragestellungsbezogen im Ergebnisteil dargestellt. Im Anschluss an die Ergebnisdarstellung folgt eine Zusammenfassung und Diskussion. Die Ergebnisse werden in Relation zur Literatur diskutiert, Stärken und mögliche Einschränkungen der Arbeit werden besprochen.

2. Auswahlverfahren für den Hochschulzugang

Verfahren, die die Zulassung, Genehmigung oder Beschränkung von Personen regeln, finden sich über die Geschichte hinweg in fast allen Lebensbereichen. Häufig waren und sind es noch immer Verfahren zur Auswahl von Personen für militärische Zwecke, Berufe, Sport, für räumliche Zugangsbeschränkungen (bspw. Einwanderung) und zur Beschränkung des Zugangs zu Bildungsstätten. Die dahinter liegende Annahme ist, dass sich Menschen in bestimmten Traits stabil unterscheiden und diese auch generalisiert werden können (vgl. DuBois, 1970). Die Ressourcen sind meist beschränkt und können daher nur einer bestimmten Anzahl von Personen zugänglich gemacht werden. Durch ein Verfahren (bspw. einen Test) wird versucht eine Auswahl an Personen die sich nach bestimmten Kriterien 'eignen' zu gewinnen.

Die Testergebnisse haben häufig für die getesteten Personen individuelle Konsequenzen. Sie entscheiden z.B. ob das Studienfach der Wahl studiert werden darf oder die gewünschte berufliche Tätigkeit erlernt oder ausgeübt werden darf. Diese Entscheidungen haben durch ihre Systematik entscheidende individuelle, bildungs- und gesellschaftspolitische Konsequenzen und dürfen daher „nicht dem Zufall oder theoretisch nicht (ausreichend) fundierten und empirisch nicht (ausreichend) erprobten Auswahlverfahren überlassen werden“ (Spiel, Litzberger & Haiden, 2007, S. 479). Die psychologische Diagnostik, die sich mit der Testung und Begutachtung von Personen beschäftigt, kann einen Beitrag leisten, um Eignungsbeurteilungen effektiver und effizienter zu gestalten. Nicht nur das Know how soll gegeben werden, sondern auch Prozesse in Gang gesetzt und auf ihre Wirksamkeit und Kosten-Nutzen-Relation hin evaluiert werden. Schmitt (2005) sieht dringenden Bedarf, die oft vorhandenen Vorstellungen selbst auf einschlägigen Informationsveranstaltungen zu korrigieren und die Einführung kontraproduktiver und unfairer Verfahren zu verhindern oder rückgängig zu machen. Die Ziele von Auswahlverfahren werden im Folgenden kurz beschrieben und reflektiert. Weiters wird auf die Anforderungen an Auswahlverfahren und deren Gütekriterien eingegangen.

2.1. Ziele von Auswahlverfahren

Nach Kubinger (2006, S. 346) soll die „Eignungsbegutachtung [...] zu Erkenntnissen führen bezüglich der Fähigkeit und der Motivation des Betroffenen zur Bewältigung jeweils bestimmter beruflicher Anforderungen...“. Im Kontext von (aus-)bildungsbezogenen Eignungsbeurteilungen im Rahmen des Hochschulzugangs, ist nach Spiel

et al. (2007) berufliche Anforderung nicht mit studienbezogener Anforderung gleichzusetzen. Das Studium ermöglicht eine Qualifikation nicht nur in einem bestimmten Beruf, sondern in verschiedenen Berufsfeldern in denen unterschiedliche Kompetenzen benötigt werden. Ziel des Verfahrens soll demnach die Feststellung der Eignung für das Studium bzw. die Studierfähigkeit sein, sowie die Prognose für den Studienerfolg. Eignungsdiagnostische Auswahlverfahren helfen Entscheidungen vorzubereiten, zu unterstützen, zu treffen und zu optimieren (Brambring, 1983). Allerdings wird es nicht einfach sein die Eignung für ein Studium festzustellen. Nach Trost et al. (1998, S. 12) ist die „Studieneignung [...] prinzipiell als Bündel von kognitiven Fähigkeiten und Fertigkeiten, psychomotorischen Fähigkeiten und Fertigkeiten, Kenntnissen, studien- und berufsbezogenen Interessen und Motiven sowie weiteren Persönlichkeitsmerkmalen zu sehen.“ Dementsprechend ist es unrealistisch und unkritisch zu glauben, dass die Eignung für ein Studium mit einem einzigen Test (d.h. einmaliges Verhalten) festzustellen sei (Kubinger, 2006; Schuler, 1991). Zusätzlich soll mit dem Verfahren nicht nur die aktuelle Eignung erfasst, sondern auch eine Prognose über eine voraussichtliche Studienzeit von 4-5 Jahren gegeben werden. Der Einfluss von Interessen, Motiven und Persönlichkeitsmerkmalen legt nahe, dass nicht nur leistungsspezifische Fähigkeiten und Fertigkeiten zu messen nötig seien.

Die prognostische Gültigkeit (siehe 2.2.2) solch einer IST-Standerhebung kann nur mit einer bestimmten Wahrscheinlichkeit gegeben werden. Systematische Veränderungen durch Entwicklung, hinzugekommenes Wissen durch die Lehrinhalte selbst und Unterstützung durch andere Studierende (Bsp. das Projekt CBM¹ – Cascaded Blended Mentoring der Universität Wien) werden nicht berücksichtigt.

Wie eingangs erwähnt, sind die Ressourcen üblicherweise beschränkt, so dass die Finanzierung einer (Aus-)bildung an den Universitäten nur für eine bestimmte Anzahl von Personen vom Staat ermöglicht werden kann. Dadurch steht die Auswahl von Personen unter dem Primat der Ökonomie. Hohe Bildung der Bevölkerung soll die Wettbewerbschancen des Staates in Zukunft sichern. Neben einer Fülle von Ungleichheitsdimensionen (wie bspw. sozialer Hintergrund, regionale Herkunft, Migrationshintergrund, Geschlecht, Einkommen und Beruf der Eltern) die auf Bildungspartizipation und Bildungserfolg einen Einfluss haben (vgl. Bacher, 2005) stellt ein Auswahlverfahren für den Hochschulzugang eine zusätzliche Hürde dar. Bis zum Beginn eines Studiums konnten die Ungleichheitsdimensionen vielfältig ihre Wirkung zeigen. Die soziale Herkunft bestimmt auch die Rahmenbedingungen während des Studiums, weil sie die sozioökonomische Situation der Eltern, die finanzielle Absicherung sowie den

¹Projekthomepage: <http://www.univie.ac.at/cbm.psychologie/>

Lebensstandard während des Studiums beeinflusst (Unger & Wroblewski, 2006).

Auch wenn primär die Prognose des Studienerfolgs das Ziel des Auswahlverfahrens sein soll, so sind weitere direkte und indirekte Ziele damit verbunden (Spiel et al., 2007). Hierzu zählen die Anwerbung leistungsfähiger und motivierter Studierender, geringe Abbruchquoten und kurze Studienzeiten (Heine, Briedis, Didi, Haase & Trost, 2006). Angestrebt werden auch ein Orientierungseffekt und die Förderung zur Selbstselektion der Studienplatzbewerber und Studienplatzberwerberinnen (Spiel et al., 2007). Weiters soll durch Auswahlverfahren die Passung zwischen personenbezogenen Voraussetzungen und den allgemeinen und studienfachspezifischen Anforderungen erhöht werden und die Frustration vieler Studierender durch einen Studienabbruch oder einen Studienwechsel verringert werden (Spiel et al., 2008). Hinsichtlich dem Wunsch finanzielle Einsparungen durch geringere Studierendenzahlen zu erreichen, wird von der Einführung von Auswahlverfahren erwartet, dass sich die Zahl der Studierenden, die sich mangels besserer Alternativen für ein Studium entscheiden, verringert (Spiel et al., 2007). Diese Zahl sei nicht zu unterschätzen. Nach einer Umfrage von Kellermann (1991) an fünf österreichischen Universitäten gab ein Drittel der Erstimmatrikulierten an, das kürzlich angefangene Studium für andere attraktive Berufsangebote wieder abzubrechen.

2.2. Anforderungen an Auswahlverfahren

Folgend werden Anforderungen, die an Auswahlverfahren gestellt werden, beschrieben und diskutiert. Zuerst folgt eine Darstellung unter allgemeiner, bildungswissenschaftlicher und psychologischer Sicht. Anschließend werden die diagnostischen Gütekriterien aufgrund ihrer Bedeutung in der eignungsdiagnostischen Beurteilung in einem eigenen Abschnitt diskutiert.

2.2.1. Anforderungen unter allgemeiner, bildungswissenschaftlicher und psychologischer Sicht

Die Anforderungen die an Auswahlverfahren an Universitäten gestellt werden sind vielfältig. Sie gehen über (eignungs-)diagnostische Gütekriterien hinaus und verfolgen teilweise konkurrierende Ziele wie beispielsweise maximaler Nutzen versus Finanzierung und führen so zu Kompromisslösungen. Spiel und Kolleginnen (2007) betonen neben den bildungswissenschaftlichen und psychologischen Aspekten auch die gesamtgesellschaftliche Relevanz von Auswahlverfahren. Als wichtigste Kriterien, neben den gängigen diagnostischen Gütekriterien, an denen sich Auswahlverfahren beim Hoch-

schulzugang messen lassen müssen nennt Trost (2005, S. 138) auszugsweise:

- die Vereinbarkeit der Verfahren mit den Vorgaben des Hochschulrahmengesetzes und der darauf beruhenden Ländergesetze
- die Transparenz
- die Fairness gegenüber definierten Gruppen von Bewerberinnen und Bewerbern
- die geringe Trainierbarkeit
- die Akzeptanz bei den betroffenen Gruppen und in der Öffentlichkeit
- die 'Gerichtsfestigkeit'

Diese Kriterien decken sich auch mit Spiel et al. (2007), die zusätzlich zur Akzeptanz und Informiertheit, Transparenz und Fairness die Kriterien Commitment (Bindung), Beratung, Self-Assessment sowie den Zeitpunkt des Verfahrens nennen. Speziell auf die Notwendigkeit eines guten Anforderungsprofils wird eingegangen (siehe unten). Im Folgenden wird zur Illustration ihrer Bedeutung auf einige Kriterien nach Spiel et al. (2007) detaillierter eingegangen.

Die *Akzeptanz*, das Annehmen des Tests durch die Testperson, zumeist mit der Erwartung eines subjektiven Nutzens verbunden, wurde vom Testkuratorium der Föderation deutscher Psychologenverbände (1986) als diagnostisches Gütekriterium beschrieben. Geringe Akzeptanz kann zu einer falschen Vorselektion führen, so dass grundsätzlich geeignete Bewerberinnen und Bewerber das Auswahlverfahren ablehnen und nicht teilnehmen. Positiv werden von den Studienbewerberinnen und Studienbewerbern Studierfähigkeitstests, strukturierte anforderungsbezogene Auswahlgespräche und Noten bewertet. Die geringste Zustimmung fanden Losverfahren und Graphologien. Es ist darauf hinzuweisen, dass die Akzeptanz des Auswahlverfahrens letztlich auch das Image der Universität bzw. der Fakultät beeinflusst. Strategien zur Förderung der Akzeptanz umfassen nach Kubinger (2006) ausreichende Erläuterungen über Zweck der Testung, Testdauer, Aufgabenstellungen, Interpretation und vor allem über die mit dem Test verbundenen Konsequenzen; sprich Erhöhung der Transparenz.

Transparenz und *Informiertheit* stellen einen wechselseitigen Bezug einerseits durch die institutionelle Verantwortung auf Seiten der Universitäten und andererseits individuelle Verantwortung auf Seiten der Studienplatzbewerberinnen und Studienplatzbewerber her. Zu den Aufgaben der Universitäten bzw. der Fakultäten zählt Informationen über Inhalte und Organisation der Studienfächer sowie über Auswahlverfahren bereitzustellen und diese leicht zugänglich zu machen. Für den Studienplatzbewerber/die

Studienplatzbewerberin ist es notwendig, die Informationsbeschaffung und -nutzung als einen eigenverantwortlichen Prozess wahrzunehmen. Eine umfassende Informiertheit über zentrale Aspekte führen zu einer adäquaten und stabilen Studienfachwahl und vermindern Studienabbruch und Studienfachwechsel. Ein in mehreren Stufen gestalteter Auswahlprozess mit Vorschaltung einer Informationsphase werden als besonders wichtig erachtet.

Ein von den Universitäten erwünschtes Ergebnis, das mit der Durchführung von Auswahlverfahren erreicht werden soll, ist die Erhöhung des *Commitments* der Studierenden. Gestützt wird diese Annahme durch die Theorie der Rechtfertigung des Aufwands (effort justification) nach Aronson und Mills (1959) welche besagt: „Je größer unter sonst gleichen Bedingungen die Anstrengung ist, die man um einer Sache willen auf sich nimmt, desto größer ist die Wertschätzung, die man dieser Sache entgegenbringt.“ (Herkner, 2001, S. 37). Da mit der Vorbereitung auf und der Teilnahme am Auswahlverfahren relativ viel Anstrengung verbunden ist, kann eine Steigerung der Wertschätzung des Studiums bzw. der Universität angenommen werden, die günstige Wirkung auf Engagement, Studienerfolg, -fortgang und -abschluss hat (Spiel et al., 2008).

Als weiteres Kriterium, dem Spiel et al. (2007) aus psychologischer und bildungswissenschaftlicher Perspektive eine hohe Bedeutung zumessen, ist der *Zeitpunkt des Auswahlverfahrens*. Hier geht es primär um die Frage ob eine Auswahl bereits vor der Zulassung zum Studium stattfinden soll oder erst nach einer zeitlich festgelegten Studieneingangsphase. Eine Selektion nach einer Studieneingangsphase bringt den Vorteil, dass die Eignungsbeurteilung auf bereits tatsächlichen Studienleistungen beruht, die auch durch eine Vielzahl an nicht fähigkeitsbezogener Variablen gebunden ist und dadurch die Prognosegültigkeit des Studienerfolgs erhöht wird (valide Ergebniskriterien siehe unten). Zu Bedenken ist jedoch, dass im Falle einer hohen Selektionsquote bereits viele Studierende Arbeits- und Lernleistung sowie Lebenszeit in ein bis zwei Semester Studium investiert haben. Hierbei wäre die Zumutbarkeit dieses Auswahlverfahrens zweifelhaft. Eine hohe Studierendenzahl in der Studieneingangsphase ist zusätzlich mit hohen Kosten verbunden. Die konkrete Abschätzung der Effizienz einer Studieneingangsphase (Nutzen in Relation zu den Kosten) steht noch aus (Spiel et al., 2007).

Auswahlverfahren sollten grundsätzlich auf wissenschaftlich fundierten Methoden basieren (Brambring, 1983) und einem zielgerichteten, strategischen Vorgehen zugrunde liegen (Spiel et al., 2007). Hierzu gehören im diagnostischen Prozess die sorgfältige Erstellung der entsprechenden (Anforderungs-)Dimensionen als auch vorab die Festlegung der notwendigen Ausprägungen (vgl. Ortner, 2003). Diese sollten im Anforderungsprofil

klar festgelegt sein. Dieses Vorgehen entspricht den Grundsätzen berufsbezogener Eignungsbeurteilung der DIN 33430 'Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen' (Deutsches Institut für Normung e.V., 2002) bzw. ÖNORM D 4000 'Anforderungen an Prozesse und Methoden in der Personalauswahl und -entwicklung' (Österreichisches Normungsinstitut, 2005).

Nach Spiel et al. (2007) sollten erstens keine Kompetenzen erfasst werden, die im Studium erst erworben werden und andererseits nur diejenigen Kompetenzen erfasst werden, die für eine erfolgreiche Bewältigung des Studiums zentral sind; zweitens für die Prognose des Studienerfolgs und nicht des Berufserfolgs aufgrund unterschiedlicher Anforderungen herangezogen werden. Die Ausbildung in den meisten Studienfächern dient auch nicht speziell der Ausübung eines bestimmten Berufs, sondern ermöglicht Tätigkeiten in unterschiedlichen Berufsfeldern. Aus den Anforderungen an das Studium können die fachspezifischen Fähigkeiten und Fertigkeiten abgeleitet werden, zusätzlich auch fächerübergreifende Grundqualifikationen die nicht fachspezifisch für die Bewältigung eines Studiums erforderlich sind (Konegen-Grenier, 2002). Spiel et al. (2007, S. 481) nennen hierzu im kognitiven Bereich analytische Fähigkeit, Abstraktionsfähigkeit, Differenzierungsvermögen, Synthese- und Transferfähigkeit und sprachliche Ausdrucksfähigkeit und bezüglich persönlicher Eigenschaften und Motive inhaltliches Interesse, Leistungsmotivation, Genauigkeit, Zielstrebigkeit und Beharrlichkeit, Eigeninitiative und Selbstorganisation, hohes Anspruchsniveau und Fähigkeit zur Selbstreflexion.

Für eine Fachrichtung müsste es grundsätzlich möglich sein, und dies wäre auch notwendig anzustreben, sich auf gemeinsame Kriterien zu einigen, die an allen Universitäten gleichermaßen getestet werden. Jäger (2005) schlägt hierfür die Etablierung eines „Fundamentums“ und eines „Additums“ vor, welche grundlegende und fachspezifische Elemente beinhalten. Die Transparenz des Fundamentums und des Additums müssen in höchstem Maße gewährleistet sein und Studienbewerber bereits bei Interessensbekundung (z.B. durch Homepage oder gedruckte spezielle Infos für Studienbewerberinnen und Studienbewerber) über Kriterien des Fachs und der Universität bei der Studierendenauswahl informieren (Schuler & Stehle, 1985 zitiert nach Jäger, 2005, S. 145). Bei Ablehnung einer Bewerberin/eines Bewerbers sollte eine (computergestützte) Rückmeldung über Kompetenz und Leistung erfolgen (Jäger, 2005). Ob sich die Kriterien für die Studierendenauswahl auch bewähren, kann am Studienerfolg der Ausgewählten gemessen werden. Die Operationalisierung des Studienerfolgs kann anhand unterschiedlicher Kriterien erfolgen (vgl. Hossiep, 1995; Schmidt-Atzert, 2004) und muss dementsprechend spezifiziert werden. Hierfür hat sich eine Differenzierung zwischen Ergebnis- und Prozesskriterien bewährt (vgl. Schüpbach & Klauer, 2005). Va-

lide Ergebniskriterien sind Prüfungsleistungen als Noten in Einzel-, Zwischen- und Abschlussprüfungen, zu den Prozesskriterien gehören die Studiendauer, Studienabbrüche und Studienfachwechsel (Spiel et al., 2008). Zur notwendigen Klärung der Wirkung (Effektivität) und der Kosten-Nutzen-Relation (Effizienz) aller Teile des Auswahlverfahrens sollten umfassende und regelmäßige Evaluationen von erfahrenen Fachleuten durchgeführt und deren Ergebnisse auch zur Modifikation und Optimierung herangezogen werden.

Zusammenfassend zeigt sich, dass aus allgemeiner, bildungswissenschaftlicher und psychologischer Sicht wichtige Bedingungen bzw. Kriterien beim Einsatz von Auswahlverfahren abgeleitet werden können, um einer validen Eignungsbeurteilung in höherem Maße gerecht zu werden: (1) Anwendung wissenschaftlich fundierter und ausreichend erprobter Verfahren, (2) hohe Transparenz des gesamten Auswahlverfahrens, (3) Maßnahmen zur Förderung der Akzeptanz, (4) ein in Stufen geführter Auswahlprozess, der auch das Kriterium der Informiertheit berücksichtigt, (5) keine Erfassung von Kompetenzen, die erst im Studium erworben werden und nur solche die für den Studienerfolg zentral sind, (6) ein im spezifischen Fach einheitliches Konzept über Eignungskriterien und Testung und (7) umfassende und regelmäßige Evaluationen zur Effektivität und Effizienz aller Teile und Stufen des Auswahlverfahrens.

Wie bereits erwähnt sollen Auswahlverfahren wissenschaftlich fundiert sein. Im Folgenden wird deshalb auf die Bedeutung der (eignungs-)diagnostischen Gütekriterien – davon ausführlicher auf die prognostische Validität – im Rahmen von Auswahlverfahren eingegangen.

2.2.2. Bedeutung eignungsdiagnostischer Gütekriterien

Dieser Abschnitt befasst sich mit den Gütekriterien psychologisch-diagnostischer Verfahren, die auch von der DIN 33430 für den Einsatz bei berufsbezogenen Eignungsbeurteilungen getragen werden. Auch das *Testkuratorium der Föderation Deutscher Psychologenvereinigungen* legte Gütekriterien fest (Testkuratorium, der Föderation deutscher Psychologenv Verbände, 1986). Die Beschreibungen der Gütekriterien beziehen sich im wesentlichen auf Lienert und Raatz (1998) sowie auf Kubinger (2006), die frei wiedergegeben werden.

Objektivität bezeichnet den Grad der Unabhängigkeit der Ergebnisse von der Untersuchungssituation und der untersuchenden Person (Durchführungsobjektivität und Testleiterunabhängigkeit), der auswertenden Person (Auswertungssicherheit) und der „Interpreten“ (Interpretationseindeutigkeit).

Die *Reliabilität* beschreibt den Grad der Genauigkeit (formale Messgenauigkeit) mit

der ein bestimmtes Merkmal gemessen wird, unabhängig davon, ob das Verfahren auch den Anspruch erhebt dieses Merkmal zu messen. Die Reliabilität ist eine Voraussetzung für die Reproduzierbarkeit (Replizierbarkeit) von Ergebnissen unter gleichen Bedingungen. Speziell die wiederholte Messung (Retest) kann über einen längeren Zeitraum Informationen über das Verfahren als auch über das gemessene Merkmal geben und als „Stabilität“ des Merkmals aufgefasst werden.

Unter *Validität* versteht man das Ausmaß der Gültigkeit mit der das zu messen gewünschte Merkmal auch wirklich erfasst wird. Während eine hohe Reliabilität eine notwendige aber nicht hinreichende Bedingung für ein Verfahren ist, stellt die Validität, unter dem Anspruch die „Welt“ nicht nur zu beschreiben sondern auch kausal zu verstehen, das wohl wichtigste Gütekriterium dar, das gleichzeitig auch das am schwersten zu messende ist. Folgende Konzepte bzw. Begriffe der Validität werden unterschieden: (1) von inhaltlicher Gültigkeit spricht man, wenn der Test/das Verfahren selbst (definitionsgemäß) das optimale Kriterium darstellt, (2) von Konstruktvalidität spricht man, wenn auf ein sog. „Konstrukt“ Bezug genommen wird und (3) Kriteriumsvalidität, wenn auf ein sog. „Außenkriterium“ Bezug genommen wird. Der Vorteil der Kriteriumsvalidität ist, dass sich mit ihr eine statistische Kennzahl und somit das Ausmaß der Validität eines Verfahrens bestimmen lässt.

Einschub Prognostische Validität

Für Auswahlverfahren ist speziell die *Vorhersagegültigkeit (prognostische Validität)*, die sich zur Kriteriumsvalidität zählen lässt, von Bedeutung. Diese bestimmt sich aus der Korrelation des Testwerts aus dem Auswahlverfahren mit dem Außenkriterium, welches zu einem späteren (ferneren) Zeitpunkt gemessen wird (z.B. Prüfungserfolg, Studienabbruchsquote). Eine effektive Auswahl maximiert die Zahl der Geeigneten. Um dies zu ermöglichen ist eine hohe Basis- bzw. Grundrate (= Anzahl geeigneter Bewerberinnen und Bewerber), eine niedrige Selektionsrate (= Anzahl nach dem Verfahren ausgewählte Personen) und eine hohe Validität des Verfahrens notwendig (Spiel et al., 2007, 2008; Stemmler, 2005). Nach Analysen von Kolland und Morgeditsch (2007) wurden an den beiden Medizinischen Universitäten (Wien und Innsbruck) im Studienjahr 2006/07 von 5611 Bewerberinnen und Bewerber bei der Anmeldung letztlich 1190 Bewerberinnen und Bewerber zugelassen. Dies entspricht einer Selektionsquote von rund 21 Prozent. Will man maximal 20 Prozent Ungeeigneter nach dem Auswahlverfahren unter den Selektierten haben, so müsste das Verfahren bei 21 Prozent Selektionsrate eine prognostische Validität von mindestens $r = .40$ aufweisen (Berechnung siehe Anhang A.1). Spiel et al. (2007) berichten, dass in der Literatur schon oft ab $r = .30$ von

einer höheren Validität gesprochen wird. Ein einzelnes Verfahren erreicht nach Schuler (1996, S. 172) nur eine Validität bis $r = .50$, bei kombinierten Verfahren scheint die Grenze bei $r = .70$ zu liegen. Ein Ziel von Auswahlverfahren im Hochschulstudium ist die Studienabbruchsquote zu senken (siehe Abschnitt 2.1). Nach einer Studie von Dell'mour & Landler (2002), anhand von Daten aus den Jahren 1975-1986 und 1980-1998, liegt in Österreich die Abbruchsquote bei inländischen Studierenden (Beginn Wintersemester) im Fach Medizin bei durchschnittlich 32 Prozent. Wollte man diese halbieren, so müsste das Auswahlverfahren eine prognostische Validität von fast $r = .50$ erreichen.

Tabelle 2.1.: Treffer- und Fehlklassifikation in der Eignungsdiagnostik (nach Spiel, Litzenberger & Haiden, 2007, S. 487)

		Erfolgsbeurteilung (Kriterium)	
		erfolgreich	nicht erfolgreich
Eignungsbeurteilung (Prädiktor)	geeignet	richtig positiv Zu dieser Gruppe zählen jene BewerberInnen, die sowohl als „geeignet“ beurteilt werden als auch das Erfolgskriterium erreichen.	falsch positiv Zu dieser Gruppe gehören KandidatInnen mit positiver Eignungsbeurteilung, die das Erfolgskriterium jedoch nicht erreichen.
	nicht geeignet	falsch negativ Zu dieser Gruppe gehören KandidatInnen mit negativer Eignungsbeurteilung, die jedoch das Erfolgskriterium erreichen.	richtig negativ Zu dieser Gruppe zählen BewerberInnen, die sowohl als „nicht geeignet“ beurteilt werden als auch das Erfolgskriterium nicht erreichen.

Durch eine Klassifikation der Bewerberinnen und Bewerber in 'geeignet' und 'nicht geeignet' kann die Trefferquote quantifiziert werden. Zur einfachen Ermittlung wird ein dichotomes Erfolgskriterium festgelegt (z.B. 'erfolgreich' vs. 'nicht erfolgreich'). Mittels des Vier-Felder-Schemas lässt sich eine Beurteilung der Qualität eignungsdiagnostischer Auswahlentscheidungen durch die beiden dichotomen Variablen Eignungsbeurteilung (Prädiktor) und Erfolgsbeurteilung (Kriterium) abschätzen (Kersting, 2003). Aus dem Vier-Felder-Schema kann die Zuordnung zu richtig und falsch positive und richtig und falsch negative Entscheidungen abgelesen werden (siehe Tabelle 2.1). Praktisch stellt sich das große Problem, dass bei der Überprüfung der Güte von Auswahlverfahren nur zensierte Datensätze vorliegen und falsch negativ beurteilte Kandidatinnen und Kandidaten nicht berücksichtigt werden können (vgl. Kubinger, 2006). Es kann somit keine Aussage über diejenigen Personen getroffen werden, die im Auswahlverfahren ausgeschieden wurden aber doch das Erfolgskriterium erreicht hätten.

Tabelle 2.2.: Überblick über die prognostische Validität verschiedener Auswahlinstrumente für Studiennoten (nach Hell, Trapmann, Weigand, Hirn & Schuler, 2005).

	k	N total	r
mittlere Schulnoten	53	48178	.46
allgemeine Studierfähigkeitstests	10	745	.43
fachspezifische Studierfähigkeitstests	37	45783	.43
Bachelornoten	9	980	.32
Essays	12	694	.29
Sprachtests	20	1333	.28
Interessenfragebogen	11	1645	.18
Interviews	44	4305	.13
Arbeitsproben	6	1987	.13

Anm: k = Anzahl der unabhängigen Stichproben

N total = (Gesamt)Anzahl der untersuchten Personen

r = Validitätskoeffizient

In einer Metaanalyse basierend auf 211 Artikeln haben Hell, Trapmann, Weigand, Hirn und Schuler (2005) die prognostische Validität einiger Auswahlinstrumente für den Hochschulzugang hinsichtlich des Erfolgskriteriums Studiennoten untersucht. Die Ergebnisse zeigen (siehe Tabelle 2.2), dass Schulnoten sowie allgemeine- und fachspezifische Studierfähigkeitstests die höchste prognostische Validität aufweisen. Für die Anforderung maximal 20 Prozent Ungeeignete auszuwählen, verfehlen die anderen Auswahlinstrumente das Kriterium deutlich (siehe Anhang Tabelle A.2).

Einschub Prognostische Validität – Ende

Neben den bereits besprochenen (klassischen) diagnostischen Gütekriterien sollen die weiteren nur überblicksartig und hinsichtlich ihres Nutzens für Auswahlverfahren beschrieben werden.

Das Gütekriterium der *Ökonomie* eines Verfahrens wird erreicht, wenn gemessen am diagnostischen Informationsgewinn, relativ wenig Ressourcen (bspw. Zeit, Geld, Personal) beansprucht werden. Die Ökonomie bestimmt sich im Sinne einer Wirtschaftlichkeit und einer Aufwandsminimierung. Der Aufwand der Universitäten bei der Durchführung von Auswahlverfahren ist aufgrund der großen Anzahl an Bewerberinnen und Bewerbern nur mit viel Personal, großen und oft extra angemieteten Räumen, Zeit (Vorbereitung, Durchführung und Auswertung), Verbrauchsmaterialien (Testbögen, evtl. Computer) und Gebühr für die Testnutzung nicht zu vernachlässigen.

Die *Nützlichkeit* eines Verfahrens wird dann erfüllt, wenn für das gemessene Merkmal praktische Relevanz besteht und die auf Verfahrensgrundlage getroffenen Entscheidun-

gen mehr Nutzen als Schaden erwarten lassen. Ob ein Verfahren einen 'Nutzen' d.h. einen Gewinn bringt, muss anhand von vordefinierten Kriterien, einer so genannten Nutzenfunktion, gemessen werden. Hierzu zählen auch andere Gütekriterien wie beispielsweise schon erwähnt die prognostische Validität.

Eng verbunden mit dem zu erwartenden 'Schaden' durch das Verfahren ist das Gütekriterium der *Zumutbarkeit*. Es wird erfüllt, wenn die Testperson in zeitlicher, psychischer und körperlicher Hinsicht, absolut oder relativ zu dem aus der Anwendung resultierenden Nutzen, geschont wird. Hinsichtlich der Auswahlverfahren an Hochschulen kann für die Zielgruppe, primär junge Erwachsene, ein Leistungstest auch über mehrere Stunden als zumutbar angesehen werden. Ein Augenmerk ist hier auf ausreichende Pausengestaltung und Arbeitsbedingungen zu legen.

Eine weitere Forderung an Auswahlverfahren ist, dass das ('Test-')Ergebnis durch die zu 'testende' Person nicht oder nur unwesentlich nach eigenem Belieben beeinflusst werden kann. Dies fordert das Gütekriterium der *Unverfälschbarkeit*, welches folglich auch die Trainierbarkeit des Verfahrens einschließt.

Ein sehr wichtiges Gütekriterium bei Auswahlverfahren ist die *Fairness*. Diese wird erfüllt, wenn es zu keiner systematischen Diskriminierung bestimmter Personen (bspw. bezüglich ihrer ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppenzugehörigkeit) führt. Ob ein Verfahren fair ist, ist nur unter Berücksichtigung der Messintention zu sehen. Als unfair ist ein Verfahren demnach anzusehen, wenn die gemessenen Unterschiede nichts mit der zu messen beabsichtigten Eigenschaft zu tun haben. Die oft verbreitete Vorstellung, dass Auswahlverfahren dann fair sind, wenn bestimmte Gruppen von Bewerberinnen und Bewerbern entsprechend ihrer Verteilung in der Grundgesamtheit ausgewählt wurden, vernachlässigt, dass zwischen den Gruppen tatsächliche Unterschiede bestehen können. Ein solches 'Quotenmodell der Fairness' wäre nicht im Sinne der Verfahren, die Personen mit der höchsten Wahrscheinlichkeit im Studium erfolgreich zu sein, auswählen. Ein in der Wissenschaft akzeptiertes Modell entspricht dem 'Modell gleicher Erfolgswahrscheinlichkeiten' (Spiel et al., 2007). Nach diesem Modell ist ein Auswahlverfahren als fair zu bezeichnen, wenn der Studienerfolg für keine Gruppe aufgrund des Ergebnisses im Verfahren systematisch über- oder unterschätzt wird. Neben dieser Definition von Fairness sind auch Aspekte zu erwähnen, die die Fairness beeinflussen. Hierzu ist beispielsweise die Testerfahrung einer Person zu zählen, die sich positiv auf das Ergebnis auswirken kann. Es geht weniger um itemrelevante Fähigkeiten im Sinne der Trainierbarkeit, sondern mehr um die grundsätzliche Gewöhnung und um die Vertrautheit mit dem Verfahren (z.B. durch mehrmalige Teilnahme).

Weiter seien noch die beiden diagnostischen Gütekriterien der Normierung und der

Skalierung erwähnt, die hinsichtlich probabilistischer testtheoretischer Ansätze auch für Auswahlverfahren bedeutend sind. Die *Normierung* wird erfüllt, wenn die individuellen (Test-)Ergebnisse in Relation zu einem Bezugssystem (sog. Normen) stehen. Diese Normen müssen an einer repräsentativen Stichprobe einer zuvor definierten Population gewonnen worden sein. Ebenso dürfen die Normen nicht veraltet sein. Zur einfachen Vergleichbarkeit werden oft Prozentränge oder Standardwerte (Mittelwert: 100, Standardabweichung: 10) herangezogen.

Bei der *Skalierung* geht es um die Angemessenheit der Verrechnungsvorschriften der einzelnen (Unter-)Testleistungen einer Testperson. Angemessenheit meint hierbei, dass im Sinne der Messintention sicher gestellt wird, dass die verrechneten Einzelwerte empirische Sachverhalte ausdrücken. Beispielsweise die Messintention wäre *ein* bestimmtes Konstrukt zu erfassen, dann dürfen nur Items verrechnet werden, die auch tatsächlich dieses Konstrukt eindimensional erfassen. Anders als die Klassische Testtheorie kann die Probabilistische Testtheorie diese Forderung prüfen. Eine besondere Bedeutung kommt hierbei dem sogenannten Rasch-Modell zu. Die Eindimensionalität ist gewährleistet, wenn das Rasch-Modell gilt (Fischer, 1995). Die Aufgaben sprechen bei allen Personen, die diese bearbeiten, die selbe Eigenschaft an. Hängt der Testwert von mehreren Eigenschaften ab, so wäre dieser nicht mehr (eindeutig) interpretierbar. Durch die Eindimensionalität kann eine klare Einbettung des Konstrukts in Modellen der Vorhersage ermöglicht und so die Wahrscheinlichkeit einer besseren Vorhersage (im Sinne diagnostischer Gütekriterien) erhöht werden. Überdies hinaus ergeben sich durch die Gültigkeit des Rasch-Modells weitere Vorteile. Die Modellgültigkeit ermöglicht „Spezifische Objektivität der Vergleiche“ (Kubinger, 2003, S. 417). Dies bedeutet einerseits, dass der Unterschied in der Fähigkeit zwischen zwei Personen unabhängig von den bearbeiteten Items des Itempools bestimmt werden kann, andererseits, dass die Itemschwierigkeit zweier Items unabhängig von der dafür verwendeten Stichprobe verglichen werden kann (vgl. Kubinger, 1989). Weiter kann die Messgenauigkeit des Testergebnisses einer Person unabhängig von anderen noch getesteten Personen bestimmt werden (vgl. Kubinger, 2006).

Im diesem Kapitel wurden die Gütekriterien zur eignungsdiagnostischen Beurteilung beschrieben. Ihre entscheidende Bedeutung erlangen die Gütekriterien dadurch, indem sie die Maßstäbe zur Bewertung und Sicherung der Qualität eignungsdiagnostischer Verfahren darstellen. Die Gütekriterien können auch zum Vergleich der Qualität von unterschiedlichen Auswahlverfahren eingesetzt werden. In 33 europäischen Ländern gibt es eine Zulassung zum Universitätsstudium nach Eignung und Leistung (Hänsgen & Spicher, 2004). Im folgenden Abschnitt soll nun speziell und mit Blick auf die hier vorliegende Evaluationsstudie das Auswahlverfahren, wie es in Österreich seit Winter-

semester 2006 zum Einsatz kommt, näher beschrieben werden. Im Anschluss folgen empirische Befunde zum zentralen Teil des Auswahlverfahrens, dem EMS-Test.

2.3. Auswahlverfahren an den Medizinischen Universitäten Wien und Innsbruck

Seit dem Studienjahr 2006/07 führen die medizinischen Universitäten Wien und Innsbruck gemeinsam eine kapazitätsorientierte Studienplatzvergabe für alle Studienbewerberinnen und Studienbewerber durch. Die Verordnung über die Zulassungsbeschränkungen wurde in den Mitteilungsblättern² der beiden Medizinuniversitäten festgelegt. Die Studienplatzvergabe erfolgt innerhalb der zugeteilten Quote anhand der Ergebnisse im Eignungstest für das Medizinstudium (EMS). Diesem sind eine elektronische Voranmeldung und eine persönliche Anmeldung verpflichtend vorgeschaltet. Konkret besteht das Aufnahmeverfahren aus folgenden Schritten:

- eine Voranmeldung (nach festgelegter Frist) mit einer verbindlichen Angabe für den gewünschten Studienort an dem die Zulassung erfolgen soll
- eine persönliche Anmeldung (nach Einladung) für den Test, bei welcher den Bewerberinnen und Bewerbern eine umfassende Informationsbroschüre über die Zulassungsvoraussetzungen, den Eignungstest und den Testablauf ausgehändigt wird
- die Teilnahme am EMS (dieser wird gleichzeitig in Wien und Innsbruck durchgeführt)
- Erklärung der Inanspruchnahme des Studienplatzes durch die Bewerberinnen und Bewerber innerhalb von 20 Tagen nach Ergebnisveröffentlichung.

Nach Auswertung der Testergebnisse durch das Département de Psychologie Fribourg (Schweiz) wird eine gemeinsame Rangliste der beiden Universitäten über den Gesamtestwert der Studienbewerberinnen und Studienbewerber getrennt für Zahn- und Humanmedizin und Quote gebildet. Studienbewerberinnen und Studienbewerber, die keine Zulassung zum Studium erhalten haben, können sich in den folgenden Studienjahren neuerlich beteiligen. Die Beteiligung an diesem Aufnahmeverfahren ist unbegrenzt

²Medizinische Universität Wien: http://www.meduniwien.ac.at/files/6/3/6_mb_20_1_06.pdf
[07.04.2010]
Medizinische Universität Innsbruck: <http://www.i-med.ac.at/mitteilungsblatt/2006/12.pdf>
[07.04.2010]

möglich. Die Medizinuniversitäten sind bemüht die Studienbewerberinnen und Studienbewerber über die Durchführung des Auswahlprozesses mittels Webseiten und Informationsmaterialien zu informieren. Inwieweit diese Informationen zu einer höheren Informiertheit der Studierenden führen und so zu einer stabilen Entscheidung für ein Studienfach (geringe Drop-out-Rate) beitragen ist noch ungeklärt. Die Transparenz des Auswahlprozesses ist hinsichtlich Ablauf der Schritte, Durchführung der Testung, Art der Fragestellungen, Auswertung der Testergebnisse, Quotenzuteilung und Studienplatzvergabe gegeben. Die Testitems selbst sind streng geheim und entziehen sich dadurch jeglicher inhaltlichen Evaluation.

Der Kern des Auswahlprozesses an den Medizinischen Universitäten Wien und Innsbruck beruht auf dem EMS. Im folgenden Abschnitt wird zuerst eine Beschreibung des Testverfahrens gegeben und anschließend seine Qualität diskutiert.

2.4. Der Eignungstest für das Medizinstudium (EMS)

Der EMS beruht auf dem deutschen Test für Medizinische Studiengänge (TMS), der über viele Jahre im Einsatz war und umfassend analysiert wurde (vgl. Trost et al., 1998; Hohner, 1997; Trost, 1994; Klieme, 1991). Bis zum Jahr 1997 gab es in der Schweiz wie auch in Österreich einen freien Hochschulzugang. Ab 1998 bis 2003 kam der EMS in der gleichen Teststruktur wie der TMS zum Einsatz. Analog zum TMS wurden die Testitems jedes Jahr (bei äquivalentem Test) verändert. Nach Erweiterung durch einen Konzentrationstest 2004 kam es 2005 zu einer Revision des Tests und ein neuer Untertest 'Planen und Organisieren' wurde aufgenommen (Hänsgen & Spicher, 2006).

Aus einer Leistungs- und Bearbeitungsgeschwindigkeits-Komponente bestehend ist der EMS in die Gruppe der Speed-and-power-Tests (vgl. Kubinger, 2006) einzuordnen, der aus 10 Untertests und insgesamt aus 198 Items besteht. Die Vorgabe der Untertests erfolgt in festgelegter Abfolge. Die Durchführung des Tests erfolgt in zwei Halbtagen mit einer Mittagspause. Von den 10 Untertests sind 9 im Single-Choice-Format – d.h. pro Item ist immer nur eine Lösung richtig. Die Testabnahme erfolgt in deutscher Sprache. Für einen Überblick über die Untertests des EMS siehe Anhang A.1.1 (Hänsgen & Spicher, 2007).

Innerhalb jedes Untertests steigt die Aufgabenschwierigkeit an. Für jedes richtig gelöste Item (unabhängig von der Itemschwierigkeit) wird ein Punkt verrechnet. Nicht richtig gelöste Aufgaben werden mit Null Punkten bewertet. Es gibt keine Minuspunkte. Dadurch wird die erschöpfende Statistik des Testscores gewährleistet (vgl. Kubinger, 2003). Aus der Gesamtpunktzahl wird der standardisierte Testwert errechnet.

Der EMS stellt den Kern des Auswahlverfahrens für die Zulassung dar. Dement-

sprechend ist seine eignungsdiagnostische Güte von zentraler Bedeutung für den Auswahlprozess. Als Studierfähigkeitstest mit einem definierten Leistungskriterium und einer eindeutigen Verrechnungsvorschrift kann der EMS als objektives Verfahren angesehen werden. Die Reliabilität des Gesamttests erwies sich mit 0.92 (Spiel et al., 2008, S. 100) als zufriedenstellend. Auf Ebene der Untertests ergab die Messgenauigkeit deutlich geringere Werte und lag im Bereich von 0.56 (Muster zuordnen) bis 0.75 (Schlauchfiguren), was für 20 (respektive 18) Items pro Test als niedrig beurteilt werden kann.

Ein Indikator für den Erfolg des Studiums sind zweifelsfrei die erzielten Prüfungsleistungen. Die Durchschnittsnote der ersten Vorprüfung³ (ein Jahr nach Studienbeginn) korrelierte mit dem Testwert mit $r = .50$ (Hänsgen & Spicher, 2000). Darüber hinaus traten Personen mit hohen Testwerten im EMS früher zur Prüfung an. Nach der zweiten Vorprüfung zeigte sich, dass Personen, die die Prüfung bestanden, im Mittel um rund 4 Punkte im EMS besser waren als diejenigen, die sie nicht bestanden (Hänsgen & Spicher, 2001). Der EMS zeigt auch Zusammenhänge mit Schulleistungen. Die Maturaleistung korreliert mit der Testleistung im EMS $r = .48$ (Hänsgen & Spicher, 2004) was als mittlerer Zusammenhang zu interpretieren ist. Die Autoren betonen, dass die Schulleistungen und der EMS zwar auf einen gemeinsamen Fähigkeitsfaktor zurückzuführen seien, aber diese nicht das Gleiche wären (Hänsgen & Spicher, 2004). Auch in Studien zum TMS zeigten sich diese Korrelationen doch sei die Überlappung beider Kriterien klein, sodass in „hinreichender Weise eigenständige Leistungsanspekte“ erfasst werden würden (Trost et al., 1998, S. 56). Die Fähigkeiten, die der EMS misst können laut Testautoren (vgl. Zentrum für Testentwicklung und Diagnostik, 2007) nur über langjährige Lern- und Entwicklungsprozesse erworben werden.

Wie bereits aus dem TMS bekannt war (vgl. Trost et al., 1998), gibt es auch im EMS systematische Geschlechtsunterschiede. Auch Spiel et al. (2008) konnten zeigen, dass Frauen im EMS deutlich schlechter abschneiden als Männer ($\eta_p^2 = .026^4$, S. 135). Weiter zeigte sich ein noch größerer Unterschied in der Nationalität. Bewerberinnen und Bewerber aus Deutschland schnitten deutlich besser ab als jene aus Österreich ($\eta_p^2 = .037$, S. 135). Neben den Einflussgrößen Geschlecht und Nationalität erwies sich auch die Schulbildung der Eltern (kategorisiert in niedrige, mittlere und hohe Schulbildung) als signifikante Einflussgröße. Bewerberinnen und Bewerber von Eltern mit hoher Schulbildung erzielten eine um durchschnittlich 4.5 Punkte höhere Testleistung als Bewerberinnen und Bewerber von Eltern mit niedriger Schulbildung. Daraus folgt, dass das Testergebnis empirisch durch eine Reihe nicht messintendierter Einflussgrößen

³In Österreich entspricht dies der 'Summativen Integrativen Prüfung 1' (SIP 1).

⁴Maß der Effektstärke.

wie Geschlecht, Nationalität und Schulbildung erklärt werden kann. Die Fairness des Verfahrens könnte verletzt sein.

Zur Prüfung der Fairness des Verrechnungsmodus haben Spiel et al. (2008) in ihrer Evaluationsstudie (Studienjahr 2007/08) testtheoretische Analysen nach dem Rasch-Modell durchgeführt. Ziel der Analysen war es zu prüfen, ob die einzelnen Untertests des EMS in unterschiedlichen Teilstichproben (geteilt nach Testwert, Geschlecht, Nationalität, Schultyp) die selbe Fähigkeit (eindimensional) erfassen oder ob Personen- oder Umgebungsfaktoren auf die Lösung der Aufgaben Einfluss haben. Die Autorinnen konnten zeigen, dass mehrere Aufgaben in den einzelnen Untertests das innere Teilungskriterium, den Testwert, verletzen. In vier von neun Untertests waren mindestens 50 Prozent der Items bezüglich des Testwerts nicht Rasch-Modell-konform. Inhaltlich bedeutet dies, dass die Aufgaben nur unzureichend zwischen 'besseren' und 'schlechteren' Personen differenzieren können. Eine eindeutige Aussage über die Fähigkeit der Person ist anhand dieser Aufgaben nicht möglich. Auch hinsichtlich der Teilungskriterien Geschlecht und Nationalität erwies sich eine Vielzahl von Items als nicht Rasch-Modell-konform. Die 'Verrechnungsfairness' im Sinne eines Gütekriteriums (vgl. Unterabschnitt 2.2.2) ist dadurch verletzt (Kubinger, 2006).

Laut der Testinformationsbroschüre (Zentrum für Testentwicklung und Diagnostik, 2007) für die Studienbewerberinnen und Studienbewerber erfasst der EMS komplexe Fähigkeiten, die nur über langjährige Lern- und Entwicklungsprozesse angeeignet wurden. Er zielt nicht auf kurzfristig erlernbares Wissen ab. Die Trainierbarkeit des Tests ist somit nach den Testautoren nicht möglich. Dennoch werden in der Informationsbroschüre eine Beschäftigung mit Aufgabenbeispielen der Informationsbroschüre und des veröffentlichten Originaltests empfohlen, um sich mit dem Test vertraut zu machen. Durch Beschäftigung mit dem Testmaterial könnte übermäßige Angst reduziert werden. Spiel und Kolleginnen (2008) konnten zudem zeigen, dass unter den Aufgenommenen die Anzahl der Antritte zum Test signifikant höher war als unter den Nicht-Aufgenommenen.

3. Prädiktoren der Leistung

Im Rahmen von Zulassungsverfahren zum Hochschulstudium nehmen Studierfähigkeitstests, neben Schulnoten und schulstoffbezogenen Kenntnistests, eine wichtige Rolle zur Eignungsbeurteilung ein (Trost, 1996). So auch im Auswahlverfahren für das Studium der Medizin in Österreich. Die Testleistung ist hierbei das entscheidende Kriterium. Grundsätzlich gilt, je höher die Testleistung, desto wahrscheinlicher bekommt eine Bewerberin oder ein Bewerber einen Studienplatz. Unter Berücksichtigung der Konsequenzen des Testergebnisses für die Einzelne/den Einzelnen sowie der damit verbundenen bildungs- und gesellschaftspolitischen Konsequenzen verlangen derartige Tests nach einer effektiven und effizienten Eignungsbeurteilung. Während schulstoffbezogene Kenntnistests vorwiegend das in der Schule erworbene Wissen prüfen, erfassen Studierfähigkeitstests kognitive Fähigkeiten, die für den jeweiligen Studiengang besonders wichtig sind (Trost, 1996). Zu diesen kognitiven Fähigkeiten zählen studienfachspezifische Fähigkeiten, die aus dem Anforderungsprofil abgeleitet werden können, sowie fächerübergreifende Grundkompetenzen, die für die erfolgreiche Bewältigung eines Hochschulstudiums erforderlich sind. Dabei sollen aber keine Fähigkeiten geprüft werden, die erst im Studium erworben werden. Eine wichtige Frage zur Sicherung der Qualität des Verfahrens ist hierbei, wo und wann diese geforderten Kompetenzen erlernt werden und inwieweit sich diese von schulstoffbezogenen Kenntnissen unterscheiden. Es besteht daher ein hohes Interesse Faktoren zu identifizieren, die die Leistungsunterschiede im Test erklären und vorhersagen können. In der Literatur werden hierzu eine Fülle von personenbezogenen und nicht-personenbezogenen Faktoren (sog. Prädiktoren) beschrieben.

Grundsätzlich ist die Schule die wichtigste Institution an der fach- und bildungsspezifische Kompetenzen vermittelt werden. Aus diesem Grund ist die Frage legitim, welchen Einfluss die Schule auf die Fähigkeits- und Kompetenzentwicklung von Kindern und Jugendlichen hat. Die dort erlernten Fähigkeiten und Kompetenzen werden primär durch die Schulnoten abgebildet. In unserer Gesellschaft sind sie die wichtigste und meist verbreitetste Form der Bewertung von Leistungen. Nach Hänsgen und Spicher (2004) sind Schulleistungen und Testleistung im EMS auf eine gemeinsame Fähigkeitsdimension zurückzuführen. Schule ist aber nicht als einheitliche institutionelle Bildungsstätte zu sehen. Vielmehr aufgrund unterschiedlicher Ausbildungswege und damit verbundenen unterschiedlichen Anforderungen ergeben sich auch unterschiedliche Fähigkeitsprofile. Dies legt die Vermutung nahe, dass der besuchte Schultyp neben den schulischen Leistungen einen Einfluss auf die Testleistung hat.

Für die Sicherung der Qualität des Verfahrens ist die Einhaltung der Fairness ein wichtiges Kriterium. Eine systematische Benachteiligung einer bestimmten Gruppe gilt als unfair. Gruppeneinteilungen können nach verschiedenen Merkmalen erfolgen, die zur Sicherung der Fairness kontrolliert werden sollten (bspw. Geschlecht und Nationalität). Hierzu würde auch die Einteilung nach der regionalen Herkunft (Bsp. Stadt versus ländliche Region) zählen. Dies bedeutet aber nicht, dass es keine tatsächlichen Unterschiede in der Leistung geben kann. Auch wenn das Verfahren selbst die Personen nicht benachteiligt, so besteht dennoch ein bildungswissenschaftliches Interesse diese Einflussgrößen auf die Leistung aufzudecken und Erklärungen für deren Wirksamkeit zu finden. Durch unterschiedliche Sozialisation haben die Kandidatinnen und Kandidaten ungleiche Voraussetzungen. Dies beeinflusst die Chancengleichheit im Eignungstest. Sind diese Ungleichheitsdimensionen mit schulischen Aspekten wie beispielsweise Schultyp bzw. regionales Schulangebot, so hat dies für bildungswissenschaftliche Erkenntnisse und bildungspolitische Entscheidungen große Bedeutung. Wie sich aus der Literatur zeigt, ist der besuchte Schultyp (auch in Österreich und Deutschland) nicht unabhängig von der ländlichen Region. Daraus folgt, dass die regionale Herkunft zusätzlich einen Erklärungswert für die Leistungsunterschiede der Personen im Test liefern könnte.

Die hier vorgestellten Einflussgrößen Schulnoten, Schultyp und regionale Herkunft lassen einen engen Zusammenhang mit den geforderten Fähigkeiten und Kompetenzen der Studierfähigkeitstests vermuten. Ziel ist es die Leistungsunterschiede des Tests anhand der Prädiktoren zu erklären. Es ist verständlich, dass dies nur eine Auswahl möglicher Prädiktoren darstellt, jedoch scheinen gerade diese einen hohen Erklärungswert aufzuweisen. Folgend werden in diesem Kapitel die genannten Prädiktoren hinsichtlich ihres Erklärungswertes im Detail erörtert und anhand empirische Befunde untermauert. Am Ende des Kapitels folgt eine Zusammenfassung der wichtigsten Befunde.

3.1. Schulnoten

Schulnoten in Form von Ziffernoten sind die am meist verbreitetste Form der Klassifikation und Bewertung von Leistungen von Schülerinnen und Schülern (Wild & Krapp, 2001). Sie werden nach bestimmten Normen/Kriterien von Lehrenden an Schülerinnen und Schüler vergeben. In Zeugnissen werden die einzelnen Leistungen über eine bestimmte Dauer (meist Semester oder Jahr) in einer einzigen Ziffer zusammengefasst. Ihre historischen Wurzeln orientieren sich an dem gesellschaftlichen Bedürfnis nach Auslese und Kontrolle sowie an der Vergabe von Berechtigungen nach dem Leistungsprinzip (Tent, 2006).

Diese *gesellschaftliche Funktion* ist noch immer ein wesentlicher Bestandteil. In ihrer

einfachsten Funktion dokumentieren sie den Schulbesuch. In einer leistungsorientierten Gesellschaft wird durch den hohen Stellenwert der Schulnoten der Lebensweg jedes Einzelnen beeinflusst (Tent, 2006). So wird über Schulnoten oft der Zugang zu weiterführenden Schulen (bspw. in Österreich von der Volksschule in die AHS, siehe 3.2.1) oder zu bestimmten Studienrichtungen (bspw. Numerus clausus für Studiengänge in Deutschland) geregelt oder zur beruflichen Ausbildung und/oder Qualifikation (bspw. Wert von Schulnoten im Zuge beruflicher Bewerbungen) herangezogen. Aus der Sicht des Einzelnen besitzen Schulnoten eine „Weichenstellfunktion“, gesamtgesellschaftlich die Funktion der „Effizienzsicherung“ (Tent, 2006, S. 874).

Neben der gesellschaftlichen Funktion haben Schulnoten auch eine *pädagogische Funktion*. Sie geben als wertende Rückmeldung Auskunft über den aktuellen Leistungsstand (Berichtsfunktion), sie machen Schülerinnen und Schüler mit Leistungsvergleichen und der vorherrschenden Methode von Leistungsbewertungen in der Gesellschaft vertraut (Sozialisationsfunktion) und sie haben eine Anreizfunktion indem sie die Schülerinnen und Schüler motivieren und disziplinieren sollen (vgl. Tent, 2006; Wild & Krapp, 2001). Speziell die Anreizfunktion von Schulnoten steht häufig in der Kritik pädagogisch-psychologischer Betrachtungen. Schulnoten bewerten das Produkt und geben keine Verbesserungsvorschläge, sie regen zum Konkurrenzdenken durch intersubjektive Vergleiche unter Schülerinnen und Schülern an und sie demotivieren mehr als sie motivieren (Brügelmann, 2006). Auch das kontinuierliche Absinken der Lernmotivation im Laufe der ersten Schuljahre (Helmke, 1993) wird als Konsequenz schulischer Leistungsrückmeldungen, durch wiederholte Misserfolgserfahrungen im sozialen Kontext, gesehen (Wild & Krapp, 2001).

Schulnoten und Zeugnisse sind für die Eltern aber auch eine Möglichkeit Einblick in die Leistungsentwicklung ihrer Kinder zu nehmen. Auch für Lehrerinnen und Lehrer können die Schulnoten unter bestimmten Voraussetzungen eine Rückmeldung über die Qualität des Unterrichts geben.

Hinsichtlich der vielfältigen Funktionen von Schulnoten und der nachhaltigen Auswirkungen für den Einzelnen und die Gesellschaft müsste von einer hohen Qualitätsanforderung in den Bereichen Lern- und Leistungsdiagnostik ausgegangen werden. Beispielsweise eine oft gängige Praxis Notendurchschnitte bis auf zwei Nachkommastellen heranzuziehen (Wild & Krapp, 2001) setzt eine hohe Messgenauigkeit voraus. Der Wert der Schulnoten zur Leistungsvorhersage ist stark abhängig davon wie genau die (Fach-)Leistungen in der Schule abgebildet werden. Dies kann nur erzielt werden, wenn störende Einflüsse ausgeschaltet oder minimiert werden. Ob die Schulnoten diesen diagnostischen Qualitätsanforderungen entsprechen soll folgend unter Betrachtung einiger wesentlicher diagnostischer Gütekriterien diskutiert werden. Speziell auf die prognos-

tische Validität wird genauer eingegangen.

3.1.1. Diagnostische Gütekriterien von Schulnoten

Zahlreiche empirische Studien haben sich mit der diagnostischen Güte von Schulnoten befasst und teilweise erhebliche Mängel und Fehlerquellen nachweisen können (Wild & Krapp, 2001). Diese Fehlerquellen liegen primär in den Prozessen der menschlichen Urteilsbildung wie selektive Wahrnehmung (z.B. Beobachtungsmängel), serielle Positioneffekte (z.B. Erinnerungsfehler), Stereotypisierung, Milde-Härte-Fehler, Attribuierungsmuster (z.B. Begabungsmängel), Kontrasteffekte sowie Erwartungseffekte um nur einige zu nennen (vgl. hierzu bei Jäger & Lissmann, 2004; Bortz & Döring, 2006; Tent, 2006). Fehlerquellen spezifisch für den Schulbereich sind beispielsweise die Verwendung verschiedener Bezugsnormen oder die nicht einheitliche Gewichtung von Teilaspekten (Tent, 2006). Weiters spielen Selbstständigkeit, Leistungsbereitschaft oder sozial angepasstes Verhalten eine wichtige Rolle (Schrader & Helmke, 2001). Bereits aus diesen grundlegenden psychologischen Überlegungen heraus wird verständlich, dass Lehrerinnen- und Lehrerurteile fehlerbehaftet sein können. Zusätzlich stützt sich die Vergabe der Schulnoten nicht an standardisierte Tests, die den Urteilsfehlern entgegenwirken könnten. Der diagnostischen Kompetenz der Lehrkräfte kommt somit ein zentrales Gewicht in der Leistungsbeurteilung zu. Die Qualität dieser diagnostischen Urteile kann anhand der klassischen diagnostischen Gütekriterien Objektivität, Reliabilität und Validität bewertet werden. Wie sehr Lehrerinnen- und Lehrerurteile nach den Gütekriterien zu bewerten sind wird folgend im Detail dargestellt.

Objektivität – mangelnde Vergleichbarkeit

Forschungen zur Notengebung haben mehrfach gezeigt, dass die Lehrerinnen- und Lehrerurteile nicht ausreichend objektiv sind (vgl. Ingenkamp & Lissmann, 2005; Tent, 2006). Nicht objektiv meint hierbei, dass verschiedene Lehrkräfte die selbe Leistung unterschiedlich beurteilen. Dies führt zu einer mangelnden Vergleichbarkeit der Schulnoten. Gerade beim Übergang in die Sekundarstufe I (in Österreich von der Volksschule zur Hauptschule oder Gymnasium) ist dies besonders aufgrund der Weichenstellfunktion problematisch, denn gleichzeitig mit dem Besuch eines Schultyps sind berufliche Chancen und Berufslaufbahnen verbunden (Wallner-Paschon, 2009). Zwischen der Testleistung in einem objektiven Schultest und der Schulnote zeigt sich innerhalb einer Klasse ein hoher empirischer Zusammenhang, zwischen den Klassen und Schulen jedoch ist der Zusammenhang niedrig (Tent, 2006). Diese mangelnde Vergleichbarkeit der Schulnoten zwischen den Klassen und zwischen den Schulen entsteht dadurch, dass

Lehrerinnen und Lehrer ihre Urteile oft auf Basis eines klasseninternen Bezugsrahmens (soziale Norm) fällen. Je nachdem ob sich eine Schülerin oder ein Schüler in einer guten oder in einer schlechteren Klasse befindet, kann bei gleicher objektiver Leistung die Schulnote variieren (vgl. Ingenkamp, 1989). Für objektive Leistungsvergleiche wären ein kriteriumsorientierter Bezugsrahmen fairer (Wallner-Paschon, 2009).

Aktuelle Ergebnisse aus PIRLS 2006 (= Progress in International Reading Literacy Study, 2006) über das objektiv gemessene Leseverständnis bestätigen die mangelnde Vergleichbarkeit der Deutschnote (Wallner-Paschon, 2009). Die in PIRLS 2006 identifizierten Risikoschülerinnen und -schüler erhalten das volle Notenspektrum von Sehr Gut bis Nicht Genügend, bei den Spitzenleserinnen und -lesern von Sehr Gut bis Befriedigend. Unter Einbezug der Streuung sind für etwa 20% der Schülerinnen und Schüler bei gleicher Leseleistung die Noten von Sehr gut bis Nicht Genügend möglich. Die Auswirkungen des sozialen Maßstab zeigen sich darin, dass Spitzenschülerinnen und -schüler in Klassen mit niedrigem Leistungsniveau eine um 10 Prozentpunkte höhere Wahrscheinlichkeit haben ein Sehr gut zu bekommen als Schülerinnen und Schüler in Klassen mit hohem Leistungsniveau. Bei den Risikoschülerinnen und -schüler ist dieser Effekt kleiner.

Ein weiterer Aspekt, der die Objektivität von Schulnoten in Frage stellt, sind geschlechtsspezifische Unterschiede in der Notengebung. Mehrere empirische Befunde stützen die Annahme, dass Mädchen und Jungen mit verschiedenen Maßstäben beurteilt werden. So haben Mädchen gegenüber den Jungen oft die besseren Schulnoten, schneiden aber in objektiven Leistungstests schlechter ab. Nach einer Studie des deutschen Bundesministeriums für Bildung und Forschung (Budde, 2008) erhalten Jungen in allen Fächern bei gleichen Kompetenzen schlechtere Noten als Mädchen. Auch Kenney-Benson, Pomerantz, Ryan und Patrick (2006) konnten zeigen, dass trotz besserer Mathematiknoten der Mädchen in zwei Jahrgangsstufen (5. und 7. Stufe) diese keine besseren Testleistungen erzielten als die Jungen. Darüber hinaus konnten die Mädchen ihren Vorsprung in den Mathematiknoten über die zwei Jahre ausbauen. Auch Wallner-Paschon (2009) konnte anhand der PIRLS 2006-Daten zur Leseleistung (Risiko- und Spitzenschülerinnen und -schüler) zeigen, dass Mädchen in Deutsch weniger streng beurteilt werden als Jungen. In der Spitzengruppe wurden signifikant mehr Sehr Gut (62% zu 74%) und in der Risikogruppe mehr Sehr Gut und Gut (24% zu 35%) an die Mädchen vergeben.

Ein deutliches Beispiel für evidente Geschlechtsunterschiede in der Notengebung zeigten Spiel et al. (2008) in ihrer Evaluation des Eignungstests für das Medizinstudium (EMS) in Österreich. Bei gleichen Schulnoten zeigten die Bewerberinnen in Wien und Innsbruck durchgängig für alle Notenstufen deutlich schlechtere Testleistungen als die

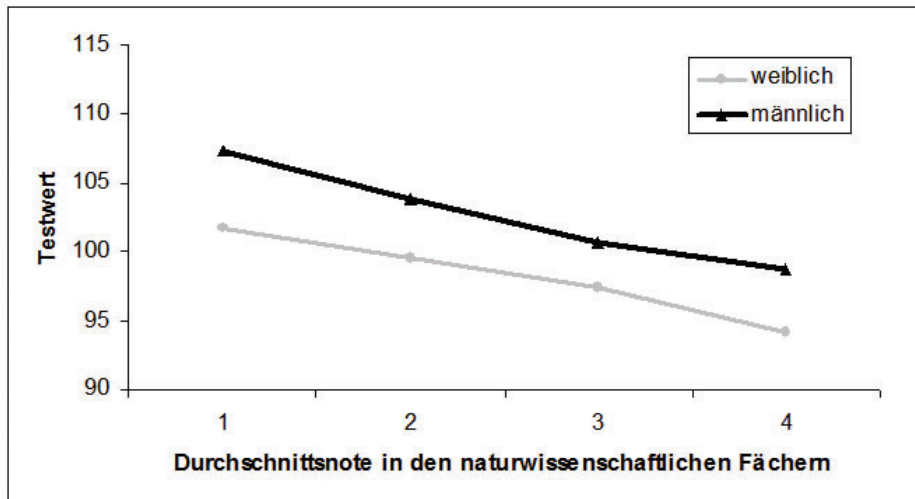


Abbildung 3.1.: Durchschnittliche Testwerte nach Durchschnittsnote in den naturwissenschaftlichen Fächern Mathematik, Chemie, Physik und Biologie getrennt nach Geschlecht (Spiel, Schober & Litzenberger, 2008)

Bewerber. Abbildung 3.1 zeigt die durchschnittlichen Testwerte nach Durchschnittsnote in den naturwissenschaftlichen Fächern Mathematik, Chemie, Physik und Biologie getrennt nach Geschlecht. Die Unterschiede sind substantiell. So erreichen die Männer im Mittel mit fast zwei Notengraden schlechterer Beurteilung die gleichen Testwerte wie Frauen. Der Vorsprung der Männer gegenüber den Frauen beträgt über alle Notenstufen durchschnittlich 4.37 Punkte (siehe Tabelle 16.1 von Spiel et al., 2008, S. 163).

Mögliche Erklärungen für die unterschiedliche Bewertung von Mädchen und Buben könnten im Einbezug nicht-fähigkeitsspezifischer Kriterien wie schulangepasstes Sozialverhalten liegen (Schrader & Helmke, 2001) in denen Mädchen üblicherweise besser abschneiden. Die Literatur zeigt, dass Mädchen fleißiger sind, mehr für die Schule arbeiten und genauer vorgehen (Wagner & Spiel, 2002).

Reliabilität

Analog zur Testreliabilität (siehe 2.2.2) ist für Schulnoten die Zuverlässigkeit der Messung im Sinne des diagnostischen Gütekriteriums von Interesse. Vorweg kann gesagt werden, dass die Reliabilitäten von Schulnoten deutlich befriedigender sind als deren Objektivität. Dies trifft einerseits sowohl auf die Retest-Reliabilität (Stabilität) über die Schuljahre hinweg zu, als auch andererseits auf die Urteils Konstanz von Lehrerinnen und Lehrern bei wiederholter Beurteilung der selben Leistungen (vgl. Tent, 2006).

Die Retest-Reliabilität in der Grundschule liegt bei Notendurchschnitten von Schuljahr zu Schuljahr bei $r > .8$ und bleibt auch über einen Zeitraum von drei Jahren mit

r bis zu .8 sehr stabil (Tent, 2006, S. 876). In der Sekundarstufe verringert sie sich ein wenig. Erwartungsgemäß sinkt die Stabilität der Abschlussnoten über die Schuljahre (vom 1. zum 8. Schuljahr $r = .2$, S. 876). Der Übergang in die Sekundarstufe ist allgemein mit einem Notenknicke verbunden – außer für die Hauptschule, in der die Noten besser werden.

Nach Tent (2006) weisen die Schulnoten einige stabile Charakteristika auf. So ist die Rangordnung der Notendurchschnittswerte als auch der Noten-Streuungen stabil. Die Notenskala wird über die einzelnen Fächer hinweg unterschiedlich stark ausgeschöpft. In der Grundstufe werden die Fächer Deutsch und Mathematik am strengsten beurteilt. Dies zeigt sich auch in PIRLS 2006 (Wallner-Paschon, 2009) in Österreich. Die Semesternoten im Unterrichtsgegenstand Deutsch werden im Vergleich zu den anderen Hauptfächern eher streng beurteilt. Die Ausschöpfung der Notenskala zeigt, dass etwa zwei Drittel der Lehrkräfte 4 Notenstufen und knapp mehr als ein Viertel 3 Notenstufen zur Beurteilung ausschöpfen. Etwa ein Prozent der Lehrkräfte nutzen nur 2 Notenstufen zur Beurteilung (siehe Abbildung 4.3.2 von Wallner-Paschon, 2009). In der Sekundarstufe werden zusätzlich die Fremdsprachen streng benotet – Religion, Musik und Kunst hingegen am mildesten (Tent, 2006).

Validität und Prognosekraft

Grundsätzlich und unabhängig davon, ob Schulnoten das zu messende Merkmal objektiv und/oder reliabel erfassen, besitzen Schulnoten eine 'logische Validität'¹, weil der Schulerfolg von Amts wegen durch diese definiert ist. Dies ist jedoch aus pädagogischer und bildungspsychologischer Sicht unbefriedigend. Durch die gesetzliche Verankerung wird der pädagogische Handlungsspielraum der Lehrerinnen und Lehrer begrenzt (Wild & Krapp, 2001). Die Einbeziehung anderer Beurteilungsformen als Ziffernoten wird dadurch erschwert. Insbesondere im unteren Skalenbereich gibt die Schulnote keine Informationen zur Behebung des Defizits (Tent, 2006). Um die vielfältigen Funktionen von Schulnoten besser zu verstehen sind gerade empirische Befunde von besonderem Interesse.

Wird für das Zustandekommen der Schulnoten ein dahinter liegendes kognitives Konstrukt vermutet, so sollten sich deutliche Übereinstimmungen mit anderen Kriterien ergeben, die das gleiche Konstrukt zu messen beanspruchen. So zeigt sich eine empirisch gut abgesicherte moderate Beziehung zwischen der mittleren Schulleistung und dem Konstrukt allgemeiner Intelligenz (mittlere Korrelation von $r = .50$ bis $r = .60$,

¹Von logischer Validität im Sinne „inhaltlicher Gültigkeit eines Tests ist zu sprechen, wenn dieser selbst, quasi definitionsgemäß, das optimale Kriterium des interessierenden Merkmals darstellt.“ (Kubinger, 2006, S. 51).

Helmke & Schrader, 2006, S. 84). Mit den einzelnen Fachnoten ist der Zusammenhang niedriger, ausgenommen Mathematik (Tent, 2006). Mehrfach zeigte sich, dass Fachnoten einem Halo-Effekt unterliegen in der Form, dass die Beurteilung in einem Fach auf Beurteilungen in anderen Fächern Einfluss hat – dies spricht für eine mangelnde Abgrenzung der Fähigkeiten aus den verschiedenen Fächern (Wild & Krapp, 2001).

In Hinblick auf die Tatsache, dass Schulnoten oft als Kriterium bei wichtigen Lebensentscheidungen (wie Bildungsweg oder Berufslaufbahnentscheidungen) herangezogen werden, kommt der Betrachtung der Vorhersagegültigkeit (*prognostische Validität*; siehe 2.2.2) von Schulnoten eine enorme Bedeutung zu. Ziel ist es, ein bestimmtes und in der Zukunft liegendes Kriterium vorherzusagen. Dass diese Leistung von den Schulnoten grundsätzlich gefordert wird verwundert nicht, sind sie ja ein Maß der in der Schule erworbenen Fähigkeiten und Kenntnisse aus verschiedenen Bereichen, der Umsetzung von Intelligenz sowie des Arbeitsverhaltens oder Lernschwierigkeiten. Mag 'für das Leben Lernen' ein zu fernes und zu breit gefasstes Ziel sein, so könnten strukturell ähnliche Leistungsfelder, wie der weitere Bildungserfolg (Sekundarstufe, Studium), ferner Berufserfolg vorhergesagt werden (vgl. Schuler, 2006).

Zur Prognose des *Sekundarschulerfolgs* durch Grundschulleistungen gibt es wenig empirische und kaum aktuelle Befunde. Sauer und Gamsjäger (1996) führten in Österreich eine Längsschnittstudie zur Vorhersage des Sekundarschulerfolgs (gemessen an den Schulnoten; Hauptschule und Gymnasiums) neben anderen Prädiktoren anhand von Schulleistungen der Volksschule durch. Mit Hilfe von Pfadanalysen prognostizierten sie den Schulerfolg nach dem ersten und vierten Jahr. Als wichtigste Prädiktoren des Gymnasialerfolgs erwiesen sich nach dem ersten Jahr die Volksschulnoten und Schulleistungstests, nach dem vierten Jahr zusätzlich die Motivation. Die Ergebnisse für die Hauptschule waren zu beiden Zeitpunkten komplexer und der Einfluss der Volksschulnoten zeigte sich nicht so deutlich wie im Gymnasium. Baeriswyl, Trautwein, Wandeler und Lüdtke (2009) versuchten anhand von Mehrebenenanalysen für kreuzklassifizierte Daten die Mathematik- und Deutschnoten am Ende der obligatorischen Schulzeit (Klassenstufe 9) im Kanton Freiburg (Schweiz) vorherzusagen. Neben den Schulnoten und standardisierten Leistungstests am Ende der Grundschulzeit wurden Lehrerinnen- und Lehrerempfehlungen untersucht. Zusammenfassend kommen die Autoren zur Einschätzung, dass die Grundschulnoten eine vorhandene jedoch begrenzte prognostische Validität aufweisen und diese durch den Einsatz von standardisierten Leistungstests verbessert werden kann.

Für die Vorhersage des *Studiumserfolgs* ist die Abiturdurchschnittsnote mit Abstand das häufigste Einzelkriterium, das bei der hochschuleigenen Studierendenauswahl herangezogen wird (Heine et al., 2006). In fast allen Ländern ist die schulische Leistung

ein wesentliches und in zahlreichen Ländern das einzige Kriterium, das bei der Zulassungsentscheidung herangezogen wird (Deidesheimer Kreis, 1997). Schul- und Studienanforderungen überschneiden sich in einigen Anforderungsbereichen, zudem sind Schulnoten ohne größeren Aufwand (bspw. einen zusätzlichen Leistungstest) vorhanden und im Rahmen eines Auswahlverfahrens für ein Hochschulstudium leicht verfügbar. Eine gute prognostische Validität von Durchschnittsnoten für den späteren Studienerfolg belegen einige Metaanalysen mit Korrelationen um $r = .40$ (vgl. Burton & Ramist, 2001; Robbins, Lauver, Le, Davis, Langley & Carlstrom, 2004; Trapmann, Hell, Weigand & Schuler, 2007). Einzelfachnoten erweisen sich hingegen mehrfach als weniger valide gegenüber den Durchschnittsnoten. Trapmann et al. (2007, S. 14) konnten in ihrer Metaanalyse mit Daten aus fünf Ländern zeigen, dass von den Einzelfachnoten Mathematik eine mittlere Validität von $r = .27$ (ungewichtet) bis $r = .32$ (reliabilitäts- und selektionskorrigiert, $k = 13$, $N = 1766$) aufweist. Die Einzelnote im Unterrichtsfach der jeweiligen Muttersprache korrelierte schon deutlich niedriger. Die Validität der Mathematiknote viel für Medizinische Disziplinen allgemein mit $r = .22$ (korrigiert, $k = 3$, $N = 314$, S. 21) niedrig aus. Speziell für das Studienfach Humanmedizin (nur für deutsche Studierende) ergab sich eine mittlere Validität der Mathematiknote von $r = .40$ (ungewichtet) bis $r = .55$ (korrigiert, $k = 7$, $N = 33576$, S. 21).

Sehr häufig wurde der Zusammenhang von Schulnoten und objektiven *Leistungstests* untersucht. Die empirisch gefundenen Korrelationen zeigen in einer Mehrzahl von Studien eine gute Einschätzung der Schülerinnen- und Schülerleistungen (vgl. Demaray & Elliott, 1998; Hoge & Coladarci, 1989). Dieser Befund ist angesichts der mangelnden Objektivität von Schulnoten höchst interessant und unerwartet. Trotz der mangelnden Vergleichbarkeit der Noten und der Anwendung einer sozialen Norm bei der Notengebung bilden diese die Leistungen der Schülerinnen und Schüler recht gut ab. Eine mögliche Erklärung hierfür wäre, dass Lehrerinnen und Lehrer bei der Benotung einer erbrachten Schülerinnen- oder Schülerleistung zusätzlich andere Informationen einfließen lassen wie beispielsweise frühere Leistungsbeurteilungen und/oder Verhaltensbeobachtungen. Dadurch wird die Objektivität der Beurteilung verringert. Bei Vergleichen der erbrachten Leistung durch verschiedene Lehrkräfte wären diese Informationen nicht verfügbar. Unter realen Bedingungen aber fließen diese Leistungsinformationen bei der Notengebung ein und scheinen die Gültigkeit des Urteils zu erhöhen. Zudem erfassen Studien über die Vergleichbarkeit des Lehrkräfteurteils meist nur einzelne oder wenige Schülerinnen- und Schülerleistungen wohingegen Zeugnisnoten aus einer Vielzahl an Leistungsbeurteilungen gebildet werden.

In ihrer Metaanalyse fanden Hoge und Coladarci (1989) für den Zusammenhang zwischen Lehrerinnen- und Lehrerurteil und Schulleistung einen Median von $r = .66$

wobei die Schwankung im Bereich von $.28 \leq r \leq .92$ lag (S. 303). In den USA gibt es langjährige Erfahrungen in der autonomen Auswahl von Studierenden mit Studierfähigkeitstests. Zu den bekanntesten allgemeinen Studierfähigkeitstests zählen der Scholastic Aptitude Test (SAT) und der American College Test (ACT). In den drei Bereichen des SAT (critical reading, mathematics und writing) zeigen sich Korrelationen mit High School Noten² im Bereich von $.45 \leq r \leq .49$, *adjusted*, für Gesamt-SAT eine Korrelation von $r = .53$, *adjusted* (Kobrin, Patterson, Shaw, Mattern & Barbuti, 2008, S. 5).

Mit dem Test für Medizinische Studiengänge (TMS) in Deutschland wurden in zehn Jahre langer Erfahrung mehrere Evaluationen zur prognostischen Validität durchgeführt. Zusammenhänge zwischen dem Kriterium Testergebnis und Abiturdurchschnittsnote zeigten Werte von minimal $r = .37$ bis maximal $r = .48$ (Trost et al., 1998, S. 56). Auch für den Eignungstest für das Medizinstudium (EMS) sind die Ergebnisse vergleichbar. Gute Maturaleistungen korrelieren auch mit guten Testleistungen im EMS $r = .48$ (Hänsgen & Spicher, 2004).

Die inhaltliche Ausrichtung des TMS und des EMS legt nahe, dass mathematisch-naturwissenschaftliche schulische Leistungen für den Testerfolg relevant sind. Kandidatinnen und Kandidaten, die sich schon in der Schule für Mathematik und Naturwissenschaft interessierten, schnitten im TMS besser ab (Trost et al., 1998). Vor allem die Mathematiknote erweist sich für den Studienerfolg am Beginn des Medizinstudiums als wichtiger Prädiktor (Haidinger, Frischenschlager & Mitterauer, 2005).

²Grade Point Average (GPA), High School GPA.

3.2. Schultypen

Im Zuge der historischen Entwicklung des Bildungssystems sind aufgrund unterschiedlicher Anforderungen verschiedene Ausbildungswege entstanden. Im österreichischen Schulsystem gibt es eine Vielzahl an möglichen Ausbildungswegen nach einer 4-jährigen Grundschulzeit zu Beginn der Sekundarstufen I und II. Abbildung 3.2 zeigt eine schematische Darstellung des österreichischen Bildungssystems.

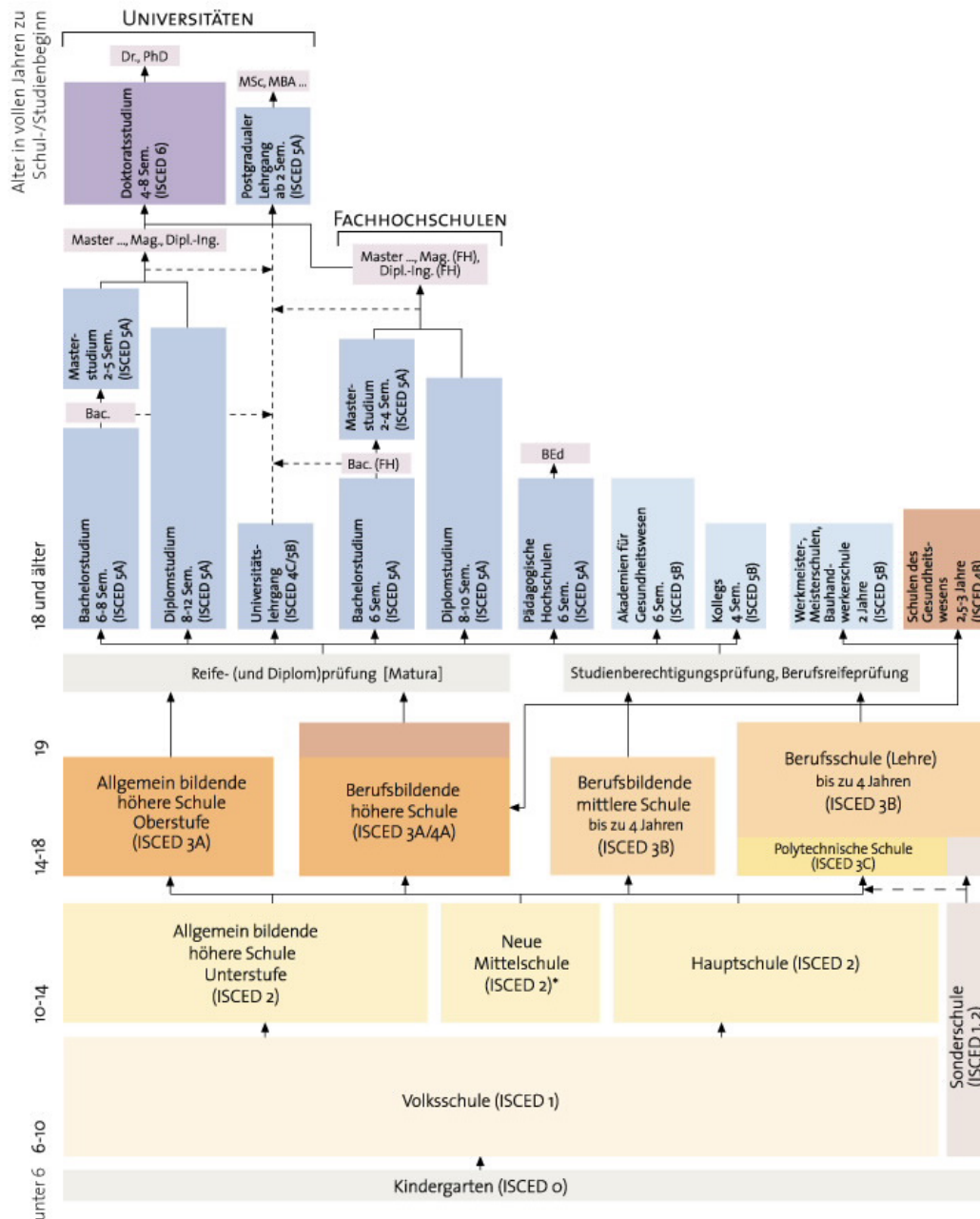
An die unterschiedlichen Schultypen werden einerseits unterschiedliche Anforderungen und andererseits auch unterschiedliche Erwartungen gestellt. Die Anforderungen unterscheiden sich beispielsweise hinsichtlich des zu erreichenden formalen Bildungsabschlusses, der zu erwerbenden Qualifikationen (Ausbildungsziele) und der Dauer des Ausbildungsweges. Die Erwartungen an den jeweiligen Schultyp sind vielfältig, beispielsweise dass der Besuch des Gymnasiums gegenüber der Hauptschule mit einem günstigeren Lernverlauf verbunden ist (Baumert, Stanat & Watermann, 2006). Die Wahl des Schultyps wird aber auch durch die soziale Schichtzugehörigkeit beeinflusst. So besuchten 2001 in Österreich 77.3% der 12-jährigen die Allgemein bildene höhere Schule – Unterstufe deren Eltern Hochschulabschluss haben aber nur 12.3% bei höchstens Pflichtschulabschluss (Bauer, 2005). Als zentrales Ergebnis von PISA 2000 (Baumert, Trautwein & Artelt, 2003) zeigte sich in Deutschland ein hoher Zusammenhang zwischen der Sozialschicht (HISEI³) und dem Schultyp in der Lesekompetenz.

Eine Verteilung der Schülerinnen und Schüler hinsichtlich ihrer Leistung auf die verschiedenen Schultypen ist nicht nur erwünscht, sondern nach Baumert et al. (2003) auch die eigentliche Idee. Dem Ziel der Homogenisierung von Schülergruppen liegt die Vorstellung zugrunde, dass der Unterricht so besser auf die individuellen Voraussetzungen abgestimmt werden kann. Da aber die Auswahl des Schultyps nicht unabhängig von der sozialen Schichtzugehörigkeit ist (siehe oben), ist die frühe Verteilung auf unterschiedliche Bildungswege mit dem Nebeneffekt der sozialen Segregation verbunden. Somit unterteilt der Schultyp nicht nur leistungsspezifisch sondern auch sozial. Weiter kommt hinzu, dass der Übertritt in einen bestimmten Schultyp der Sekundarstufe auch an regionale Gegebenheiten gebunden ist (siehe Abschnitt 3.3).

Folgend werden die wichtigsten Ausbildungswege im österreichischen Schulsystem nach der Grundschulstufe (Volksschule) und nach der Sekundarstufe I mit Fokus auf die unterschiedlichen Lehrziele kurz vorgestellt. Auf die Sonderschulformen wird nicht eingegangen. Im Anschluss werden Lehrpläne ausgewählter Schultypen in einigen Merkmalen verglichen. Die folgenden Informationen über die einzelnen Schultypen stammen vom Bundesministerium für Unterricht, Kunst und Kultur – bm:ukk (2010, April).

³HISEI = highest ISEI; International Socio-Economic Index of Occupational Status

3 Prädiktoren der Leistung



ISCED: International Standard Classification of Education (UNESCO)

* ...Beginn Schuljahr 2008/09

Abbildung 3.2.: Aufbau des österreichischen Bildungssystems (Statistik Austria, 2010).

3.2.1. Schultypen und Lehrziele

Nach der 4-jährigen Volksschule kann mit Eintritt in die Sekundarstufe I (5.Schulstufe) zwischen dem Besuch der Hauptschule und dem Besuch einer Allgemeinbildenden höheren Schule – Unterstufe (AHS-Unterstufe) gewählt werden. Die Ausbildungszeit in beiden Schultypen dauert vier Jahre. Lehrziele der *Hauptschule* ist die Vermittlung einer grundlegenden Allgemeinbildung. Diese soll für eine Berufsausübung als auch zum Übertritt in eine mittlere und höhere Schule befähigen. Für den Übertritt in die *AHS-Unterstufe* sind nach erfolgreich abgeschlossener Volksschule einige Voraussetzungen verknüpft. Sehr gute bis gute Noten der Fächer Deutsch+Lesen und Mathematik oder eine Empfehlung der Schulkonferenz müssen vorliegen. Andernfalls kann eine Aufnahmeprüfung abgelegt werden. Ziel ist die Vermittlung einer umfassenden und vertiefenden Allgemeinbildung sowie die Befähigung zum Übertritt in die AHS-Oberstufe. Während in der 1. und 2. Stufe die selben Fächer wie in der Hauptschule unterrichtet werden, kann ab der 3. Stufe der Sekundarstufe I zwischen drei Ausbildungsrichtungen gewählt werden: (1) Gymnasium (mit Latein), (2) Realgymnasium (Latein, zweite lebende Fremdsprache, dazu Darstellende Geometrie, mehr Mathematik und Naturwissenschaften) und (3) Wirtschaftskundliches Realgymnasium (Latein oder zweite lebenden Fremdsprache, dazu mehr Chemie, Wirtschaftskunde, Psychologie und Philosophie) (Bundesministerium für Unterricht, Kunst und Kultur, 2010, März).

Mit Beginn der 9. Schulstufe kann zwischen dem Besuch der AHS-Oberstufe, Oberstufen-Realgymnasien, einer berufsbildenden mittleren und höheren Schule und einer polytechnischen Schule gewählt werden. Die *AHS-Oberstufe* kann nach der AHS-Unterstufe oder auch nach der Hauptschule (mit einigen Voraussetzungen) besucht werden. Neben einer fundierten Allgemeinbildung sollen die Voraussetzungen für ein Universitätsstudium geschaffen werden. Nach der Hauptschule kann in der 9. Schulstufe auf ein *Oberstufen-Realgymnasium* gewechselt werden. Dieses bietet die Möglichkeit einer „kreativen oder naturwissenschaftlichen Schwerpunktsetzung“ sowie die Wahl zwischen einer zweiten lebenden Fremdsprache oder Latein. Die *polytechnische Schule* dauert 1 Jahr und wird vorzugsweise von den 14-15 jährigen zum Besuch der 9. Schulstufe und anschließender Berufsausbildung genutzt. Ziel ist es einen qualifizierten Übertritt in eine Lehrlingsausbildung (duale Berufsausbildung: Lehrbetrieb und Berufsschule) oder den Übertritt in weiterführende Schulen zu ermöglichen.

Weiter kann ab der 9. Schulstufe (Beginn der Sekundarstufe II) zwischen berufsbildenden mittleren Schule (BMS) und berufsbildenden höheren Schule (BHS) gewählt werden. Die *BMS* (oder Fachschule) dauert 3 bis 4 Jahre und endet mit einer Abschluss-

prüfung und liefert neben einer fundierten Allgemeinbildung eine berufliche Erstausbildung. Eine Weiterqualifizierung mit der Berufsreifeprüfung oder einem Aufbaulehrgang sind möglich. Die Ausbildungsdauer der *BHS* ist 5 Jahre und somit um 1 Jahr länger als die der *AHS-Oberstufe*. Angestrebt werden neben einer fundierten Allgemeinbildung eine höhere berufliche Ausbildung, die mit der Reife- und Diplomprüfung abschließt. Neben fachspezifischen beruflichen Qualifikationen wird die allgemeine Hochschulreife erworben (Doppelqualifikation).

Die unterschiedlichen Lehrziele und angestrebten Qualifikationen, unterschiedliche Ausbildungsdauer sowie einzelne Zugangsvoraussetzungen rechtfertigen eine Annahme, dass die allgemeinen und fachlichen Kenntnisse der Absolventen ja nach Schultyp unterschiedlich sind. Bei jeder Übertrittsmöglichkeit (5. und 9. Schulstufe) fächert sich das Schulsystem immer mehr auf. Über postsekundäre Ausbildungen (Berufsreifeprüfung und Aufbaulehrgang) kann die Hochschulreife auch erworben werden. Bis zum Erwerb der Hochschulreife liegen daher sehr heterogene Bildungswege zurück.

Das deutsche Schulsystem ist dem österreichischen in seiner vertikalen Gliederung ähnlich. Auch hier gibt es Wahlmöglichkeiten an den Übergängen zu den Sekundarstufen I und II. Vor dem Übergang in die Sekundarstufe I gibt es in einigen Bundesländern eine zweijährige Orientierungsphase. Die Gliederung in Haupt-, Realschule und Gymnasium stellt eine externe Differenzierung nach Leistung dar. Innerhalb eines Schultyps findet der Unterricht in leistungshomogenisierten Gruppen statt. Im Unterschied zu Österreich wird in Deutschland (in fast allen Bundesländern) das Abitur zentral durchgeführt. Dies soll eine höhere Transparenz in der Leistungsbeurteilung und eine bessere Vergleichbarkeit der Bildungsqualifikation schaffen. Ein Überblick über die Struktur des deutschen Bildungssystems siehe Anhang Abbildung A.2.

Einen weiteren Einblick in die Unterschiede zwischen den Schultypen können die Lehrpläne liefern. Am Beispiel zweier mit Matura abschließenden Schultypen 'Ausbildungszweige der *AHS-Oberstufe*' und 'Höherer Technische und Gewerbliche Lehranstalt (HTL)' werden ein paar ausgewählte Unterschiede exemplarisch aufgezeigt. Die HTL wurde aufgrund ihres größten Anteils an Einstiegsschülern der *BHS-Formen* mit 43.3% für das Schuljahr 2006/07 (Statistik Austria, 2009) gewählt.

3.2.2. Lehrpläne – *AHS-Oberstufe* und HTL

Die höheren technischen und gewerblichen Lehranstalten (HTL) zählen zu den berufsbildenden höheren Schulen. Die Ausbildungsdauer beträgt 5 Jahre. Neben dem Erwerb der Hochschulreife vermitteln sie „eine hochwertige fachtheoretische und fachpraktische Bildung als Grundlage für einen effizienten Einstieg in das Berufsleben und für

eine erfolgreiche Tätigkeit in verschiedenen Einsatzbereichen.“ (Bundesministerium für Unterricht, Kunst und Kultur – bm:ukk, April, 2010). Allgemeine und soziale Qualifikationen befähigen zum Selbststudium und am Prozess des lebenslangen Lernens teilzunehmen.

Tabelle 3.1.: Wochenstundenzahl in den Schultypen Gymnasium, Realgymnasium, Wirtschaftskundliches Realgymnasium, Oberstufenrealgymnasium und Höhere Technische Lehranstalt (HTL). In Klammer die Mindest-Wochenstundenzahl bei schulautonomer Lehrplanänderung.

Schulfach	Gymnasium	Wirtschafts-			HTL
		Real- gymn.	kundl. Realgymn.	Oberstufen- realgymn. ¹	
Deutsch	12 (11)	12 (11)	12 (11)	13 (12)	11
1. lebende Fremdsprache	12 (11)	12 (11)	12 (11)	12 (11)	11
Mathematik	12 (11)	14 (13)	12 (11)	15 (13)	13.5 ²
Biologie	6 (6)	9 (7)	7 (7)	7 (7)	
Chemie	4 (4)	5 (5)	4 (4)	6 (5)	
Physik	7 (5)	9 (7)	7 (5)	8 (7)	
gesamt					
Wochenstunden	130	130	130	130	195

¹ ... mit Darstellender Geometrie oder ergänzendem Unterricht in Biologie und Umweltkunde, Physik sowie Chemie

² ... Mittelwert der Wochenstundenzahl in Mathematik aus 8 verschiedenen HTL-Fachrichtungen

Die Gesamt-Wochenstundenzahl in den Ausbildungsrichtungen der AHS-Oberstufe beträgt inklusive Wahlpflichtgegenstände 130 Stunden, in den HTLs sind es in 5 Jahren 195 Wochenstunden. Tabelle 3.1 gibt einen Überblick über die Wochenstundenzahl in den Haupt- und naturwissenschaftlichen Fächern AHS-Oberstufe und HTL im Vergleich. Die Wochenstundenzahl der Fächer Deutsch und 1. lebenden Fremdsprache (in der HTL Englisch) sind in den Schultypen sehr ähnlich. Unterschiede zeigen sich vorwiegend in Mathematik und in den Ausbildungszweigen der AHS in den naturwissenschaftlichen Fächern Biologie, Chemie und Physik. Die höchste Wochenstundenzahl in Mathematik hat das Oberstufenrealgymnasium gefolgt vom Realgymnasium. Im Realgymnasium zeigt sich der Schwerpunkt auf naturwissenschaftliche Fächer in der Wochenstundenzahl am deutlichsten. Je nach schulautonomer Entscheidung kann es Abweichungen in der Wochenstundenzahl geben, jedoch ein Mindestmaß an Stunden muss in der Schulzeit angeboten werden. Für die Höheren Technischen Lehranstalten können keine Wochenstunden zu den naturwissenschaftlichen Fächern in vergleichbarer Form gegeben werden, da die Ausbildung durch eine Vielzahl an Fächern fächerüber-

greifend zu einer naturwissenschaftlich fachspezifischen Ausbildung führt.⁴

Die mit den unterschiedlichen Ausbildungszielen verbundene Schwerpunktsetzung in den Lehrplänen, lassen eine gerechtfertigte Annahme von Qualifikationsunterschieden zwischen den einzelnen Schultypen zu. Insbesondere in Mathematik und den naturwissenschaftlichen Fächern gibt es deutliche Unterschiede. Querschnittliche Leistungsunterschiede zwischen den Schultypen sind aufgrund der leistungsspezifischen Aufteilung (siehe oben) zu erwarten und können daher nicht primär als Einfluss des Schultyps interpretiert werden. Dieser zeigt sich über den Längsschnitt. Folgend werden daher die unterschiedliche Leistungsentwicklung in den verschiedenen Schultypen anhand einiger empirischer Befunde dargestellt.

3.2.3. Leistungsunterschiede zwischen Schultypen – Empirische Evidenzen

Empirische Evidenzen für Schereneffekte, im Sinne dass die Leistungsunterschiede der Schülerinnen und Schüler zwischen den Schultypen über die Schuljahre hinweg immer mehr auseinander gehen, konnten bereits mehrfach erbracht werden (z.B. Köller & Baumert, 2001; Baumert et al., 2003; Klieme et al., 2006; Becker, Lüdtke, Trautwein & Baumert, 2006; Neumann et al., 2007). Die Unterschiede können aber nach Becker et al. (2006) nicht direkt als insitutionale Effekte interpretiert werden. Als weitere Erklärungsansätze für die gefunden Unterschiede werden Eingangselektivität und Kompositionseffekte genannt, die in den empirischen Untersuchungen berücksichtigt werden müssen. Eingangselektivität bedeutet, dass die Schülerinnen und Schüler sich bereits vor dem Eintritt in die Sekundarstufe in ihren Leistungen und Lernraten unterscheiden haben. Als Kompositionseffekte werden Einflüsse die auf „leistungsmäßige, soziale, kulturelle und lernbiographische Zusammensetzung der Schülerschaft“ zurückzuführen sind bezeichnet (Becker et al., 2006, S. 234).

Für das deutsche mehrgliedrige Schulsystem ab der Sekundarstufe I sind die Schereneffekte, die primär zugunsten des Gymnasiums ausfallen, empirisch gut gesichert. Abbildung 3.3 zeigt Ergebnisse in drei Unterrichtsfächern getrennt nach Schultyp aus der Kohorten-Längsschnittstudie „Bildungsprozesse und psychosoziale Entwicklung im Jugendalter und jungen Erwachsenen“ (BIJU) vom Beginn der 7. bis zum Ende der 10. Jahrgangsstufe. Es zeigt sich, dass die Zuwächse im Gymnasium stärker sind (steilerer Anstieg) als in den drei anderen Schultypen Real-, Gesamt und Hauptschule. Hinweise auf die Leistungsschere in der Sekundarstufe I finden sich auch in Befunden der Studie Deutsch-Englisch-Schülerleistungen-International (DESI, vgl. Klieme et al. 2006). Im

⁴Details zu den Lehrplänen der HTLs finden sich unter http://www.bmukk.gv.at/medienpool/11755/ahs_lp_broschuere.pdf und <http://www.htl.at/de/home/lehrplaene.html>

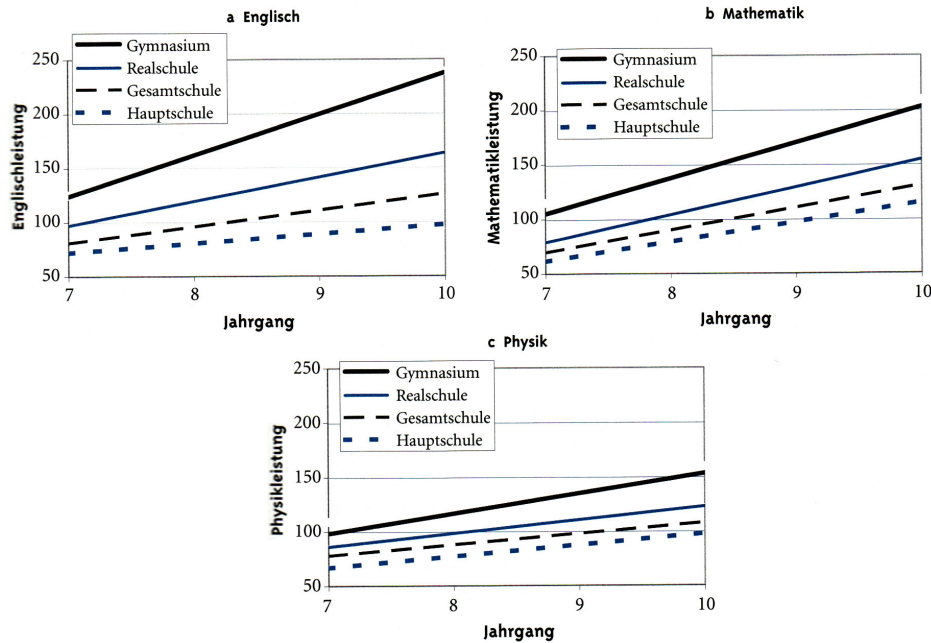


Abbildung 3.3.: Entwicklung der Schulleistungen in den Fächern Englisch, Mathematik und Physik von der 7. bis zur 10. Jahrgangsstufe getrennt nach Schultyp (Köller & Baumert, 2002).

Laufe der 9. Jahrgangsstufe erzielten im Kompetenzbereich Deutsch (Teilbereich Lesekompetenz und Sprachbewusstheit) als auch in Englisch (Teilbereich Hörverstehen) Schülerinnen und Schüler im Gymnasium höhere Lernzuwächse als Schülerinnen und Schüler in den anderen Schultypen.

Einen weiteren Befund erbrachten Becker und Mitarbeiter (2006) anhand der Daten aus der deutschen Erweiterung der dritten internationalen Mathematik- und Naturwissenschaftsstudie (TIMSS). Mittels eines Mehrgruppen-Strukturgleichungsmodells zeigten die Autoren zwischen der 7. und 8. Jahrgangsstufe deutliche höhere latente Zuwächse in den Lernständen zwischen allen drei Schultypen. Dabei fielen die Zuwächse (in Standardabweichungen) im Gymnasium ($d = .79$) und in der Realschule ($d = .60$) mehr als doppelt so hoch aus wie die in der Hauptschule ($d = .25$) (S. 238). Die unterschiedliche Eingangsselektivität wurde nicht kontrolliert, jedoch kommen Becker und Mitarbeiter (2006) zur Auffassung, dass das Auseinanderklaffen der Leistungen auf differenzielle Entwicklungsverläufe in den Schultypen zurückzuführen sei und nicht allein aufgrund von Ausgangsunterschieden erklärt werden kann.

Eine neuere Studie die speziell auf das Problem der Eingangsselektivität sowie auf Kompositionseffekte eingeht stammt von Neumann et al. (2007). Auf Basis zweier repräsentativer Stichproben der schweizer Kantone Wallis ($N = 923$) und Fribourg ($N = 778$) wurden die Leistungszuwächse im Fach 'Französisch als Fremdsprache'

während der 8. Jahrgangsstufe untersucht. Neumann et al. (2006) fanden unter Berücksichtigung der Eingangsselektivität (einschließlich des Vorwissens) bedeutsame Unterschiede in den Leistungszuwächsen zwischen den Schultypen (Gesamtstichprobe: $d = .32$, S. 410). Es ist zu erwarten, dass die Unterschiede über längere Untersuchungszeiträume stärker ausfallen würden. Aus den Befunden lässt sich weiter ableiten, dass die schultyp-spezifischen Entwicklungsverläufe nicht allein auf Schülerzusammensetzungen und Gruppenprozesse zurückzuführen sind, sondern stärker durch „institutionell verankerte und über das Unterrichtsgeschehen vermittelte curriculare und didaktische Vorgaben geprägt werden“ (S. 415).

Für den Bereich der Sekundarstufe II gibt es deutlich weniger Befunde als für die Sekundarstufe I. Für Deutschland zeigt sich in den nationalen Ergebnissen der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie (Third International Mathematics and Science Study – TIMSS), dass sich die gefundene Leistungsschere zwischen den Schultypen der Sekundarstufe I in der Sekundarstufe II tendenziell weiter öffnet (Baumert, Bos, Brockmann, Gruehn, Klieme, Köller et al., 2000). Besonders scheinen die Leistungsunterschiede zwischen Real- und Hauptschülern zu wachsen. Selbst unter Kontrolle der abgeschlossenen Schulbildung und der erreichten fachlichen Grundbildung zeigen sich beim Übergang in die Sekundarstufe II institutionelle und individuelle Selektionsprozesse die ineinander greifen (Baumert et al., 2000). An Bildungsabschlüssen gebundene Berechtigungssysteme begünstigen zusätzlich die Leistungsschere.

Für Österreich zeigte sich im naturwissenschaftlichen Bereich der PISA 2003 Analysen (Haider & Reiter, 2004), dass Schülerinnen und Schüler aus Allgemeinbildenden Höheren Schulen (AHS-Oberstufe) um 28 Punkte besser waren als Schülerinnen und Schüler der Berufsbildenden Höheren Schulen (BHS). Die Gruppe der BHS ist jedoch hinsichtlich der fachlichen Ausbildungsziele sehr heterogen. So sind neben den naturwissenschaftlich-technisch orientierten Höheren Technischen Lehranstalten (HTLs) auch höhere Schulen mit beispielsweise Mode, künstlerischer Gestaltung, Tourismus, Kindergarten- und Sozialpädagogik als Fachrichtung enthalten. Auch in den Kompetenzprofilen in PISA 2006 (Schreiner & Schwantner, 2009) zeigt sich, dass im Mittel die AHS gegenüber den BHS höhere Leistungen in den Kompetenzbereichen Naturwissenschaft, Mathematik und Lesen erzielen. Die Unterschiede jedoch zwischen den Schulen des selben Schultyps sind sehr hoch.

An einer Stichprobe von insgesamt 7393 Schülerinnen und Schüler wurden die Englischkenntnisse der 12. und 13. Jahrgangsstufe aus allgemeinbildenden und beruflichen Gymnasien sowie integrierten Gesamtschulen in den beiden Bundesländern Baden-Württemberg und Nordrhein-Westfalen getestet (Köller, Baumert, Cortina, Trautwein,

& Watermann, 2004). Die Englischkenntnisse wurden mit dem „Test of English as a Foreign Language“ (TOEFL) erfasst. Anhand der kritischen Werte, die amerikanische Universitäten als Aufnahmekriterium für ausländische Studentinnen und Studenten benutzen, wurden die Leistungen bewertet. Die Ergebnisse zeigten deutlich, dass der Großteil an Absolventinnen und Absolventen von allgemeinbildenden Gymnasien die kritischen Werte im TOEFL erreichten, wobei die anderen Schultypen deutlich geringere Leistungen erzielten.

3.3. Regionale Herkunft - Stadt/Land

In der Sozialstrukturanalyse wird die regionale Herkunft einer Person neben Geschlecht, ethnischer- und nationaler Zugehörigkeit zu den so genannten „neuen“ Ungleichheitsdimensionen gezählt während Bildung, Einkommen und Beruf der Eltern zu den klassischen Ungleichheitsdimensionen zählen (Geißler, 1990; Hradil, 1987). Im Gegensatz zu den klassischen Ungleichheitsdimensionen wird nur von einer horizontalen Wirkung in Teilbereichen ausgegangen (Geißler, 1990). Räumliche Analysen differenzieren meist zwischen ländlichem und städtischem Raum wobei der Regionsbegriff offen ist (vgl. Weishaupt, 2009). Die Daten hierfür stammen oft aus administrativen Einrichtungen der Stadtbezirke, Gemeinden und Bundesländer. Folgend wird die regionale Herkunft mit Blick auf Unterschiede in den Bildungschancen und der Bildungspartizipation für primär städtische versus ländliche Lebensumgebung anhand von empirischen Befunden thematisiert.

Bereits seit den 60er Jahren des vorigen Jahrhunderts wird am Zusammenhang zwischen Bildung und Region geforscht, wobei der Einfluss der regionalen Herkunft auf die Bildungsbeteiligung mehrfach gezeigt werden konnte (Weishaupt, 2009). Im Unterschied zur Bundesrepublik Deutschland hat in Österreich in den 1990er keine kritische Bildungsdiskussion stattgefunden (Bacher, 2003). Die Möglichkeiten an Analysen von Daten der großen internationalen Schulvergleichsstudien (Bsp. PISA) und flächen-deckenden Lernstandserhebungen halfen bei der Re-Thematisierung von Bildungsungleichheiten. So wurden im letzten Jahrzehnt die soziale und regionale Selektivität des österreichischen Schul- und Bildungssystems gut erforscht (Bacher, 2006). Einen speziellen Fokus auf regionale Unterschiede in Österreich haben Fassmann (2002) und Schmid (2004) gelegt. Weitere Studien über den Einfluss durch regionale Ungleichheiten stammen von Bacher (2003, 2005), bezüglich des Übergangs von Sekundarstufe I zur Sekundarstufe II Spielauer, Schwarz & Schmid (2002) und Schwarz, Spielauer & Städtner (2003), speziell für das Bundesland Oberösterreich Neuhofer (2007). Für den Hochschulbereich die Studien von Bauer (2005), Wroblewski & Unger (2003) und Unger & Wroblewski (2006).

Schwabe und Gumpoldsberger (2008) haben sich mit der Verteilung der Schülerinnen und Schüler auf Schultypen der Sekundarstufe (5. und 9. Schulstufe) in Österreich beschäftigt. An den Schnittstellen der Bildungswahl zeigen sich in den vergangenen Jahrzehnten ein stetiger Trend zu maturaführenden Schulen. Der Anteil der Hauptschülerinnen und -schüler ging in den letzten 25 Jahren zugunsten der AHS zurück. Abbildung 3.4 zeigt, dass die Hauptschule noch immer die 'Haupt-Schule' ist. Die Bundesländer Vorarlberg, Tirol, Salzburg und Oberösterreich haben den geringsten

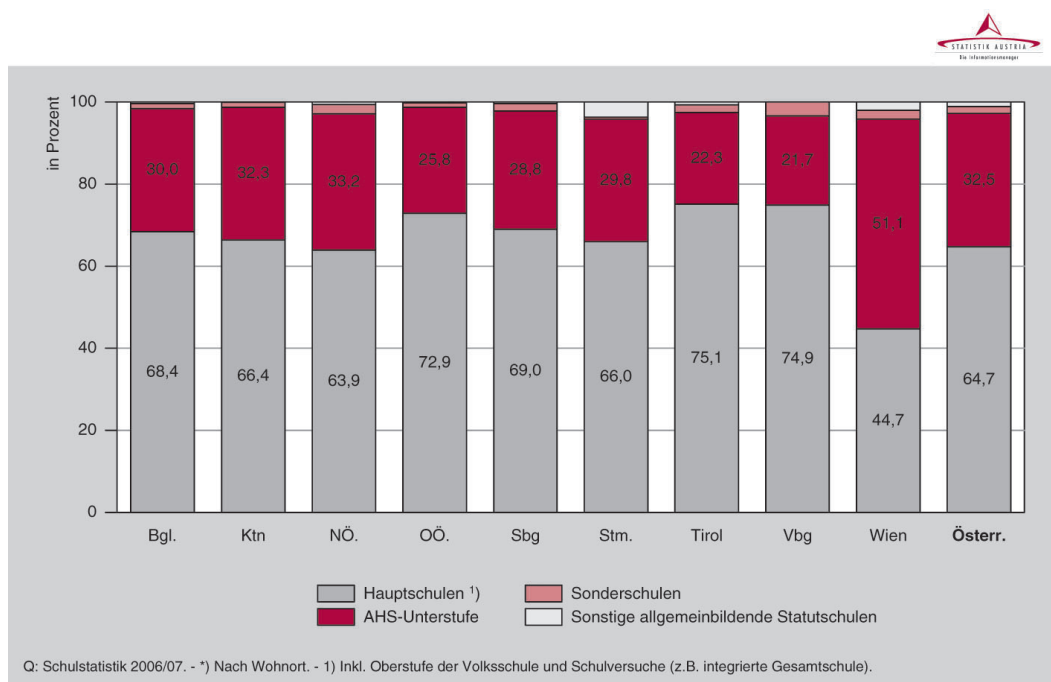


Abbildung 3.4.: Verteilung der Schüler und Schülerinnen in der 5. Schulstufe nach Bundesländern* (Schwabe & Gumpoldsberger, 2008).

Anteil an AHS-Unterstufe (Ost-West-Disparität; vgl. bei Fassmann, 2002). In Wien ist der Anteil an der AHS-Unterstufe mit 51.1% so hoch wie in keinem anderen Bundesland. Der österreichische Durchschnitt liegt bei 32.5%. Während sämtliche österreichische Bezirke mit Volks- und Hauptschulen gut ausgestattet sind (Machhold & Tamme, 2009) ist das Angebot im ländlichen Raum an AHS und BHS oft eingeschränkt und lange Schulwege sind die Folge (Fassmann, 2002; vgl. auch Schwabe & Gumpoldsberger, 2008). Mehr Bildungseinrichtungen bedeutet mehr Spielraum für Handlungsmöglichkeiten (Wagner, 1990; zitiert nach Weishaupt, 2009, S. 227). Gefordert hierfür ist die Regionalplanung und Raumordnungspolitik für die Verteilung der Bildungseinrichtungen. In der BRD ist die flächendeckende einheitliche Bildungsversorgung noch immer unzureichend verwirklicht (Weishaupt, 2009).

Die Studien kommen übereinstimmend zum Ergebnis, dass die Bildungswahl beim Übergang zur Sekundarstufe I (Hauptschule oder AHS) aufgrund der unterschiedlichen Verfügbarkeit der Schultypen (räumliche Disparität) mit der regionalen Herkunft verbunden ist. Auch in Deutschland sind räumliche Disparitäten im Schulsystem deutlich vorhanden (siehe Weishaupt, 2009). Dieser Befund ist von zentraler Bedeutung, sind doch bereits mit der Wahl der Sekundarstufe I der weitere Bildungsweg und die damit einhergehende unterschiedliche Leistungszuwächse (siehe Schereneffekt 3.2.3) festgelegt.

Je höher die Ausbildung wird, desto mehr Wahlmöglichkeiten bestanden bereits im Ausbildungsweg, die nachhaltig auch die Partizipation an den Fachhochschulen und Universitäten mitbestimmen. Die Schwankung der Zusammensetzung der Studienanfängerinnen und -anfänger nach regionaler Herkunft ist erheblich und bleibt auch über die Studienjahre für die Universitäten in größerem Ausmaß als bei den Fachhochschulen bestehen (siehe Tabelle 3.2: Unger & Wroblewski, 2006, S. 38). Über Vergleiche mit der Wohnbevölkerung aus dem Jahre 2005 zeigt sich für das Studienjahr WS 2005/06 an den Universitäten eine ungleiche Verteilung von Studienanfängerinnen und Studienanfängern aus den verschiedenen Bundesländern. Studierende aus Wien sind dabei am stärksten mit +5.4% überrepräsentiert, während Studierende aus Niederösterreich (-1.7%), Tirol (-0.9%) und Vorarlberg (-0.7%) unterrepräsentiert sind. An den Fachhochschulen ist eine Überrepräsentation von Studierenden aus den östlichen Bundesländern Wien (+2.2%), Niederösterreich (+2.2%) und Burgenland(+1.1%) zu erkennen, während die südwestlichen Bundesländer, am stärksten Tirol (-2.2%) zu wenig Studienanfängerinnen und Studienanfänger haben.

Tabelle 3.2.: Regionale Herkunft der Studienanfängerinnen und Studienanfänger im Vergleich zur Wohnbevölkerung 2005 in Österreich (Unger & Wroblewski, 2006).

	WS 1994/95			WS 2000/01			WS 2005/06			2005 Wohn- bev.
	Univ.	FH	Ges.	Univ.	FH	Ges.	Univ.	FH	Ges.	
Burgenland	3,3%	14,5%	3,7%	2,9%	3,7%	3,1%	3,1%	4,7%	3,5%	3,6%
Kärnten	7,7%	1,5%	7,5%	8,7%	7,2%	8,4%	8,0%	6,2%	7,5%	7,2%
NÖ	17,1%	27,2%	17,4%	16,6%	20,3%	17,2%	17,1%	22,0%	18,4%	19,8%
OÖ	15,5%	23,3%	15,8%	17,1%	17,2%	17,1%	15,8%	16,5%	16,0%	17,5%
Salzburg	6,2%	1,5%	6,1%	6,0%	6,0%	6,0%	6,3%	6,1%	6,2%	6,2%
Steiermark	15,8%	6,7%	15,5%	15,7%	13,9%	15,5%	15,0%	14,1%	14,7%	15,4%
Tirol	7,7%	1,8%	7,5%	7,3%	7,0%	7,3%	7,4%	5,7%	6,9%	8,3%
Vorarlberg	3,5%	5,0%	3,6%	3,3%	4,7%	3,6%	3,4%	3,5%	3,4%	4,1%
Wien	22,3%	18,2%	22,2%	22,2%	19,9%	21,8%	23,3%	20,1%	22,5%	17,9%
Summe	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Erstmals zugelassene inländische, ordentliche Studierende im Vergleich zur inländischen Wohnbevölkerung.
Siehe Anhangtabelle 33, Anhangtabelle 35 und Anhangtabelle 37.
Quelle: Statistik Austria, BMBWK, Mikrozensus 2005.

Die ländliche Herkunft kann eindeutig als Benachteiligungsfaktor für das Ausschöpfen von Bildungsqualität angesehen werden (Unger & Wroblewski, 2006). Personen, die in größeren Städten aufwachsen, haben eine bessere Ausbildung (Wagner, 1989; zitiert nach Weishaupt, 2009, S. 227). Die Kombination der Determinanten wie beispielsweise sozialer Hintergrund, regionale Herkunft, Migrationshintergrund, Geschlecht, Einkommen und Beruf der Eltern legt die entscheidenden Weichen der Ausbildung und sind nach Bacher (2005) immer noch wirksam. Frauen aus ländlichen Re-

gionen sind besonders benachteiligt (Unger & Wroblewski, 2006). Die Bildungsbenachteiligung durch eine ländliche Herkunft zeigt sich (selbst unter Berücksichtigung des familiären Hintergrunds, Geschlechts und des Studienfachs) auch in Kanada und den USA wobei in den USA aufgrund der städtischen Struktur die vorstädtische Bevölkerung zu den Bildungsgewinnern zählt (vgl. hierzu Andres & Looker, 2001; Roscigno, Tomaskovic-Devey, & Crowley, 2006).

Während sich der Einfluss der Region auf die Bildungspartizipation deutlich zeigt, sind die Befunde hinsichtlich des regionalen Einflusses auf die Leistung unterschiedlich (vgl. Young, 1998). Young (1998) vermutet, dass der Einfluss von anderen Indikatoren (bspw. sozioökonomischer Status und/oder Ethnizität) nicht ausreichend kontrolliert wurden. Zudem gibt es nur sehr wenige empirische Studien, die direkt den Einfluss der regionalen Herkunft auf Schulleistungen oder Leistungen in Studierfähigkeitstests messen.

Die Unterschiedlichkeit der empirischen Befunde soll kurz exemplarisch an den folgenden drei empirischen Studien verdeutlicht werden. Young (1998) fand in Australien bessere Leistungen in Mathematik/Wissenschaft bei Schülern aus der städtischen Region. Eine kanadische Studie (Randhawa & Hunt, 1987) zeigt eine deutliche Überlegenheit in mathematischen Fertigkeiten der Schüler aus ländlichen Regionen. In einer amerikanischen Studie fanden Borland und Howsen (1999) eine Überlegenheit in der Leistung von Studenten aus Regionen mit mittlerer Dichte gegenüber Studenten aus Regionen mit sehr hoher und sehr niedriger Dichte. Für die USA ist die Überlegenheit der vorstädtischen Region eindeutiger und bereits mehrfach gezeigt worden (vgl. Cakir, Deliaglioglu, Dennis & Duffy, 2009).

3.4. Zusammenfassung

Zur Kontrolle und Sicherung der Qualität von Auswahlverfahren, dazu zählt auch die Fairness, ist es notwendig Leistungsunterschiede im Testverfahren (Studierfähigkeitstests) genauer zu analysieren. Gründe hierfür sind einerseits mögliche Diskriminierungen von bestimmten Gruppen. Andererseits prüfen Studierfähigkeitstests bestimmte Fähigkeiten und Kompetenzen der Bewerberinnen und Bewerber, die für das Studium zentral sein sollen. Hierbei stellt sich die zentrale Frage wann und wo diese Fähigkeiten und Kompetenzen erworben wurden. Ziel der Analysen für die Kontrolle und Sicherung der Qualität der Testverfahren ist, die Testleistungen anhand von geeigneten Faktoren (sogenannten Prädiktoren) vorherzusagen bzw. die Leistungsunterschiede zwischen den Personen aufzuklären.

In diesem Kapitel wurden hierzu eine Auswahl von drei wichtigen Faktoren vorgestellt, die mit dem Erwerb dieser Fähigkeiten und Kompetenzen in der Literatur in engem Zusammenhang gebracht werden. Als wichtigste Einflussgröße ist hierbei die Institution Schule zu erwähnen, die die Fähigkeiten und Kompetenzen vermitteln und fördern soll. Konkret wurde auf die Prädiktoren Schulnoten (in Form von Ziffernoten) und der besuchte Schultyp im Detail eingegangen. Auch der mit dem besuchten Schultyp eng in Verbindung stehende Prädiktor 'regionale Herkunft' wurde genauer dargestellt.

Zusammenfassend kann zu den Schulnoten gesagt werden, dass sie deutliche Mängel in der Güte Schulleistungen abzubilden aufweisen, jedoch besser sind als ihr Ruf (vgl. Tent, 2006). Aufgrund ihrer vielfältigen Funktionen und der nachhaltigen Auswirkungen für den Einzelnen und die Gesellschaft wäre eine bessere Qualität wünschenswert. Trotz ihrer Mängel speziell im Gütekriterium Objektivität ist die Prognosegüte von Schulleistungen für Testleistungen in objektiven Leistungstests und Studierfähigkeitstests relativ hoch. Speziell (Zeugnis-)durchschnittsnoten sind sehr gute Prädiktoren, um die Testleistungen vorherzusagen. Von den Einzelfachnoten erweist sich die Mathematiknote als besonders valide.

Die Institution Schule ist aber nicht als ein einheitliches System zu sehen. Es gibt unterschiedliche Schultypen, die zur Matura und damit zur Studienberechtigung führen. Diese unterschiedlichen Schultypen haben unterschiedliche Ausbildungsziele, die sich auch in den Lehrplänen widerspiegeln. So werden durch die Auswahl von bestimmten Fächern sowie deren Quantität (Wochenstudenzahl) unterschiedliche (fachspezifische) Fähigkeiten und Kompetenzen vermittelt und gefördert. In Österreich zeigen sich diese Unterschiede speziell zwischen den Allgemeinbildenden höheren Schulen (AHS) und der Berufsbildenden höheren Schulen (BHS). Je nach geforderten Fähigkeiten und Kompe-

tenzen sind Absolventinnen und Absolventen in Studierfähigkeitstest im Rahmen von Auswahlverfahren aus den verschiedenen Schultypen kommend unterschiedlich erfolgreich. In einen Test der vorwiegend mathematisch-naturwissenschaftliche Fähigkeiten prüft, während hier Absolventinnen und Absolventen speziell aus HTLs mit einschlägigem Schwerpunkt wahrscheinlich erfolgreicher wären.

Der Einfluss der regionalen Herkunft zeigt sich speziell in der Wahl der Schule. Die Schulwahl ist primär durch das regionale Schulangebot beeinflusst. Neben der erbrachten Schulleistung, die Einfluss auf die Schulwahl in den Sekundarstufen hat, kommt zusätzlich die mangelnde Verfügbarkeit bestimmter Schultypen hinzu. So fällt die Schulwahl in der Sekundarstufe I in den ländlichen Gebieten primär auf die Hauptschule und seltener auf das Gymnasium. Auch in der Sekundarstufe II und an Universitäten sind gemessen an der Wohnbevölkerung zu wenige Personen aus ländlichen Regionen vertreten. Die ländliche Herkunft kann somit eindeutig als Benachteiligungsfaktor für das Ausschöpfen von Bildungsqualität gesehen werden. Die Befunde bezüglich dem direkten Einfluss von regionaler Herkunft auf Leistungstests sind nicht eindeutig. Zudem gibt es auch nur wenig empirische Studien, die den Einfluss der Region auf die Leistung untersuchen. Um den Einfluss der regionalen Herkunft eindeutiger fassen und verstehen zu können, wären noch mehrere Forschung, die diesen Sachverhalt klären können notwendig.

4. Fragestellungen und Hypothesen

Seit dem Studienjahr 2006/07 wird an den Medizinischen Universitäten in Wien und Innsbruck ein Auswahlverfahren zur Zulassung zum Studium der Humanmedizin durchgeführt. Kern dieses Auswahlverfahrens ist der 'Eignungstest für das Medizinstudium' (EMS). Ziel des EMS ist es die geeignetsten Kandidatinnen und Kandidaten für das Medizinstudium auszuwählen. Die Auswahlentscheidung durch den EMS hat individuelle Konsequenzen für die Bewerberinnen und Bewerber aber auch Konsequenzen für die Gesellschaft (vgl. hierzu Kapitel 2). Die individuellen Konsequenzen hierbei sind, dass das gewünschte Studienfach im aktuellen Semester nicht studiert werden darf und die Wartezeit bis zum nächsten Antritt (sinnvoll) überbrückt werden muss. Auch kann die negative Rückmeldung zur Wahl eines anderen, nicht primär favorisierten Studienfachs führen, oder gänzlich auf ein Studium zugunsten anderer Tätigkeiten verzichtet werden. Aufgrund der Systematik der Auswahlentscheidung über mehrere Jahre können Fehlentscheidungen beträchtliche Konsequenzen für die Gesellschaft nach sich ziehen. Demnach ist es absolut notwendig, dass der Test auch tatsächlich für den Studienerfolg relevante Fähigkeiten und Kompetenzen prüft. Zur Sicherung und Kontrolle der Qualität des Verfahrens ist es daher notwendig das Auswahlverfahren regelmäßig zu evaluieren und zu optimieren. Ein sehr wichtiges Qualitätskriterium ist die Fairness. Fairness bedeutet, dass keinen Gruppen durch den Test systematisch benachteiligt werden.

Die für den Studienerfolg relevanten Fähigkeiten sollen mit dem EMS gemessen werden. Dabei sollen aber keine Fähigkeiten geprüft werden, die erst im Studium erworben werden. Eine wichtige Frage hierbei ist, wo und wann diese geforderten Kompetenzen erlernt werden. Aufgrund unterschiedlicher Bildungsverläufe der Bewerberinnen und Bewerber kann von unterschiedlichen Ausprägungen der geforderten Fähigkeiten ausgegangen werden (vgl. hierzu Kapitel 3). Zur Beurteilung der Qualität des EMS allgemein sowie speziell zur Prüfung der Fairness, ist es daher wichtig die Einflussgrößen auf die Testleistungen differenziert zu betrachten. Für den Erwerb von fach- und bildungsspezifischen Fähigkeiten und Kompetenzen kommen schulischen Aspekten wie Schulnoten und absolvierter Schultyp die größte Bedeutung zu. Da die Schulwahl durch die regionale Herkunft beeinflusst wird, könnte sie einen Beitrag zur Erklärung der Leistungen im EMS beitragen. Nach den vorliegenden empirischen Studien zum EMS (vgl. Spiel et al., 2008) zeigen sich deutliche Unterschiede in den Prädiktoren Geschlecht und Nationalität. In dieser empirischen Arbeit soll daher folgend der Fragestellung nachgegangen werden, ob *schulische Aspekte* (Schulnoten und Schultyp), die *regionale*

Herkunft sowie *Geschlecht* und *Nationalität* die *Leistungen im EMS* erklären können.

Die Qualität von Auswahlverfahren kann anhand der eignungsdiagnostischen Gütekriterien bewertet werden. Diesen wissenschaftlich fundierten und für die Eignungsdiagnostik allgemein akzeptierten Kriterien (vgl. DIN 33430) hat ein gutes Verfahren zu genügen (vgl. Unterabschnitt 2.2.2).

Speziell zur Prüfung der Fairness des Verrechnungsmodus haben Spiel et al. (2008) testtheoretische Analysen nach dem Rasch-Modell durchgeführt. Wie die Autorinnen mittels Rasch-Analysen zeigen konnten, sind eine Reihe von Aufgaben der einzelnen Untertests gegenüber dem Testwert (dem sogenannten 'inneren Teilungskriterium', vgl. Kubinger, 2006) nicht verrechnungsfair. Inhaltlich bedeutet dies, dass die Aufgaben nur unzureichend zwischen 'besseren' und 'schlechteren' Personen differenzieren können. Eine eindeutige Aussage über die Fähigkeit einer Person ist anhand dieser Aufgaben nicht möglich. Die Verrechnungsfairness im Sinne eines Gütekriteriums ist dadurch verletzt (Kubinger, 2006). Auch hinsichtlich der Teilungskriterien *Geschlecht* und *Nationalität* erwies sich eine Vielzahl von Items als nicht Rasch-Modell-konform. Die übliche Vorgehensweise ist es diese nicht modellkonformen Aufgaben auszuschließen (vgl. Spiel et al., 2008; Kubinger, 2006). Dies erhöht die Testqualität, da einerseits nur Aufgaben einer Dimension miteinander verrechnet werden und andererseits Aufgaben, die bestimmte Gruppen benachteiligen, ausgeschlossen werden.

Die Eindimensionalität eines Untertests ist eine wichtige Voraussetzung für die Interpretation des Messergebnisses. Wenn das zu messende Konstrukt 'sauber' gemessen wird, das heißt störende Einflussgrößen ausgeschaltet oder minimiert werden, dann lassen sich auch reliablere und validere Zusammenhänge mit anderen empirisch erfassbaren Einflussgrößen finden. Wenn störende Faktoren das Messergebnis beeinflussen, so können tatsächlich vorhandene Effekte von bestimmten und grundsätzlich geeigneten Einflussgrößen nicht oder nur schwerer in empirischen Untersuchungen gefunden werden (vgl. Bortz & Döring, 2006). Dies gilt analog auf für die Prädiktoren, die in dieser Studie untersucht werden sollen (siehe oben). Durch den Ausschluss der nicht verrechnungsfaireren Aufgaben könnten mögliche Störfaktoren minimiert werden. Der Beitrag der Prädiktoren zur Aufklärung der Leistungsunterschiede zwischen den Personen könnte dadurch verbessert werden. Zudem könnte die Effizienz des Verfahrens durch eine höhere Treffersicherheit gesteigert werden. Die Anzahl der Geeigneten unter den Ausgewählten würde sich erhöhen und gleichzeitig die Anzahl der Geeigneten unter den Nicht-Ausgewählten verringern. *Es wird daher der Frage nachgegangen, ob der Ausschluss von nicht verrechnungsfaireren Aufgaben die Prädiktorleistung der Einflussgrößen Schulnoten, Schultyp, regionale Herkunft, Geschlecht und Nationalität erhöht?*

Im Folgenden werden die Fragestellungen zu den einzelnen Prädiktoren genauer beschrieben sowie die Richtung der Hypothesen festgelegt.

4.1. Schulische Aspekte

Es werden hierfür zwei unterschiedliche schulische Aspekte als Einflussgrößen in die Analyse einfließen. Zum einen sind dies die schulischen Leistungen in Form von Schulnoten, zum anderen Einflüsse, die auf den besuchten Schultyp zurückzuführen sind.

4.1.1. Schulnoten

Die *Schulnoten* gelten als beste Prädiktoren für den Studienerfolg (vgl. Tabelle 2.2). Sie sind zudem auch Merkmale der Leistung wie das Testergebnis des EMS. Weiter sind sie ein Indikator für Sozialisation und Bildung. Der EMS prüft laut Autoren komplexe und längjährig erworbene Lern- und Entwicklungsprozesse (Hänsgen & Spicher, 2006). Diese komplexen Prozesse sollten sich durch die Schulbildung im Laufe der Jahre geformt haben. Die Schule die wichtigste Institution an der fach- und bildungsspezifische Kompetenzen vermittelt werden. Die dort erlernten Fähigkeiten und Kompetenzen werden primär durch die Schulnoten abgebildet. Die Autoren betonen, dass die Schulleistungen und der EMS auf einen gemeinsamen Fähigkeitsfaktor zurückzuführen seien (Hänsgen & Spicher, 2004). *Konkret soll daher die Frage untersucht werden, ob die Schulnoten die Leistungen im EMS erklären bzw. vorhersagen können?*

Spiel et al. (2008) fanden bereits zu den Schulnoten konträre Testleistungen; die österreichischen Frauen hatten im Mittel (Durchschnittsnote der naturwissenschaftlichen Fächer) die besten Schulnoten, waren jedoch im Mittel die schlechtesten im EMS (im Vergleich zu den Gruppen deutsche und österreichische Männer und deutsche Frauen). Grundsätzlich wäre zu erwarten, dass bessere schulische Leistungen auch zu besseren Leistungen im EMS führen. Dieser Befund zeigte sich auch, jedoch nicht so deutlich in der Durchschnittsnote aus den beiden Fächern Deutsch und Englisch.

Wie auch die Literatur zeigt, sind speziell Zeugnisdurchschnittsnoten die besten Prädiktoren für die Leistung in objektiven Leistungstests (vgl. Hodge & Coladarci, 1989). Grundsätzlich gilt, dass mehrere Leistungsbeurteilungen eine höhere Prognosegüte gegenüber einzelnen Beurteilungen aufweisen. Sehr häufig wird in Zulassungsverfahren zum Hochschulstudium die Durchschnittsnote aller Fächer des Matura-/Abiturzeugnisses zur Eignungsfeststellung herangezogen. Es scheint daher auch für diese Studie angebracht eine Durchschnittsnote aus allen Fächern des Matura-/Abiturzeugnisses zu bilden und in die Analysen mit aufzunehmen.

Unter den Einzelfachnoten zeigt sich der Literatur nach Mathematik als guter Prädiktor für den Studienerfolg (vgl. Trapman et al., 2007). Wenn die Leistungen des EMS mit dem Studienerfolg in Zusammenhang stehen, so müsste die Mathematiknote auch als Prädiktor für die Testleistung geeignet sein. Für eine differenzierte Betrachtung ist nicht nur der Einfluss der Durchschnittsnote, sondern auch der Einfluss der einzelnen Fachnoten von Interesse. Die Einzelfachnoten können als Maß für die fachspezifischen Fähigkeiten einer Person angesehen werden. Wenn die Untertests des EMS inhaltlich mit einigen fachspezifischen Fähigkeiten in Zusammenhang stehen (bspw. mathematische und/oder naturwissenschaftliche Inhalte), dann sollte dies in gewissem Grade durch die Schulnoten auch abgebildet werden können. Diese Studie soll daher auch die Frage klären, wie gut die Einzelfachnoten die Leistungen im EMS erklären und vorhersagen können.

Für eine differenziertere Betrachtung erscheint es sinnvoll den Einfluss des *'Notenprofils'* einer Person zu untersuchen. Die Einzelfachnoten einer Person stellen statistisch betrachtet eine Häufigkeitsverteilung dar. Jede Häufigkeitsverteilung besitzt sogenannte Lagemaße. Die Durchschnittsnote bildet hierbei nur den zentralen Wert der Verteilung ab. Zu jeder Verteilung gehört aber nicht nur ein Zentralwert, sondern auch ein Abweichungs- bzw. Streumaß (vgl. bei Hartung, 2005). Das Notenprofil wird somit über den zentralen Wert der Verteilung sowie über das zugehörige Abweichungsmaß beschrieben. Das Abweichungsmaß kann inhaltlich als Homogenitätsmaß aufgefasst werden. Je kleiner die Streuung der Noten innerhalb einer Person ist, desto größer ist die Homogenität des Notenprofils. Dies soll folgend anhand eines einfachen Beispiels gezeigt werden. Der selbe Notendurchschnitt kann aus unterschiedlichen Noten gebildet werden. Beispielsweise für eine Durchschnittsnote von 2 kann diese nur aus Zweiern resultieren oder auch aus vier Einser und zwei Vierer. Das bedeutet, dass durch den Einbezug des Abweichungsmaßes zusätzliche Information über die Schulleistungen gewonnen werden kann.

Es macht somit einen Unterschied aus welchen Schulnoten die Durchschnittsnote gebildet wird. Je einheitlicher (homogener) die Schulnoten sind, desto stabiler ist die schulische Leistung zu bewerten. Je inhomogener das Notenprofil ist, desto mehr spricht für eine Annahme, dass fachspezifische Fähigkeiten unterschiedlich stark ausgeprägt sind. Die einzelnen Untertests im EMS zielen auf unterschiedliche Fähigkeiten der Kandidatinnen und Kandidaten ab. Eine unterschiedliche Ausprägung der fachspezifischen Fähigkeiten der Kandidatinnen und Kandidaten könnte eine Erklärung für die Leistungsunterschiede im EMS sein. *Es soll somit der Frage nachgegangen werden, ob das Notenprofil, d.h. das Ausmaß der Homogenität der Schulnoten die Vorhersage der Leistungen im EMS verbessern kann?*

Die Bedeutung des Zusammenspiels von der Durchschnittsnote und dem Abweichungsmaß für die im EMS-Test erbrachte Leistung ist vorab nicht ganz klar zu sehen. Entsprechend den oben formulierten Hypothesen ist zu erwarten, dass Personen mit einem absolut homogenen Notenprofil und der besten Durchschnittsnote auch die höchsten Testleistungen zeigen. Es stellt sich jedoch die Frage, wie sich eine gute Durchschnittsnote bei einem inhomogenen Notenprofil oder eine schlechtere Durchschnittsnote bei einem sehr homogenen Notenprofil auf die Testleistungen auswirkt. Es gibt vorab keine Information darüber, ob es sich hierbei um einen linearen Zusammenhang handelt. Speziell könnte es sein, dass sich Personen mit nur Einsern von den anderen Personen in ihren Fähigkeiten qualitativ unterscheiden oder dass die Skalierung in Schulnoten (bspw. durch einen Deckeneffekt) die Fähigkeiten nicht adäquat abbilden kann. Aus dieser Überlegung heraus ist empirisch zu prüfen, ob ein Interaktionseffekt zwischen den beiden Maßen des Notenprofils vorliegt.

4.1.2. Schultyp

Die Institution Schule ist nicht als ein einheitliches System zu sehen. Die verschiedenen Schultypen haben unterschiedliche Ausbildungsziele und vermitteln dadurch unterschiedliche (fachspezifische) Fähigkeiten und Kompetenzen. Dies legt die Vermutung nahe, dass der besuchte Schultyp neben den schulischen Leistungen einen Einfluss auf die Testleistung hat.

Der EMS enthält eine Reihe von Untertests die auf mathematische und naturwissenschaftliche Fähigkeiten abzielen (Spiel et al., 2008). Der Untertest 'Quantitative und formale Probleme' zeigt in seinen Aufgaben Ähnlichkeiten zum Mathematikunterricht. Andere Untertests wie 'Textverständnis', 'med.-naturwissenschaftliches Grundverständnis' und 'Diagramme und Tabellen' zielen auf naturwissenschaftlich-medizinische Fähigkeiten ab. In der Gesamtleistung des EMS konnten signifikante Unterschiede zwischen naturwissenschaftlich und nicht naturwissenschaftlich orientierten Schultypen bei deutschen Bewerberinnen und Bewerbern gefunden werden (Spiel et al., 2008). Für die hier angestrebte differenziertere Betrachtung ist es sinnvoll den Schultyp in feinere Kategorien zu unterteilen. Dadurch werden Schultypen, die doch im Detail unterschiedliche Ausrichtungen und Schwerpunkte haben, nicht zu einer gemeinsamen Gruppe zusammengefasst. Diese Vorgehensweise ermöglicht eine konkrete Aussage, ob sich Personen aufgrund ihres absolvierten Schultyps in der Leistung des EMS unterscheiden.

In Österreich zeigen sich Unterschiede in den Lehrplänen und im Absolventinnen- und Absolventenprofil speziell zwischen den Schultypen Allgemeinbildene höhere Schu-

le (AHS) und Berufsbildende höhere Schule (BHS). Im deutschen Bildungssystem gibt es die Berufsbildenden höheren Schulen wie sie in Österreich Tradition haben nicht. Wie bereits erwähnt, kann der EMS als Test mit teilweiser naturwissenschaftlich-mathematischer Ausrichtung beurteilt werden. Dies lässt erwarten, dass Personen aus Schultypen mit mathematisch- naturwissenschaftlichen Schwerpunkt oder Absolventinnen und Absolventen aus facheinschlägigen Höheren Technischen Lehranstalten (HTL) gegenüber Personen aus anderen Schultypen bessere Testleistungen zeigen. *Daraus ergeben sich für eine österreichische Stichprobe die Fragen (1) ob der absolvierte Schultyp einen Einfluss auf die Testleistung im EMS hat und (2) ob HTL-Absolventinnen und -Absolventen im EMS erfolgreicher sind, als Personen aus anderen Schultypen.*

4.2. Regionale Herkunft

Wie aus der Literatur gezeigt werden konnte (siehe Abschnitt 3.3) kann die Herkunft aus einer ländlichen Region als Benachteiligungsfaktor angesehen werden. Der Einfluss der regionalen Herkunft zeigt sich speziell in der Wahl des Schultyps (vgl. Bacher, 2003, 2005). Als beeinflussender Faktor ist die regionale Verfügbarkeit der einzelnen Schultypen zu sehen. Nach Unger und Wroblewski (2006) sind Frauen aus ländlichen Regionen besonders benachteiligt. Sollte sich dies auch auf die Leistung auswirken so würde sich ein Interaktionseffekt zeigen.

Empirische Befunde über den Einfluss von regionaler Herkunft auf die Leistung sind nicht eindeutig. Zudem gibt es auch nur wenig empirische Studien, die den Einfluss der Region auf die Leistung untersucht haben. Um den Einfluss der regionalen Herkunft eindeutiger fassen und verstehen zu können, wären noch mehrere Forschung, die diesen Sachverhalt klären können notwendig. *Es wird daher der Fragestellung nachgegangen, ob die regionale Herkunft einen Einfluss auf die Testleistungen des EMS hat?* Aufgrund der nicht eindeutigen Ergebnisse der wissenschaftlichen Literatur wird die Hypothese zweiseitig formuliert.

4.3. Geschlecht und Nationalität

Einen bedeutenden Einfluss von Geschlecht und Nationalität (Österreich/Deutschland) auf die Gesamtleistung im EMS konnten bereits Spiel et al. (2008) zeigen. Frauen schnitten deutlich schlechter ab als Männern und österreichische Kandidatinnen und Kandidaten schlechter als deutsche. Dieser Befund ist nicht unerwartet basiert doch der EMS auf dem 'Test für medizinische Studiengänge (TMS)' indem bereits mehrfach Geschlechtsunterschiede beobachtet werden konnten (vgl. Trost et al., 1998).

Für die Sicherung und Kontrolle der Fairness des EMS ist es notwendig Analysen auf mögliche systematische Benachteiligungen von bestimmten Gruppen durchzuführen. Eine systematische Benachteiligung bestimmten Gruppen (z.B. nach Geschlecht oder Nationalität) gilt als unfair. Dies bedeutet aber nicht, dass es keine tatsächlichen Unterschiede in der Leistung geben kann. Auch wenn das Verfahren selbst die Personen nicht benachteiligt, so besteht dennoch ein bildungswissenschaftliches Interesse die Einflussgrößen auf die Leistung aufzudecken und Erklärungen für deren Wirksamkeit zu finden. Den empirisch gefundenen Geschlechtseffekt erklären die Autorinnen damit, dass Bewertungen der Schulleistungen von Mädchen und Knaben etwas qualitativ unterschiedliches sind. Dadurch ist die Chancengleichheit der Frauen aufgrund der Sozialisationsvoraussetzungen deutlich schlechter als die der Männer.

Im Faktor Nationalität fanden sie einen deutlichen Effekt zwischen österreichischen und deutschen Kandidatinnen und Kandidaten. Zusätzlich fanden Spiel und Kolleginnen (2008) eine signifikante Interaktion zwischen Geschlecht und Nationalität. Die Geschlechtsunterschiede bei den Österreicherinnen und Österreicher sind etwas deutlicher als bei den Deutschen. Ausgehend von diesen empirischen Ergebnissen werden Geschlecht und Nationalität weiter als Prädiktoren in die Analysen aufgenommen. Die formulierten Hypothesen lauten: (1) Männer erzielen höhere Leistungen als Frauen und (2) deutsche Bewerberinnen und Bewerber höhere Leistungen als österreichische.

Spiel et al. (2008) fanden auch einen sehr kleinen Interaktionseffekt zwischen Geschlecht und Nationalität. Dieser zeigt, dass Geschlechtsunterschiede bei den Österreicherinnen und Österreichern deutlicher sind als bei den Deutschen. Dieser Interaktionseffekt wird auch in der hier vorliegenden Studie geprüft.

5. Methode

Die hier vorliegende Arbeit basiert auf den Daten der Evaluationsstudie von Spiel et al. (2008) über das Auswahlverfahren an den Medizinischen Universitäten, das am 06. Juli 2007 durchgeführt wurde. Im Folgenden werden die genauen Umstände der Durchführung der Untersuchung dargestellt und Informationen zur Stichprobe gegeben. Die Darstellung der eingesetzten Erhebungsinstrumente beschränkt sich auf die in meiner Arbeit relevanten Teile. Im Anschluss folgt eine Beschreibung der Vorgehensweise bei der Auswertung sowie der eingesetzten Auswertungsverfahren.

5.1. Untersuchungsdurchführung/Design

Die Evaluationsstudie wurde als eine Kombination von Testergebnissen aus den Eignungstests, soziodemographischen Daten über die Bewerberinnen und Bewerber sowie einer Fragebogenerhebung konzipiert. Die Fragebogenerhebung erfolgte im Anschluss an die Testung am 06. Juli 2007 an den Medizinischen Universitäten. Die Durchführung wurde von den Medizinischen Universitäten organisiert. Die vorausgegangene Testdauer des EMS betrug hierbei in zwei Teilen (Vormittag und Nachmittag; siehe 2.4) bereits insgesamt 315 Minuten (=5.25 Stunden). Für die Vorgabe des für die Beantwortung der Fragestellungen notwendigen Fragebogens (zur Erfassung der theoretisch begründeten Erklärungsvariablen), mussten Einschränkungen in der Konzeption sowie in der Vorgabe berücksichtigt werden (Spiel et al., 2008, S. 80):

Aufgrund des vorausgegangenen, sehr zeitintensiven Auswahlverfahrens und der knappen Ressourcen im Organisationsplan durfte der Fragebogen nicht mehr als 5 Minuten in Anspruch nehmen. Die zeitlich knappe Bemessung der Befragung ermöglichte auch keine offenen Fragen zur Beantwortung der Fragestellungen. Der Fragebogen wurde daher vollständig in der Form von Auswahl durch Ankreuzen vorgegeben (siehe 5.2). Weiters war zur Bewältigung der sehr großen Stichprobe nur ein automatisiertes Erfassen der ausgefüllten Fragebögen möglich. Bei der Gestaltung des Fragebogens musste daher die Kompatibilität mit den Anforderungen des automatischen Beleglesers berücksichtigt werden. Die Befragung der Testteilnehmerinnen und Testteilnehmer fand erst nach dem EMS statt, wobei aus Sicht der Evaluatorinnen eine Befragung *vor der Testung* für die Beantwortung einiger Fragestellungen sinnvoller gewesen wäre, da der „fraglos sehr anstrengende und für viele eher frustrierende Test“ die Ergebnisse beeinflusste (S. 87).

Obwohl es eine schriftliche Instruktion gab, kam es zu unterschiedlicher „Ausfüll-

qualität“ in den einzelnen Sektoren (die Testung wurde nicht in einem einzigen Raum durchgeführt, sondern war auf mehrere Sektoren parallel aufgeteilt, in denen unterschiedliche Testleiterinnen und Testleiter tätig waren). Eine Heterogenität in der Instruktion führte zu einer schlechteren Datenqualität durch „unsauberes Ankreuzen, falsches Kleben der Codeetiketten, häufig durchgestrichene Antworten etc.“ (S. 86). Neben dem knappen Zeitbudget für die Datenerhebung gab es auch Limitierungen im Zeitfenster für die Vorbereitung der Studie.

Die Ergebnisse des Auswahltests (Gesamttestleitung und alle Untertestleistungen) sowie einige Angaben zu den Bewerberinnen und Bewerbern wurden von den Medizinischen Universitäten in Form von Dateien anonymisiert (nur EMS-Code) zur Verfügung gestellt. Als ein wesentlicher Aspekt in der Evaluation sei angemerkt, dass den Evaluatorinnen kein Einblick in die Fragen der Tests ermöglicht wurde bzw. im Falle des EMS sie diese Informationen nicht für die Evaluation verwenden durften (S. 80). Unter diesen Bedingungen wurde auf die Einsicht verzichtet.

5.2. Erhebungsinstrumente

5.2.1. Eignungstest für das Medizinstudium - EMS

Der EMS ist ein Speed-and-power-Test bestehend aus 10 Untertests (insgesamt 198 Aufgaben). Er wird in einer festgelegten Abfolge (siehe Anhang A.1.1; Hänsgen & Spicher, 2007) an zwei Halbtagen mit Mittagspause vorgegeben. Die Netto-Testbearbeitungszeit beträgt vormittags 175 Minuten, nachmittags 140 Minuten. Neun Untertests bestehen aus 178 Aufgaben. Bis auf den Untertest 'Textverständnis' (18 Aufgaben) enthalten die anderen Untertests je 20 Aufgaben. Es gibt zwei Typen von Aufgaben: (1) bestehend aus einem Fragetext, der richtigen Antwort und 3-4 falschen Antwortalternativen, (2) so genannte 'Kombinationsaufgaben' bestehend aus einem Fragetext, der richtigen Antwort und 4-5 falschen Antwortmustern. Zu jeder Aufgabe gibt es vier bis fünf Antwortmöglichkeiten, jedoch nur jeweils eine Antwort ist im Sinne der Aufgabenstellung richtig. Im Untertest 'Konzentriertes und sorgfältiges Arbeiten' (Testdauer 8 Minuten) sind aus 1600 Zeichen nach Vorgabe nur ganz Bestimmte mittels Durchstreichtechnik zu markieren.

Innerhalb jedes Untertests steigt die Aufgabenschwierigkeit an. Für jedes richtig gelöste Item wird ein Punkt verrechnet. Nicht richtig gelöste Aufgaben werden mit Null Punkten bewertet. Es gibt keine Minuspunkte. Aus dem Testcore (= Gesamtpunktzahl) wird der standardisierte Testwert in Z-Werten ($\mu = 100, \sigma = 10$) errechnet. Die Auswertung erfolgt für beide Medizinischen Universitäten (Wien und Innsbruck) ge-

meinsam am Zentrum für Testentwicklung und Diagnostik an der Universität Fribourg.

In Tabelle 5.1 sind die Fähigkeiten auf welche die einzelnen Untertests abzielen kurz zusammengefasst (Hänsgen & Spicher, 2007).

Tabelle 5.1.: Zu messende Fähigkeiten auf welche die Untertests des EMS abzielen.

Untertest	zu messende Fähigkeit
Quantitative und formale Probleme	Umgang mit Zahlen, Einheiten, Formeln im Rahmen med./naturwiss. Fragestellungen
Schlauchfiguren	räumliches Vorstellungsvermögen
Textverständnis	umfangreiches und komplexes Textmaterial aufzunehmen und zu verarbeiten
Planen und Organisieren	effiziente Selbstorganisation
Konzentriertes und sorgfältiges Arbeiten	rasches, sorgfältiges und konzentriertes Arbeiten
Med.-naturwiss. Grundverständnis	Verständnis für Fragen aus Med. und Naturwiss.
Figuren Lernen	Einzelheiten von Gegenständen wahrnehmen, einprägen und merken
Fakten Lernen	Fakten einprägen und merken können
Muster zuordnen	Ausschnitte eines komplexen Bildes wieder zu erkennen
Diagramme und Tabellen	Diagramme und Tabellen richtig analysieren und interpretieren

5.2.2. Fragebogen

Der in der Evaluation von Spiel et al. (2008) eingesetzte Fragebogen umfasste insgesamt 36 Fragen auf 2 DIN A4 Seiten (siehe die komplette Darstellung des Erhebungsinstruments im Anhang A.3) sowie die Erfassung des „Testcodes“, der die Zuordnung der Bewerberin bzw. des Bewerbers zu den Testergebnissen im Auswahlverfahren ermöglichte. Aufgrund der in 5.1 dargestellten Vorgaben und Einschränkungen konnte kein bereits vorhandener standardisierter Fragebogen vorgegeben werden, sondern musste eigens für die Untersuchung ein Fragebogen mit den interessierenden Fragenbereichen konstruiert werden. Soziodemographische Daten der Kandidatinnen und Kandidaten wie Geschlecht, Alter/Geburtsjahr, Nationalität/Staatsbürgerschaft und Maturaland (=Land in dem die Matura/Abitur abgelegt wurde) wurden bereits von den Medizinischen Universitäten festgestellt und deshalb nicht nochmals im Fragebogen erhoben. Folgend werden nur jene Teile des Fragebogens in Text und Abbildungen dargestellt, die für die statistische Auswertung in dieser Arbeit herangezogen wurden.

Instruktion

Am oberen Bereich der ersten Seite des Fragebogens (siehe Abbildung 5.1) wurden die Kandidatinnen und Kandidaten des Auswahlverfahrens zusätzlich zur mündlichen Instruktion, die sie von den jeweiligen Testleiterinnen bzw. Testleitern erhalten haben (siehe 5.1), schriftlich informiert über: (1) das Ziel das Auswahlverfahren genauer zu analysieren und daraus Optimierungsmöglichkeiten abzuleiten, (2) die Anonymität der Befragung und die Unabhängigkeit der Befragung von einer Aufnahme für das Medizinstudium sowie (3) die Durchführung der Evaluation von einer externen Arbeitsgruppe.

Liebe Kandidatinnen und Kandidaten,
Ziel dieses Fragebogens ist es, die Ergebnisse des Auswahlverfahrens genauer zu analysieren und daraus Optimierungsmöglichkeiten abzuleiten. Wir bitten Sie daher um ehrliche und vollständige Antworten. Der Fragebogen wurde von einer Arbeitsgruppe für Evaluation entwickelt. Es wird gewährleistet, dass die Auswertung völlig anonym erfolgt und unabhängig von der Entscheidung über Ihre Aufnahme ins Medizinstudium ist.

Abbildung 5.1.: Instruktion zu Beginn des Fragebogens.

Regionale Herkunft

Die Frage zur Regionalen Herkunft bezieht sich auf die Größe des Wohngebietes (nach Einwohner) in der der Großteil der Schulzeit im Alter von 14-19 Jahren verbracht wurde.

Wo haben Sie den Großteil Ihrer Schulzeit im Alter zwischen 14 und 19 Jahren verbracht?	
<input type="checkbox"/>	Großstadt (> 50.000 Einwohner)
<input type="checkbox"/>	mittel große Stadt (10.000-50.000 Einwohner)
<input type="checkbox"/>	Kleinstadt bzw. ländliche Region (< 10.000 Einwohner)

Abbildung 5.2.: Erfassung der regionalen Herkunft.

Schulnoten

Um Testleistungen anhand der Schulnoten vorhersagen zu können (siehe 3.1), wurde die jeweils letzte Schulnote in folgenden Fächern durch Ankreuzen erfasst (siehe Abbildung 5.3): Erstens in den sogenannten Hauptfächern (Deutsch, Englisch, Mathematik) und zweitens, in den mit den Inhalten des Medizinstudiums und des EMS eng

in Zusammenhang stehenden naturwissenschaftlichen Nebenfächern (Biologie, Chemie, Physik). Entsprechend dem österreichischen Bewertungssystem von schulischen Leistungen (1=sehr gut; 2=gut; 3=befriedigend; 4=genügend; 5=nicht genügend) mussten die Kandidatinnen und Kandidaten aus anderen Ländern (vorwiegend aus der Bundesrepublik Deutschland) ihre Leistungen zuordnen. Dadurch wurde eine bessere Vergleichbarkeit der Noten ermöglicht.

Bitte geben Sie Ihre letzte Schulnote in den untenstehenden Fächern an. Falls Sie ein anderes Notensystem hatten, ordnen Sie Ihre Leistung bitte entsprechend zu (1 = sehr gut; 2 = gut; 3 = befriedigend; 4 = genügend; 5 = nicht genügend) !											
	1	2	3	4	5		1	2	3	4	5
Deutsch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Biologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Englisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Chemie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mathematik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Physik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 5.3.: Erfassung der Schulnoten

Schultyp

In welchem Schultyp haben Sie Ihre Matura/ Ihr Abitur gemacht? (Bitte <u>keine</u> Mehrfachnennungen)	
<input type="checkbox"/>	Naturwissenschaftliches Gymnasium
<input type="checkbox"/>	Gymnasium mit anderem Schwerpunkt (z.B. humanistisch, sprachlich)
<input type="checkbox"/>	Gymnasium f. Berufstätige (Studienberechtigungsprüfung)
<input type="checkbox"/>	Berufsbildende höhere Schule <u>mit</u> technischem Schwerpunkt (z.B. HTL)
<input type="checkbox"/>	Berufsbildende höhere Schule <u>ohne</u> techn. Schwerpunkt (z.B. HAK, HBLA)
<input type="checkbox"/>	Fach-/Berufsoberschule/Kollegschule (gilt nur für Deutschland)
<input type="checkbox"/>	Berufliches Gymnasium (gilt nur für Deutschland)
<input type="checkbox"/>	Gesamtschule
<input type="checkbox"/>	anderer Schultyp <u>mit</u> mathemat.-naturwissenschaftl. Schwerpunkt
<input type="checkbox"/>	anderer Schultyp <u>ohne</u> mathemat.-naturwissenschaftl. Schwerpunkt

Abbildung 5.4.: Erfassung des Schultyps

Die Erfassung des Schultyps indem die Bewerberinnen und Bewerber ihre Matura/Abitur abgelegt haben erfolgte auf 2 unterschiedlichen Ebenen: (1) Im Rahmen des Auswahlverfahrens durch die Medizinischen Universitäten und (2) durch den Fragebogen, der im Anschluss an den EMS vorgegeben wurde. Von den Medizinischen Universitäten wurde der Schultyp in Form einer 31-stufigen Variable zur Verfügung

gestellt (siehe Anhang A.4.3). Diese Kategorisierung des Schultyps orientiert sich am Österreichischen Schulsystem. Für Personen aus anderen Ländern als Österreich wurde die Kategorie „ausländische Reifeprüfung“ vergeben. Da diese Kategorie uninformativ über den Schultyp ist, werden die Daten aus der Fragebogenerhebung herangezogen.

Im Fragebogen wurde der Schultyp mittels einer 10-stufigen kategoriellen Variable erhoben (siehe Abbildung 5.4). Hierzu wurden die Bewerberinnen und Bewerber gebeten jenen Schultyp auszuwählen, indem sie ihre Matura bzw. ihr Abitur abgelegt haben (ohne Mehrfachnennung). Auf eine Erfassung naturwissenschaftlich orientierter Ausbildungszweige, sowie dass nicht alle Bewerberinnen und Bewerber aus Österreich sind, wurde bei der Erstellung der Kategorien geachtet (Spiel et al., 2008).

5.2.3. Soziodemographische Daten im Rahmen des Auswahlverfahrens

Im Rahmen des Auswahlverfahrens wurden einige soziodemographische Daten der Bewerberinnen und Bewerber von den Medizinischen Universitäten erhoben und zur Verfügung gestellt. Für diese Studie werden jedoch mit Berücksichtigung der formulierten Fragestellungen nur die Informationen über Geschlecht und Nationalität herangezogen.

5.3. Beschreibung der Stichprobe

Untersucht wurden die Testteilnehmerinnen und Testteilnehmer, die am 06. Juli 2007 am Eignungstest für das Medizinstudium (EMS) an den Medizinischen Universitäten Wien und Innsbruck teilgenommen haben. In die Analyse wurden nur die Studienbewerberinnen und Studienbewerber für das Studium der Humanmedizin aufgenommen. Eine Unterscheidung zwischen den Universitäten Wien und Innsbruck wird nicht vorgenommen, da (1) das identische Auswahlverfahren angewandt wurde und eine gemeinsame Rangreihung der Testleistungen erfolgte und (2) die Zuteilung zum Testungsort sowie die Zuteilung zum Studienort von den Medizinischen Universitäten je nach vorhandenen Ressourcen vorgenommen wurde. Die Bewerberinnen und Bewerber konnten bei der Anmeldung zwar eine Präferenz des Studienortes angeben - Daten darüber liegen aber nicht vor. Im Folgenden kann die Stichprobengröße zwischen den Analysen aufgrund fall- bzw. listenweiser Ausschlüsse von Missings in einigen Verfahren in der Ergebnisdarstellung abweichen.

5.3.1. Repräsentativität

Insgesamt nahmen am Auswahlverfahren in Wien und Innsbruck 3704 Personen teil. Von diesen Personen lagen auch alle Informationen über die Testleistungen und Unter-testleistungen sowie die im Rahmen der Anmeldung erhobenen Daten vor, die von den Medizinischen Universitäten zur Verfügung gestellt wurden. Den Fragebogen haben insgesamt 3623 Personen ausgefüllt, deren auch je ein EMS-Code und dadurch die Daten aus dem Auswahlverfahren zugeordnet werden konnten. Die Teilnahmequote am Fragebogen betrug somit fast 98 Prozent. Spiel et al. (2008) fanden, dass jene Kandidatinnen und Kandidaten, die den Fragebogen nicht ausgefüllt haben, im Durchschnitt häufiger Männer waren und schlechtere Testwerte erzielten, als jene Personen, die an der Evaluation teilgenommen haben; die Daten wurden jedoch nicht weiter auf systematische Stichprobenausfälle hin analysiert. Insgesamt kann gesagt werden, dass die Ausfüllrate des Fragebogens sehr hoch war, wobei die erste Seite meist fast vollständig ausgefüllt wurde. Die zweite Seite des Fragebogens wurde von über 76 Prozent der Kandidatinnen und Kandidaten zu mindestens 90 Prozent (=24 von 27 Fragen) ausgefüllt.

5.3.2. Geschlecht und Nationalität

Die Geschlechterverteilung der Personen, die am Eignungstest an den Medizinischen Universitäten Wien und Innsbruck teilnahmen zeigt, dass sich deutlich mehr Frauen (2075; 57.3 Prozent) als Männer (1548; 42.7 Prozent) für das Studium der Humanmedizin beworben haben.

Tabelle 5.2.: Absolute (relative) Häufigkeiten der Nationalitäten getrennt nach Geschlecht und Gesamt.

Staatsbürgerschaft (kategorisiert)	Frauen	Männer	Gesamt
Österreich	1104 (57.0)	832 (43.0)	1936
Deutschland	765 (56.0)	601 (44.0)	1366
Italien (Südtirol)	59 (68.6)	27 (31.4)	86
EU	75 (64.1)	42 (35.9)	117
Nicht-EU	72 (61.0)	46 (39.0)	118
Gesamt	2075 (57.3)	1548 (42.7)	3623

Für eine übersichtlichere Darstellung wurden die häufigsten Staatsangehörigkeiten (Häufigkeiten größer als 1 Prozent) einer eigenen Gruppe zugeordnet (vgl. Spiel et al., 2008, S. 89). Die häufigsten Gruppen waren Bewerberinnen und Bewerber aus Österreich bzw. der österreichischen Staatsbürgerschaft gleichgestellt, Deutschland und Italien (Südtirol). Die anderen Staatsangehörigkeiten wurden den beiden Kategorien

'Länder der Europäischen Union' (EU) und 'Länder nicht in der Europäischen Union' (Nicht-EU; Stand August 2007) zugeordnet. Tabelle 5.2 zeigt die absoluten und relativen Häufigkeiten (in Klammer) der Nationalitäten nach Geschlecht. Eine Tabelle mit den Häufigkeiten der nicht weiter zusammengefassten Staatszugehörigkeiten befindet sich im Anhang A.4.1.

5.3.3. Regionale Herkunft

Tabelle 5.3 zeigt, dass fast die Hälfte (45.1%) der österreichischen Kandidatinnen und Kandidaten den Großteil ihrer höheren Schulbildung in der Großstadt verbrachten, unter den deutschen Kandidatinnen und Kandidaten sind es 37.3 Prozent. Aus beiden Ländern stammen etwa ein Drittel aus ländlichen Regionen. Österreicherinnen und Österreicher stammen zu etwa einem Fünftel aus mittelgroßen Städten während es unter den Deutschen knapp unter einem Drittel sind. Unterschiede in der Verteilung

Tabelle 5.3.: Absolute (relative) Häufigkeiten der Regionalen Herkunft für Österreich, Deutschland und Gesamt (=alle Nationalitäten).

Regionale Herkunft	Österreich	Deutschland	Gesamt
Großstadt (>50.000 Einwohner)	868 (45.1)	506 (37.3)	1518 (42.2)
mittlere Stadt (10.000-50.000 Einwohner)	418 (21.7)	457 (33.7)	957 (26.6)
Kleinstadt/ländliche Region (<10.000 Einwohner)	637 (33.1)	395 (29.1)	1124 (31.2)
Gesamt	1923 (100)	1358 (100)	3599 (100)
Missings	9	8	24

in der regionalen Herkunft sind jedoch aufgrund der unterschiedlichen Bevölkerungsverteilung der beiden Länder Österreich und Deutschland zu erwarten. Ausgehend vom Bevölkerungsstand am 31.12.2008 zeigt sich, dass die Österreicherinnen und Österreicher häufiger als erwartet aus der Großstadt und mittelgroßen Städten kommen und zu selten aus kleinstädtischen/ländlichen Regionen ($\chi^2 = 412.8; p < .01$). Für Deutschland konnte keine signifikante Abweichung von der erwarteten Verteilung beobachtet werden ($\chi^2 = 4.4; p = .112$). Die genaue Berechnung siehe Anhang A.4.2. Dieser Befund zeigt deutlich, dass in Österreich die universitäre Bildungsbeteiligung deutlich durch die regionale Herkunft beeinflusst wird. Dieses Ergebnis ist übereinstimmend mit den Befunden aus der Literatur (vgl. Unger & Wroblewski, 2006).

5.3.4. Schultyp

Wie bereits in 5.2.2 beschrieben, wurde für Personen mit der Kategorie „ausländische Reifeprüfung“ die Angaben aus der Fragebogenerhebung herangezogen. Die Kategorie „ausländische Reifeprüfung“ wurde vor allem (in 506 von 584 Fällen; entspricht 86.6 Prozent) an Bewerberinnen und Bewerber mit deutscher Staatsbürgerschaft vergeben. Inhaltlich ähnliche Schultypen wurden wie folgt zu drei neuen Kategorien zusammengefasst: (1) Neusprachliches- & Humanistisches Gymnasium & Gymn. mit anderem Schwerpunkt, (2) Realgymnasium & Naturwiss.- & Mathemat. Realgymnasium und (3) Wirtschaftskundl. Realgymnasium & Handelsakademie (HAK) & Höhere Lehranstalt für wirtschaftliche Berufe (HLW). Zur übersichtlicheren Darstellung einerseits, andererseits aus statistischen Überlegungen für weitere Berechnungen wird wie folgt nur über Kategorien mit einer Häufigkeit von mindestens 41 ($\geq 1.2\%$) berichtet. Tabelle 5.4 zeigt die Häufigkeiten in den einzelnen Kategorien für die Gruppen Gesamt, Österreich und Deutschland. Dabei ist zu sehen, dass drei Kategorien praktisch nur mit österreichischen Kandidatinnen und Kandidaten besetzt wurden.

Tabelle 5.4.: Absolute (relative) Häufigkeiten von Kandidatinnen und Kandidaten im Schultyp nach Zusammenfassung und ≥ 41 Fällen für Gesamt, Österreich und Deutschland.

Schultyp	Gesamt	Österreich	Deutschland
Gymnasium (ohne spez. Schwerpunkt)	1325 (39.7)	469 (26.3)	745 (58.5)
Hum./Neusprach./Anderes Gymnasium	684 (20.5)	342 (19.2)	295 (23.2)
Naturwiss.- /Math.- /Realgymnasium	737 (22.1)	416 (23.3)	231 (18.1)
Oberstufenrealgymnasium	269 (8.1)	259 (14.5)	2 (.2)*
Wirtsch. Realgym./HAK/HLW	242 (7.3)	227 (12.7)	1 (.1)*
H. techn. u. gewerbl. Lehranstalt	77 (2.3)	70 (3.9)	0 (0)
Gesamt	3334 (100)	1783 (100)	1274 (100)
Missings	289	153	92

*...deutsche Kandidatinnen/Kandidaten mit österreichischem Schulabschluss

Der überwiegende Anteil der Bewerberinnen und Bewerber für das Humanmedizinstudium kommt aus einem Gymnasium ohne spezifischen Schwerpunkt. Je ein Fünftel kommt aus einem naturwissenschaftlich-mathematischen Realgymnasium und aus einem Gymnasium mit humanistisch/neusprachlichen oder anderen Schwerpunkt. Bewerberinnen und Bewerber aus Berufsbildenden Höheren Schulen sind mit etwa 8 Prozent deutlich seltener vertreten.

5.4. Vorgehen bei der Testverkürzung

Wie Spiel et al. (2008) mittels Rasch-Analysen zeigen konnten, sind eine Reihe von Aufgaben der einzelnen Untertests gegenüber dem Testwert (dem sogenannten inneren Teilungskriterium, vgl. Kubinger, 2006) nicht verrechnungsfair (ausgenommen der Untertest 'Konzentriertes und sorgfältiges Arbeiten – Konz'). Diese Aufgaben sind nicht Rasch-Modell-konform und sollten aus dem Test ausgeschieden werden, da sie nur unzureichend zwischen 'besseren' und 'schlechteren' Bewerberinnen und Bewerbern differenzieren können. In Tabelle 5.5 sind die Aufgabennummern angegeben, die aus dem Test ausgeschieden werden. In jenen Untertests in denen mehr als 50 Prozent der Aufgaben nach dem Testwert ausgeschlossen werden sollten (vgl. Spiel et al., 2008, S. 110), wurden nur die von Spiel und Mitarbeiterinnen angegebenen Aufgaben (d.h. maximal 50 Prozent) ausgeschlossen. Insgesamt wurden 66 Aufgaben in der verkürzten Testversion ausgeschlossen. Die Testwerte der Untertests wurden wie in der ungekürzten Testversion zur Vergleichbarkeit auf Z-Werte ($\mu = 100, \sigma = 10$) standardisiert. Ein neuer Testwert der verkürzten Testversion wurde aus den einzelnen Untertests berechnet.

Tabelle 5.5.: Testverkürzung: Ausgeschiedene Aufgaben nach Untertest; die fettgedruckten Zahlen in der Klammer geben die Anzahl der ausgeschiedenen Aufgaben pro Untertest an.

Untertest	ausgeschiedene Aufgabennummern
Quantitative und formale Probleme (QFP)	4, 20, 8, 10, 18, 6, 19, 15, 16, 9 (10)
Schlauchfiguren (SFIG)	20, 15, 16, 17, 14 (5)
Textverständnis (TV)	14, 9, 6, 5, 16, 12, 18, 10, 2, 4 (10)
Planen und Organisieren (PO)	2, 3, 4, 6, 9, 17, 12, 8 (8)
Med.-naturwiss. Grundverständnis (MNGV)	8, 17, 10, 16, 19, 18, 11, 12, 15, 20 (10)
Figuren Lernen (Figuren)	8, 10, 19, 7 (4)
Fakten Lernen (Fakten)	19, 17, 5 (3)
Muster zuordnen (Muster)	20, 11, 12, 17, 18, 15, 16, 19, 9, 14 (10)
Diagramme und Tabellen (DuT)	5, 9, 18, 20, 4, 19 (6)

5.5. Vorgehensweise bei der Auswertung und eingesetzte Verfahren

Ein Ziel der Auswertung ist eine möglichst simultane Betrachtung der Erklärungsvariablen. Die Analysemethoden werden daher so gewählt, dass die zu untersuchen-

den Variablen, soweit möglich, gleichzeitig berücksichtigt werden. Diese gleichzeitige Berücksichtigung erhöht die Aussagekraft der Analyse.

Die ungekürzte und die verkürzte Testversion werden immer voneinander getrennt berechnet. Für jede der beiden Testversionen ergeben sich folglich Tabellen¹ mit den Mittelwerten, Standardabweichungen und Stichprobenumfängen der einzelnen Faktorstufen. Mittels mehrerer t-Tests für unabhängige Stichproben wird die Mittelwertsdifferenz (d) auf Signifikanz geprüft. Da es sich hierbei um Einzelvergleiche (mehrere unabhängige t-Tests) handelt, wurde das Signifikanzniveau $\alpha = .025$ nach Bonferoni (siehe Bortz, 1999, S. 261) korrigiert. Die Varianzhomogenität wurde angenommen.

Für die Einflussgrößen Schulnoten, regionale Herkunft, Geschlecht und Nationalität auf den Testwert werden univariate mehrfaktorielle Kovarianzanalysen berechnet. Die sechs Schulnoten Deutsch, Englisch, Mathematik, Biologie, Chemie und Physik werden als Kovariaten eingebunden. Die Intervallskala der Schulnoten wird angenommen. Für die Untertests erfolgt die Analyse mittels multivariater mehrfaktorieller Kovarianzanalysen, da apriori nicht von der Unabhängigkeit der Untertests ausgegangen werden kann. Die Bedeutung des Einflusses der Schulnoten (Kovariaten) erfolgt über die standardisierten Koeffizienten (beta-Koeffizienten: β) aus den Kovarianzanalysen.

Für weitere Analysen zur Aufklärung der Unterschiede im EMS durch die Schulnoten wurde ein *Notendurchschnitt* (arithmetisches Mittel) aus allen sechs Schulnoten berechnet. Diese mittleren schulischen Leistungen der Kandidatinnen und Kandidaten wurden gruppiert in Schulnoten 1 bis 5 (genaue Vorgehensweise siehe Anhang A.5). In den einzelnen Analysen wurde die Durchschnittsnote 5 auf Grund zu geringer Fallzahlen (1 Fall) ausgeschlossen. Die Analysen hinsichtlich des Einflusses der Durchschnittsnote werden für den Testwert univariat und für die Untertests multivariat berechnet.

Für die Beantwortung der Fragestellung nach der Vorhersagekraft der Testleistungen durch das *Notenprofil* wurde wie folgt vorgegangen: Das Notenprofil wird durch ein Zentralmaß (Notendurchschnitt) und ein Streumaß (Range = Maximum minus Minimum) charakterisiert. Als Zentralmaß wurde, wie bereits erwähnt, das in Schulnoten kodierte arithmetische Mittel herangezogen. Das Streumaß charakterisiert den Grad der Homogenität des Schulnotenprofils. Der Range wurde gegenüber der Standardabweichung vorgezogen, weil dieser im Gegensatz zur Standardabweichung (welche deutlich rechtsschief ist) zu einer normalverteilten Variable führt und auch nicht weiter rekodiert werden musste.

Leistungsunterschiede im EMS hinsichtlich des *Schultyps* werden nur für den Testwert beider Testversionen nicht jedoch für die Untertests berechnet. Aufgrund der im

¹Hierfür wurden die deskriptive Daten der einzelnen realisierten Faktorstufen aus den varianzanalytischen Methoden herangezogen

Schultyp ungleichen Besetzung von österreichischen und deutschen Kandidatinnen und Kandidaten wurden nur Österreicherinnen und Österreicher in die Analysen aufgenommen. Speziell waren es die Schultypen HTL, Oberstufenrealgymnasium und wirtschaftl. Realgymnasium & HAK & HLW die nicht oder nur mit maximal 2 Fällen unter den deutschen besetzt waren (siehe 5.5). Die beiden Einflussgrößen Notendurchschnitt und Geschlecht können aufgrund zu geringer Zellenbesetzungen nicht gleichzeitig in die Varianzanalysen aufgenommen werden. Es werden daher 2 univariate mehrfaktorielle Varianzanalysen (für den Notendurchschnitt 6x4-fach, für Geschlecht 6x2-fach) getrennt voneinander berechnet.

6. Ergebnisse

In dieser Studie soll der Fragestellung nachgegangen werden, ob schulische Aspekte (Schulnoten, Notenprofil und Schultyp), die regionale Herkunft sowie Geschlecht und Nationalität die Leistungsunterschiede im EMS zwischen den Personen erklären können. Weiter wird die Fragestellung untersucht, ob der Wert der Erklärungsvariablen durch eine verkürzte Testversion (nach Ausschluss nicht verrechnungsfairer Aufgaben) erhöht werden kann. Als Leistungskriterien (abhängige Variablen) werden der standardisierte Testwert sowie die standardisierten Leistungen in den einzelnen Untertests herangezogen. Die Analysen werden in vier Schritte aufgebaut: Leistungsunterschiede im EMS in Abhängigkeit...

1. von Einzelfachnoten (Testwert und Untertests)
2. von der Durchschnittsnote (Testwert und Untertests)
3. vom Notenprofil
4. vom Schultyp

Zuerst folgt immer die Darstellung der Ergebnisse zum Testwert, danach zu den Untertests. Die Analysen werden jeweils für die ungekürzte als auch für die verkürzte Testversion durchgeführt. In der Ergebnisdarstellung werden beide Testversionen nebeneinander berichtet. Die zentralen Ergebnisse werden als Text und in Tabellenform berichtet. Deskriptive Ergebnisse zu den einzelnen Variablen sowie statistische Detailergebnisse und Einschränkungen der Verfahren sind dem Anhang zu entnehmen.

Aufgrund des Einflusses der Stichprobengröße auf die Signifikanz des empirischen Ergebnisses werden die Ergebnisse ihrer Effektgröße (=praktische Bedeutsamkeit) nach bewertet. Für die varianzanalytischen Untersuchungen erfolgt die Einteilung der Effektstärke (η_p^2) nach Cohen (1988). Sie werden in kleine ($\eta_p^2 > .0099$), mittlere ($\eta_p^2 > .0588$) und große ($\eta_p^2 > .1379$) Effekte eingeteilt. Alle signifikanten Ergebnisse ($p \leq .05$) werden berichtet. In den multivariaten Varianzanalysen wird das Wilks' Lambda interpretiert.

6.1. Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität

6.1.1. Testwert (Einzelfachnoten)

Mittels 3x2x2 univariater mehrfaktorieller ANCOVA (je für die ungekürzte und die verkürzte Testversion) wurde der Einfluss der Prädiktorvariablen Einzelfachnoten,

regionale Herkunft, Geschlecht und Nationalität auf den Testwert geprüft. Die sechs Schulnoten Deutsch(D), Englisch(E), Mathematik(M), Biologie(B), Chemie(C) und Physik(P) wurden als Kovariaten in die Analyse aufgenommen. Als unabhängige feste Faktoren waren dies:

- regionale Herkunft (Großstadt, mittelgroße Stadt, Kleinstadt/ländliche Region)
- Geschlecht (weiblich, männlich)
- Nationalität (Österreich, Deutschland)

Tabelle 6.1.: Signifikante Effekte für den Testwert nach Einzelfachnoten, Geschlecht, regionale Herkunft und Nationalität; ungekürzte Version.

	β	$F(1, 2930)$	P	η_p^2
KOVARIATEN				
Deutsch	-.045	5.54	.019	.002
Englisch	-.083	18.63	< .001	.006
Mathematik	-.232	147.35	< .001	.048
Biologie	-.064	10.92	.001	.004
Chemie	-.078	14.18	< .001	.005
Physik	-.042	5.02	.025	.002
FESTE FAKTOREN				
Geschlecht		159.322	< .001	.052
Nationalität		301.05	< .001	.093
Geschlecht*Nationalität		24.03	< .001	.008

β ... standardisierter Koeffizient

Tabelle 6.2.: Signifikante Effekte für den Testwert nach Einzelfachnoten, Geschlecht, regionale Herkunft und Nationalität; verkürzte Version.

	β	$F(1, 2930)$	P	η_p^2
KOVARIATEN				
Deutsch	-.057	8.89	.003	.003
Englisch	-.084	19.31	< .001	.007
Mathematik	-.216	125.49	< .001	.041
Biologie	-.060	9.87	.002	.003
Chemie	-.071	11.57	.001	.004
Physik	-.040	4.54	.033	.002
FESTE FAKTOREN				
Geschlecht		141.74	< .001	.046
Nationalität		327.99	< .001	.101
Geschlecht*Nationalität		22.74	< .001	.008

β ... standardisierter Koeffizient

Die Ergebnisse der ANCOVA (siehe Tabelle 6.1 und Tabelle 6.2) zeigen für beide

Testversionen vergleichbare Ergebnisse. Alle Einzelfachnoten hatten einen signifikanten Einfluss auf den Testwert. Den stärksten Einfluss zeigte die Mathematiknote mit einer Effektstärke von $\eta_p^2 > .048$ (ungekürzte Testversion) bzw. $\eta_p^2 > .041$ (verkürzte Testversion) gefolgt von den Einzelfachnoten Englisch, Chemie und Biologie. Ein negativer β -Koeffizient (standardisiert) bedeutet, dass Personen mit besserer Mathematiknote (1 ist die beste Note) auch eine höhere Testleistung erzielten. In der verkürzten Testversion sind die β -Koeffizient der Einzelfachnoten tendenziell niedriger als in der ungekürzten Testversion.

Unter den festen Faktoren zeigte sich ein mittlerer Effekt in der Nationalität wobei Personen aus Deutschland besser als jene aus Österreich waren (ungekürzte V.: $M(\text{Deutschl.}) = 103.5, SD = 9.31, M(\text{Österr.}) = 98.8, SD = 9.76$; verkürzte V.: $M(\text{Deutschl.}) = 104.0, SD = 9.13, M(\text{Österr.}) = 99.0, SD = 9.77$). Weiter zeigte sich in beiden Testversionen ein signifikanter Geschlechtseffekt. Männer erzielten bessere Leistungen als Frauen (ungekürzte V.: $M(\text{Männer}) = 102.9, SD = 9.84, M(\text{Frauen}) = 98.86, SD = 9.52$; verkürzte V.: $M(\text{Männer}) = 103.02, SD = 9.74, M(\text{Frauen}) = 99.3, SD = 9.62$). Zusätzlich zeigte sich ein schwacher aber signifikanter Interaktionseffekt zwischen Geschlecht und Nationalität. Inhaltlich bedeutet dieser, dass der nationale Unterschied in der Gruppe der Frauen größer ist als in der Gruppe der Männer. Weiter bedeutet dieser, dass der Geschlechtsunterschied zwischen den österreichischen Kandidatinnen und Kandidaten größer ist, als zwischen den deutschen. Abbildung 6.1 zeigt den Interaktionseffekt für die ungekürzte Testversion. Die deutschen Männer waren im Test die erfolgreichsten ($M = 104.73, SD = 9.26$), die Östertreicherinnen schnitten am schlechtesten ab ($M = 96.63, SD = 8.98$). Es gibt in

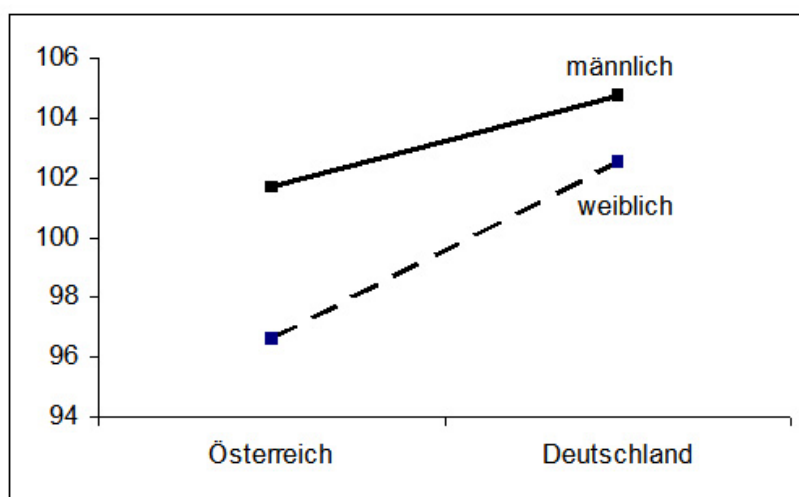


Abbildung 6.1.: ANCOVA, signifikanter Interaktionseffekt Geschlecht*Nationalität, ungekürzte Testversion.

beiden Testversionen keinen signifikanten Einfluss der regionalen Herkunft (ungekürzte Version: $p = .676$; verkürzte Version: $p = .605$).

Mittelwertsunterschiede zwischen den beiden Testversionen zeigen, dass die deutschen Bewerberinnen und Bewerber in der verkürzten Testversion um 0.57 Punkte besser abschneiden als in der ungekürzten Testversion ($t = 4.072, p < .025$); bei den österreichischen Bewerberinnen und und Bewerbern zeigt sich dieser Effekt nicht. Die Frauen (0.44 Punkte) und insbesondere die deutschen Frauen (0.69 Punkte) profitieren von der Testverkürzung. Detaillierergebnisse beider Analysen hinsichtlich Mittelwerte, Standardabweichungen und Stichprobenumfänge sind dem Anhang Tabelle A.8 zu entnehmen.

6.1.2. Untertests (Einzelfachnoten)

Zur Prüfung des Einflusses der Prädiktoren Einzelfachnoten, regionale Herkunft, Geschlecht und Nationalität auf die Untertestleistungen wurde eine 3x2x2 multivariate mehrfaktorielle ANCOVA gerechnet. Entsprechend der Vorgehensweise wie für den Testwert wurden die Einzelfachnoten als Kovariaten und regionale Herkunft, Geschlecht und Nationalität als feste Faktoren in die multivariate ANCOVA einbezogen. Abhängige Variablen waren die zehn standardisierten Untertestleistungen (in der verkürzten Testversion nur neun, ohne Untertest 'Konzentriertes und sorgfältiges Arbeiten – Konz'). Die Berechnungen erfolgten wieder getrennt für die ungekürzte und die verkürzte Testversion. Zunächst werden die multivariaten Effekte, anschließend die univariaten Effekte berichtet.

Für alle Einzelfachnoten bis auf Physik konnten kleine und für Mathematik ein mittlerer multivariater Effekt in beiden Testversionen gefunden werden. Die Effektstärken in der verkürzten Testversion waren tendenziell kleiner. Für den Faktor Geschlecht konnten in beiden Testversionen große multivariate Effekte ($\eta_p^2 \geq .16$) gefunden werden. Auch der Faktor Nationalität zeigte einen praktisch bedeutsamen mittleren Effekt. Der multivariate Interaktionseffekt (Nationalität*Geschlecht) war in beiden Testversionen als klein zu beurteilen (ungekürzte V.: $\eta_p^2 = .013$; verkürzte V.: $\eta_p^2 = .010$). Es gab keinen signifikanten Effekt der regionalen Herkunft. Die Prüfgröße Wilks' Lambda, die Signifikanzen und die Effektstärken der multivariaten Effekte beider Testversionen sind in Tabelle 6.3 dargestellt.

Die univariaten Analysen zeigen, dass von den Einzelfachnoten nur Mathematik nennenswerte Effekte ($\eta_p^2 > .01$) aufweist (genaue Ergebnisse der ungekürzten und der verkürzten Version siehe Anhang Tabelle A.9 und Tabelle A.10). In fast allen Untertests konnten signifikante Effekte der Mathematiknote beobachtet werden. Lediglich die

6.1 Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität

Tabelle 6.3.: Multivariate Effekte ANCOVA; Untertests; ungekürzte und verkürzte Testversion.

	UNGEKÜRZTE V.			VERKÜRZTE V.		
	Wilks'λ	p	η_p^2	Wilks'λ	p	η_p^2
KOVARIATEN						
Deutsch	.987	< .001	.013	.987	< .001	.013
Englisch	.979	< .001	.021	.986	< .001	.014
Mathematik	.921	< .001	.079	.933	< .001	.067
Biologie	.990	< .001	.010	.992	< .001	.008
Chemie	.989	< .001	.011	.988	< .001	.012
FESTE FAKTOREN						
Geschlecht	.832	< .001	.168	.840	< .001	.160
Nationalität	.866	< .001	.134	.874	< .001	.126
Nationalität*Geschlecht	.987	< .001	.013	.990	< .001	.010

Tabelle 6.4.: Effektstärken (η_p^2) und standardisierte Beta-Koeffizienten (β) der Kovariate Mathematiknote in den einzelnen Untertests (alle $p < .001$); ungekürzte und verkürzte Testversion.

	UNGEKÜRZTE V.		VERKÜRZTE V.	
	η_p^2	β	η_p^2	β
QFP	.061	-.268	.050	-.243
SFIG	.012	-.120	.010	-.112
TV	.014	-.125	< .01	-.097
PO	.037	-.217	.029	-.196
MNGV	.010	-.106	< .01	-.084
Figuren	.013	-.133	.012	-.127
Fakten	< .01	-.083	< .01	-.076
Muster	< .01	-.097	< .01	-.092
DuT	.029	-.181	.026	-.175
Konz	.014	-.140	–	–

–... nicht berechnet

beiden Untertests 'Fakten lernen' und 'Muster erkennen' zeigten in beiden Testversionen keine praktisch bedeutsamen Effekte. Die Mathematiknote war auch in den beiden Untertests 'Textverständnis (TV)' und 'Med.-naturwiss. Grundverständnis (MNGV)' in der verkürzten Testversion nicht praktisch bedeutsam ($\eta_p^2 > .01$). Da sich nur die Mathematiknote in beiden Testversionen von den Einzelfachnoten als praktisch bedeutsamer Prädiktor gezeigt hat, werden folgend die Effektgrößen (η_p^2) und standardisierten beta-Koeffizienten nur von der Mathematiknote berichtet. Die Werte sind in Tabelle 6.4 dargestellt (alle Signifikanzen $p < .001$). Die Untertestleistung im Untertest 'Quantitative und formale Probleme (QFP)' konnte durch die Mathematiknote von allen Un-

tertests am besten vorhergesagt werden (ungekürzte Version: $\beta = -.268$; $\eta_p^2 = .061$; ungekürzte Version: $\beta = -.243$; $\eta_p^2 = .050$), gefolgt von den Untertests 'Planen und Organisieren' und 'Diagramme und Tabellen'.

Hinsichtlich des Prädiktors Geschlecht zeigten sich in beiden Testversionen praktisch bedeutsame Einflüsse auf die Untertestleistungen des Untertests 'Diagramme und Tabellen' (ungekürzte Testversion: $\eta_p^2 = .102$; verkürzte Testversion: $\eta_p^2 = .091$; Männer sind um ca. 6 Punkte besser) und 'Quantitative und formale Probleme (QFP)' und 'Schlauchfiguren (SFIG)' (Männer sind um ca. 5 Punkte besser) gefunden. Für die beiden Untertests QFP und SFIG liegen nach Cohen mittlere Effektstärken vor. Die Effektstärken in der verkürzten Testversion sind vergleichbar mit der ungekürzten Testversion, fallen jedoch tendenziell niedriger aus. Die genauen Effektstärken, Mittelwerte, Standardabweichungen und Stichprobenumfänge der signifikanten Faktoren in den Untertests sind dem Anhang Tabelle A.11 und Tabelle A.12 zu entnehmen. Der Vergleich der Mittelwerte beider Testversionen (t-Tests) zeigt, dass nur die Frauen von der Testverkürzung in den Untertests 'Textverständnis', 'Planen und organisieren', 'Med.-naturwiss. Grundverständnis' und 'Diagramme und Tabellen' profitieren.

Effekte hinsichtlich des Faktors Nationalität zeigten sich in den Untertests in der ungekürzten Testversion 'Textverständnis' (Deutsche sind um 5.8 Punkte besser), 'Med.-naturwiss. Grundverständnis' (Deutsche sind um 5.2 Punkte besser), 'Diagramme und Tabellen' (Deutsche sind um 3.9 Punkte besser) und 'Quantitative und formale Probleme' (Deutsche sind um 3.3 Punkte besser). Tendenziell waren die Effekte in der verkürzten Testversion niedriger. Mittelwertsvergleiche der beiden Testversionen (t-Tests) zeigen, dass nur die Österreicherinnen und Österreicher von der Testverkürzung in den Untertests 'Textverständnis', 'Med.-naturwiss. Grundverständnis' und 'Diagramme und Tabellen' profitieren.

6.2. Leistungsunterschiede in Abhängigkeit von der Durchschnittsnote, regionaler Herkunft, Geschlecht und Nationalität

6.2.1. Testwert (Durchschnittsnote)

Zur Berechnung des Einflusses der Durchschnittsnote (unter Einschluss der Prädiktoren regionale Herkunft, Geschlecht und Nationalität) auf den Testwert wurde eine $4 \times 3 \times 2 \times 2$ univariate mehrfaktorielle ANOVA getrennt für die ungekürzte und die verkürzte Testversion gerechnet.

In beiden Testversionen zeigt sich ein signifikanter Effekt des Faktors Durchschnitts-

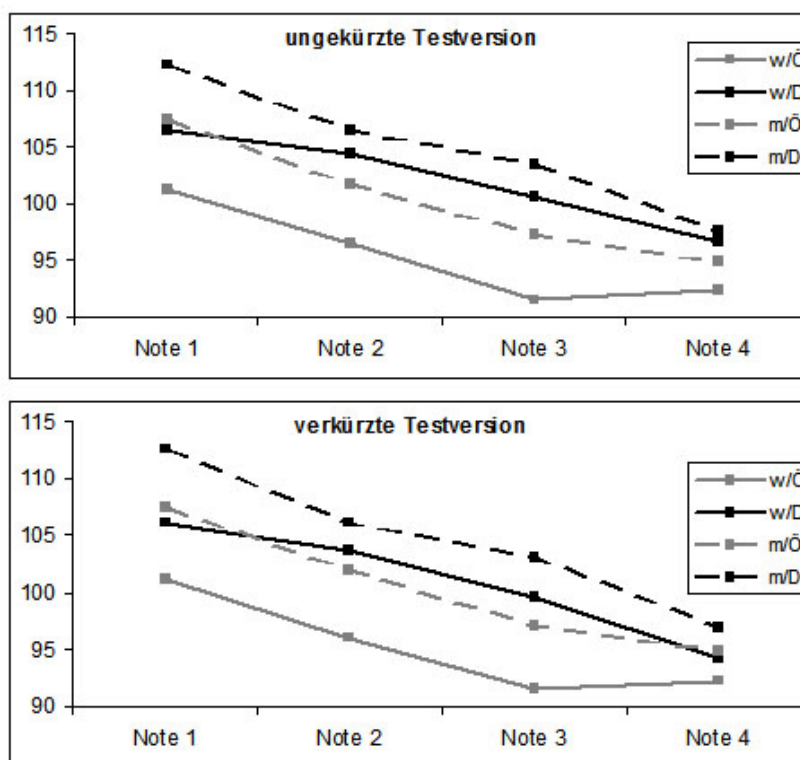


Abbildung 6.2.: Durchschnittliche Testleistung nach Durchschnittsnote getrennt nach Geschlecht und Nationalität.

note. Je besser die Durchschnittsnote einer Person ist, desto besser ist auch die Leistung im Test (ungekürzte V.: $F(3, 3200) = 61.91, p < .001; \eta_p^2 = .055$; verkürzte V.: $F(3, 3200) = 59.50, p < .001; \eta_p^2 = .053$). Dieser Zusammenhang ist relativ linear (ausgenommen bei den Österreicherinnen in den Durchschnittsnoten 3 und 4). Zur besseren Veranschaulichung wurden die Ergebnisse in Abbildung 6.2 dargestellt. Deutliche Effekte auf die Testleistung zeigten sich auch im Faktors Geschlecht (ungekürzte V.: $F(1, 3200) = 31.24, p < .001; \eta_p^2 = .010$; verkürzte V.: $F(1, 3200) = 25.25, p < .001; \eta_p^2 = .008$) und des Faktors Nationalität (ungekürzte V.: $F(1, 3200) = 42.78, p < .001; \eta_p^2 = .013$; verkürzte V.: $F(1, 3200) = 53.72, p < .001; \eta_p^2 = .017$). Männer hatten höhere Testwerte als Frauen und deutsche Kandidatinnen und Kandidaten waren erfolgreicher als österreichische. Ein signifikanter Effekt der regionalen Herkunft konnte nicht gefunden werden ($p = .143$). Es gab auch keine signifikanten Interaktionseffekte. Die genauen Mittelwerte, Standardabweichungen sowie die Stichprobenumfänge beider Testversionen finden sich im Anhang in Tabelle A.13.

Der Vergleich beider Testversionen (t-Tests) zeigt, wie auch schon in den Ergebnissen zu den Einzelfachnoten, dass die Frauen (0.48 Punkte) und speziell die deutschen Frauen (0.75 Punkte) von der Testverkürzung profitieren. Zudem zeigten sich auch

innerhalb der Notenstufen 2 und 3 der Durchschnittsnote signifikante Unterschiede in der Untertestleistung der beiden Testversionen. Frauen mit der der Durchschnittsnote 2 profitieren um 0.54 Punkte von der verkürzten Testversion. Deutsche Frauen mit einer Durchschnittsnote von 3 erreichen in der verkürzten Testversion um 0.88 Punkte mehr als in der ungekürzten Testversion (siehe Anhang Tabelle A.13).

6.2.2. Untertests (Durchschnittsnote)

Zur Berechnung des Einflusses der Durchschnittsnote (unter Einschluss der Prädiktoren regionale Herkunft, Geschlecht und Nationalität) auf die einzelnen Untertestleistungen wurde eine 4x3x2x2 mehrfaktorielle MANOVA getrennt für beide Testversionen gerechnet. Zunächst werden die multivariaten, dann die univariaten Effekte dargestellt. Tabelle 6.5 zeigt je kleine multivariate Effekte der Durchschnittsnote, des Geschlechts

Tabelle 6.5.: multivariate Effekte ($\eta_p^2 \geq .01$); Untertests; ungekürzte und verkürzte Testversion

	UNGEKÜRZTE V.		GEKÜRZTE V.	
	P	η_p^2	P	η_p^2
Durchschnittsnote	< .001	.024	< .001	.022
Geschlecht	< .001	.042	< .001	.039
Nationalität	< .001	.025	< .001	.025

und der Nationalität. Es gab keinen signifikanten multivariaten Effekte der regionalen Herkunft ($p = .393$) und auch keine signifikanten multivariaten Interaktionseffekte. Die Effektstärken der beiden Testversionen sind vergleichbar.

In den univariaten Ergebnissen beider Testversion zeigten sich vergleichbare kleine Effekte der Durchschnittsnote, des Geschlechts und der Nationalität. Tendenziell waren die Effekte in der verkürzten Testversion kleiner als in der ungekürzten Version. Die Durchschnittsnote war in allen Untertests signifikant (für alle Untertests: $p < .001$). Die Effektstärken werden aufgrund der Ähnlichkeit zwischen den beiden Testversionen folgend nur für die ungekürzte Testversion berichtet. Detailergebnisse sind dem Anhang A.8 zu entnehmen. Die stärksten Effekte hinsichtlich der Durchschnittsnote zeigten sich in den Untertests 'Quantitative und formale Probleme' ($\eta_p^2 = .039$), 'Textverständnis' ($\eta_p^2 = .039$), 'Diagramme und Tabellen' ($\eta_p^2 = .038$) und 'Med.-naturwiss. Grundverständnis' ($\eta_p^2 = .033$). Dabei hatten Personen mit besseren Noten auch höhere Testwerte. Im Geschlecht zeigten sich praktisch relevante Effekte nur in den Untertests 'Schlauchfiguren' ($\eta_p^2 = .019$), 'Diagramme und Tabellen' ($\eta_p^2 = .017$) und 'Quantitative und formale Probleme' ($\eta_p^2 = .014$). Die Mittelwerte zeigen eine Überlegenheit

der Männer. Hinsichtlich der Nationalität zeigten sich Effekte nur in den beiden Untertests 'Textverständnis' ($\eta_p^2 = .016$), 'Med.-naturwiss. Grundverständnis' ($\eta_p^2 = .014$) und 'Diagramme und Tabellen' ($\eta_p^2 = .011$). In allen Untertests sind die deutschen Kandidatinnen und Kandidaten den österreichischen überlegen.

Mittelwertsvergleiche zwischen den beiden Testversionen (t-Tests) traten in den Faktoren Durchschnittsnote, Geschlecht und Nationalität nur vereinzelt auf. Personen mit einer Durchschnittsnote von 3 profitieren durch die Testverkürzung im Untertest 'Textverständnis' (0.62 Punkte). Hinsichtlich des Faktors Geschlecht profitieren signifikant nur die Frauen von der Testverkürzung in den Untertests 'Textverständnis' (0.73 Punkte), 'Medizinisch-naturwissenschaftliches Textverständnis' (0.62 Punkte) und 'Planen und Organisieren' (0.46 Punkte). Im Faktor Nationalität profitieren die Österreicherinnen und Österreicher in den Untertests 'Textverständnis' (0.56 Punkte) und 'Diagramme und Tabellen' (0.42 Punkte) von der Testverkürzung.

6.3. Einfluss des Schulnotenprofils

In diesem Abschnitt soll die Frage nach dem Einfluss des Notenprofils auf die Leistungen im EMS untersucht werden. Wie bereits in Abschnitt 5.5 beschrieben, wird das Notenprofil über den Mittelwert und den Range charakterisiert. Der Range gibt den Grad der Homogenität des Notenprofils an. Kombinationen, die zu einer Zellenbesetzung von ≤ 5 Fällen führten, wurden von der Analyse ausgeschlossen (insgesamt waren dies 13 Fälle).

Für den Testwert in Abhängigkeit von der Durchschnittsnote und dem Range wurde eine 4x4 univariate mehrfaktorielle ANOVA gerechnet. Es zeigte sich in beiden Testversionen ein signifikanter Effekt in der Durchschnittsnote (ungekürzte Version: $F(3, 3238) = 15.36, p < .001; \eta_p^2 = .014$; verkürzte Version: $F(3, 3238) = 14.91, p = .001; \eta_p^2 < .014$). Der Range zeigte keinen signifikanten Haupteffekt (ungekürzte V.: $p = .795$; verkürzte V.: $p = .874$). Weiter zeigte sich ein signifikanter, jedoch praktisch nicht bedeutsamer Interaktionseffekt Durchschnittsnote*Range (ungekürzte V.: $F(7, 3238) = 2.75, p = .008; \eta_p^2 = .006$; verkürzte V.: $F(7, 3238) = 2.54, p = 0.013; \eta_p^2 = .005$). Abbildung 6.3 veranschaulicht die Beziehung zwischen der Durchschnittsnote und dem Range in der ungekürzten Testversion. Nur innerhalb der Durchschnittsnote 1 haben Personen mit homogenerem Notenprofil (=niedrigerer Range) im Mittel tendenziell auch höhere Testwerte. Die anderen Notenstufen der Durchschnittsnote entsprechen nicht der erwarteten Hypothese. Es gab keine signifikanten Mittelwertsunterschiede zwischen beiden Testversionen. Detailangaben zu beiden Testversionen finden sich im Anhang Tabelle A.17.

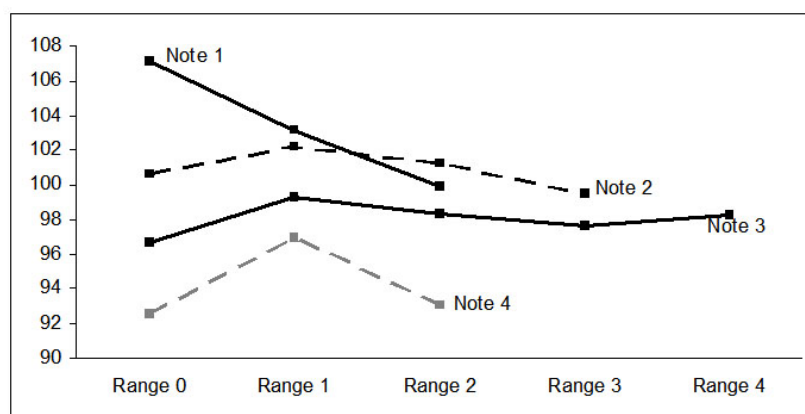


Abbildung 6.3.: Durchschnittliche Testwerte nach Durchschnittsnote (Note) und Range; ungekürzte Testversion.

6.4. Leistungsunterschiede in Abhängigkeit von Schultyp, Durchschnittsnote und Geschlecht

In diesem Abschnitt soll der Einfluss des Schultyps auf den Testwert untersucht werden. Aufgrund der im Schultyp ungleichen Besetzung von österreichischen und deutschen Kandidatinnen und Kandidaten (siehe 5.5) wurden nur Österreicherinnen und Österreicher in die Analysen aufgenommen. Die Analysen zur ungekürzten und verkürzten Testversion wurden getrennt berechnet. Es folgen zuerst die Ergebnisse in Relation zum Notendurchschnitt, dann zum Geschlecht.

6.4.1. Leistungsunterschiede in Abhängigkeit von Schultyp und Durchschnittsnote

Getrennt für beide Testversionen wurde eine 6x4 univariate mehrfaktorielle Varianzanalyse berechnet. Es zeigte sich in beiden Versionen ein kleiner Haupteffekt im Schultyp (ungekürzte V.: $F(5, 1744) = 7.02; p \leq .001; \eta_p^2 = .020$; verkürzte V.: $F(5, 1744) = 7.18; p \leq .001; \eta_p^2 = .020$) und ein mittlerer Haupteffekt im Notendurchschnitt (ungekürzte V.: $F(3, 1744) = 48.36; p \leq .001; \eta_p^2 = .077$; verkürzte V.: $F(3, 1744) = 47.71; p \leq .001; \eta_p^2 = .076$). Es gab keinen signifikanten Interaktionseffekt (ungekürzte V.: $p = .855$; verkürzte V.: $p = .747$). Signifikante Mittelwertsunterschiede (t-Tests) zwischen den beiden Testversionen konnten nicht gefunden werden (Detailergebnisse siehe Anhang Tabelle A.18). Die weitere Darstellung der Ergebnisse beziehen sich nur auf die ungekürzte Version. Erwartungsgemäß führte eine bessere Durchschnittsnote zu besseren Leistungen im EMS. Im Mittel waren HTL-Absolventinnen und -Absolventen mit einer Durchschnittsnote von 1 am erfolgreichsten ($M=109.39$;

n=19). Die schlechteste Leistung erbrachten Personen aus den Schultypen 'Oberstufengymnasium' (M=95.88; n=256) und 'wirtschaftl. Realgymnasium & HAK & HLW' (M=96.15; n=222). Zur Veranschaulichung wurden die Ergebnisse grafisch in Abbildung 6.4 dargestellt. Post-hoc Paarvergleiche (nach Scheffé; siehe Anhang Tabelle A.19) zeigten, dass sich diese beiden (im Mittel schlechtesten) Schultypen von den anderen Schultypen signifikant unterscheiden ($p < .05$). Zwischen den anderen Schultypen konnten keine signifikanten Unterschiede beobachtet werden. Es gab keine signifikanten Mittelwertsunterschiede (t-Tests) zwischen den beiden Testversionen.

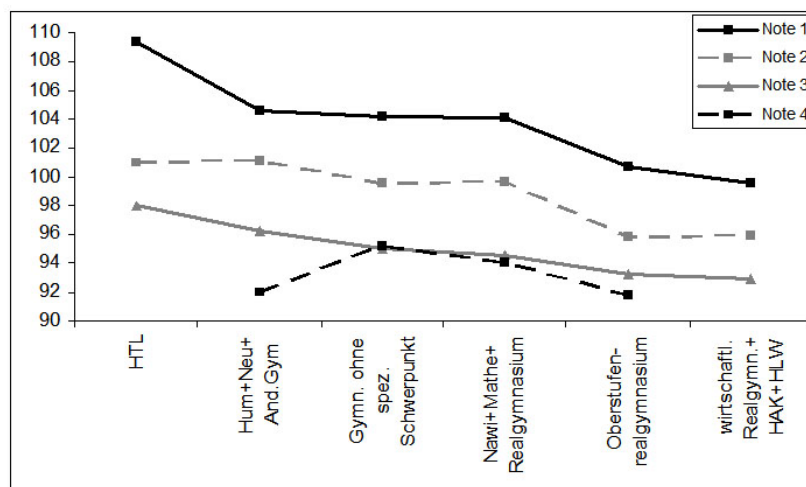


Abbildung 6.4.: Mittlere Testwerte in Abhängigkeit von Schultyp und Durchschnittsnote (Note); ungekürzte Testversion.

6.4.2. Leistungsunterschiede in Abhängigkeit von Schultyp und Geschlecht

Für beide Testversionen wurde je eine 6x2 univariate mehrfaktorielle Varianzanalyse berechnet. Die Ergebnisse der Varianzanalyse zeigen signifikante Haupteffekte des Schultyps und des Geschlechts in beiden Testversionen. Hinsichtlich des Schultyps konnte je ein kleiner Effekt von $\eta_p^2 = .033$ ($F(5, 1771) = 12.01; p < .001$) in der ungekürzten und von $\eta_p^2 = .031$ ($F(5, 1771) = 11.21; p \leq .001$) beobachtet werden. Die schlechteste Leistung erbrachten Personen aus den Schultypen 'Oberstufengymnasium' und 'wirtschaftl. Realgymnasium & HAK & HLW' (Detailergebnisse siehe Anhang Tabelle A.20. Post-hoc Paarvergleiche (nach Scheffé) zeigten (wie schon bei den Analysen zur Durchschnittsnote), dass sich diese beiden Schultypen von den anderen Schultypen signifikant unterscheiden ($p < .05$). Zwischen den anderen Schultypen konnten keine signifikanten Unterschiede beobachtet werden. Tendenziell hatten Personen mit HTL-Abschluss die höchste Testleistung mit einem Vorsprung von etwa 1.5 Punkten. Hinsichtlich des Geschlechts waren die Männer im Mittel um fast 5 Punkte besser

als die Frauen (ungekürzte V.: $F(1,1771) = 38.74; p < .001; \eta_p^2 = .021$; verkürzte V.: $F(1,1771) = 29.86; p < .001; \eta_p^2 = .017$). Im Schultyp HTL fiel der Geschlechtsunterschied geringer aus, als in den anderen Schultypen. Es gab keine signifikanten Mittelwertsunterschiede (t-Tests) zwischen beiden Testversionen. Tendenziell erreichten die Frauen in der verkürzten Testversion im Mittel höhere Testwerte als in der ungekürzten Testversion; speziell im Schultyp HTL erreichten die Frauen nach der Testverkürzung tendenziell (jedoch nicht signifikant) die höchsten Werte. Es konnten kein Interaktionseffekt Schultyp*Geschlecht gefunden werden (ungekürzte V.: $p = .202$; verkürzte V.: $p = .142$).

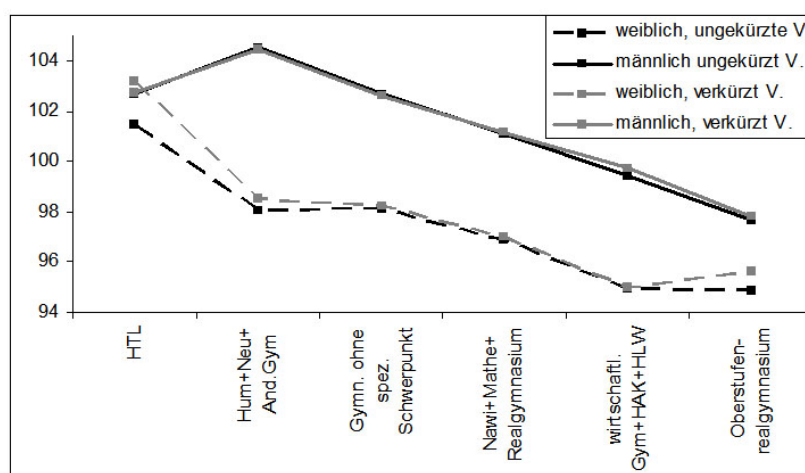


Abbildung 6.5.: Mittlere Testwerte in Abhängigkeit von Schultyp und Geschlecht); ungekürzte und verkürzte Testversion.

7. Zusammenfassung der Ergebnisse

In diesem Kapitel werden die Ergebnisse der hier vorliegenden empirischen Studie entsprechend den Fragestellungen (siehe Kapitel 4) zusammengefasst und die Fragestellungen beantwortet.

1. Fragestellung:

Erhöht der Ausschluss von nicht verrechnungsfairer Aufgaben die Prädiktorleistung der Einflussgrößen Schulnoten, Schultyp, regionale Herkunft, Geschlecht und Nationalität?

Ziel des Ausschlusses der nicht verrechnungsfairer Aufgaben war, störende Einflussgrößen zu minimieren. Die Idee dahinter ist das Konstrukt 'sauber' zu messen, um so reliablere und validere Zusammenhänge mit anderen empirisch erfassbaren Einflussgrößen finden. Die Leistungsunterschiede zwischen den Personen könnten dann durch die Prädiktoren besser erklärt werden.

Die Ergebnisse zeigen, dass der Ausschluss der nicht verrechnungsfairer Aufgaben (verkürzte Testversion) nicht zu einer verbesserten Aufklärung der Leistungsunterschiede zwischen den Personen führte. Die gefundenen Effekte sind in beiden Testversionen praktisch gleich. Tendenziell waren die Effekte in der verkürzten Testversion meist kleiner als in der ungekürzten Testversion. Dies gilt für die Testleistung gesamt als auch für die einzelnen Untertests.

Zwar konnte durch die verkürzte Testversion die Leistungen im EMS nicht besser erklärt werden, jedoch zeigte sich, dass einzelne Gruppen von der Testverkürzung profitieren. Mittelwertsunterschiede zwischen beiden Testversionen zeigen, dass die Frauen und speziell die deutschen Frauen von der Testverkürzung profitieren würden. Auch innerhalb einiger Untertests konnten signifikante Mittelwertsunterschiede zwischen den beiden Testversionen gefunden werden. In den Untertests 'Textverständnis', 'Medizinisch-naturwissenschaftliches Grundverständnis' und 'Diagramme und Tabellen' profitieren Frauen und österreichische Kandidatinnen und Kandidaten. Im Untertest 'Planen und Organisieren' scheiden die Frauen in der verkürzten Testversion besser ab.

2. Fragestellung:

Können die Schulnoten die Leistungen im EMS erklären und vorhersagen?

Für die Einzelfachnoten (Deutsch, Englisch, Mathematik, Biologie, Chemie und Physik) als auch für die Durchschnittsnote konnten signifikanten Einflüsse auf die Test-

leistung gezeigt werden. Bis auf einen mittleren Effekt (nach Cohen, 1988) in der Mathematiknote, zeigten die anderen Einzelfachnoten nur kleine Effekt. Die Schulnoten können somit die Leistungen im EMS teilweise erklären und vorhersagen.

Von den sechs Einzelfachnoten erwies sich Mathematik als der beste Prädiktor der Testleistung gefolgt von Englisch, Chemie und Biologie. Die Prädiktorkraft der Physiknote war am geringsten. Zwischen der Durchschnittsnote und der Testleistung wurde ein weitgehend linearer Zusammenhang gefunden (ausgenommen bei den Österreicherinnen in den Durchschnittsnoten 3 und 4). Grundsätzlich gilt, je besser die Durchschnittsnote ist, desto höher ist auch die erbrachte Testleistung.

Auch hinsichtlich der Untertestleistungen konnten signifikante Effekte fast aller Einzelfachnoten (bis auf Physik) und der Durchschnittsnote gefunden werden. Von den Einzelfachnoten konnte bei Mathematik ein mittlerer Effekt, bei anderen Einzelfachnoten nur kleine Effekte beobachtet werden. In fast allen Untertests bis auf 'Fakten lernen' und 'Muster erkennen' konnten signifikante Effekte der Mathematiknote beobachtet werden. In den Untertest 'Quantitative und formale Probleme' zeigte sich ein mittlerer Effekt, in den Untertests 'Planen und Organisieren' und 'Diagramme und Tabellen' kleine Effekte. Praktisch bedeutsame Effekte hinsichtlich der Durchschnittsnote zeigten sich in den Untertests 'Quantitative und formale Probleme', 'Textverständnis', 'Diagramme und Tabellen' und 'Med.-naturwiss. Grundverständnis'.

3. Fragestellung:

Kann das Notenprofil, d.h. das Ausmaß der Homogenität der Schulnoten die Vorhersage der Leistungen im EMS verbessern?

Die Einzelfachnoten einer Person stellen statistisch betrachtet eine Häufigkeitsverteilung dar. Diese Verteilung wurde über die beiden Maße Durchschnittsnote und dem Abweichungsmaß Range beschrieben. Diese beiden Lagemaße der Verteilung kennzeichnen das Notenprofil. Der Range kann inhaltlich als Homogenitätsmaß aufgefasst werden. Durch den Einbezug des Ranges können Information zusätzlich zur Durchschnittsnote über die Schulleistungen gewonnen und in den Analysen berücksichtigt werden. Die dahinter liegende Idee ist: Je einheitlicher (homogener) die Schulnoten sind, desto stabiler ist die schulische Leistung zu bewerten. Hingegen spricht eine Inhomogenität des Notenprofils für unterschiedlich starke Ausprägungen in den fachspezifischen Fähigkeiten der Person. Diese unterschiedlichen Ausprägungen der fachspezifischen Fähigkeiten könnten eine Erklärung für die Leistungsunterschiede im EMS liefern.

Es konnte kein Effekt des Grades der Homogenität des Notenprofils auf die Testleistung empirisch gezeigt werden. Tendenziell, jedoch nicht signifikant, zeigte sich nur in der Gruppe der Personen mit einer Durchschnittsnote von 1, dass sich

mit abnehmendem Grad der Homogenität auch die Testleistung verringert. Nur die Durchschnittsnote zeigte einen kleinen Effekt. Ein signifikanter, jedoch praktisch nicht bedeutsamer Interaktionseffekt Durchschnittsnote*Range konnte beobachtet werden. Nur innerhalb der Durchschnittsnote 1 haben Personen mit homogenerem Notenprofil (=niedrigerer Range) im Mittel tendenziell auch höhere Testwerte. Die anderen Notenstufen der Durchschnittsnote entsprechen nicht der erwarteten Hypothese.

4. Fragestellung:

- (1) Hat der absolvierte Schultyp einen Einfluss auf die Testleistung im EMS?*
- (2) Sind HTL-Absolventinnen und -Absolventen im EMS erfolgreicher als Personen aus anderen Schultypen?*

Die verschiedenen Schultypen haben unterschiedliche Ausbildungsziele und vermitteln auch durch ihre Lehrpläne unterschiedliche (fachspezifische) Fähigkeiten und Kompetenzen. Dies legt die Vermutung nahe, dass der besuchte Schultyp einen Einfluss auf die Testleistung im EMS hat. Der EMS enthält eine Reihe von Untertests die auf mathematische und naturwissenschaftliche Fähigkeiten abzielen. Parallelen zum Mathematikunterricht sind erkennbar. Dies lässt erwarten, dass Personen aus Schultypen mit mathematisch-naturwissenschaftlichen Schwerpunkt, speziell facheinschlägigen Höheren Technischen Lehranstalten (HTL) gegenüber Personen aus anderen Schultypen bessere Testleistungen zeigen.

Es konnte ein kleiner Effekt des Schultyps zur Erklärung der Testleistung gefunden werden. Der besuchte Schultyp hat somit einen Einfluss auf die Leistung im EMS. Absolventinnen und -Absolventen aus dem 'Oberstufenrealgymnasium' und 'wirtschaftl. Realgymnasium & HAK & HLW' schnitten gegenüber allen anderen Schultypen signifikant am schlechtesten ab. Zwischen den anderen Schultypen konnten keine signifikanten Unterschiede beobachtet werden. Tendenziell jedoch erreichten die HTL-Absolventinnen und -Absolventen, entsprechend der formulierten Hypothese, im Mittel die höchste Testleistung. Der Effekt der Durchschnittsnote, dass eine bessere Note auch zu einer höheren Testleistung führt, konnte über die einzelnen Schultypen hinweg weitestgehend bestätigt werden. In den Schultypen 'Gymnasium ohne spezifischen Schwerpunkt' und 'Realgymnasium mit naturwissenschaftl. oder mathematischen Schwerpunkt' zeigte sich dieser Effekt nicht. Die Geschlechtsunterschiede innerhalb der Schultypen (bis auf HTL) zeigten sich deutlich. Männer schnitten durchschnittlich um 5 Punkte besser ab. Innerhalb des Schultyps HTL fiel der Geschlechtsunterschied deutlich geringer aus. Zudem erreichen im Schultyp HTL die Frauen nach der Testverkürzung tendenziell (jedoch statistisch nicht signifikant) die höchsten Testwerte.

5. Fragestellung:

Hat die regionale Herkunft einen Einfluss auf die Testleistungen im EMS?

Es konnte kein signifikanter Effekt der regionalen Herkunft zur Erklärung der Leistungsunterscheide in der Testleistung des EMS gefunden werden. In der ungekürzten Testversion im Untertests 'medizinisch-naturwissenschaftliches Grundverständnis' zeigte sich ein signifikanter (jedoch praktisch nicht bedeutsamer) Effekt der regionalen Herkunft. Weiter zeigt sich ein ebenso praktisch nicht bedeutsamer Interaktionseffekt regionale Herkunft*Nationalität im Untertest 'Muster Lernen'.

Geschlecht und Nationalität

Aufgrund der bereits vorliegenden empirischen Befunde (vgl. Spiel et al., 2008) waren Geschlechtseffekte und Effekte in der Nationalität (Österreich/Deutschland) zu erwarten. Diese Befunde konnten auch mit der hier vorliegenden Studie repliziert werden. Frauen schnitten deutlich schlechter ab als Männern und österreichische Kandidatinnen und Kandidaten schlechter als deutsche. Ebenso der erwartete Interaktionseffekt zwischen Geschlecht und Nationalität konnte repliziert werden. Inhaltlich bedeutet dieser, dass der nationale Unterschied (Österreich/Deutschland) in der Gruppe der Frauen stärker ausfällt als in der Gruppe der Männer. Der Geschlechtsunterschied zwischen den österreichischen Kandidatinnen und Kandidaten ist stärker ausgeprägt als zwischen den deutschen.

8. Diskussion

Als ein zentrales Ergebnis der hier vorliegenden Studie kann festgehalten werden, dass die Schulnoten die Testleistungen teilweise erklären können. Dies gilt einerseits für die Testleistungen des gesamten Tests, andererseits auch für die einzelnen Untertests. Von den Einzelfachnoten erwies sich Mathematik als bester Prädiktor (mittlerer Effekt). Die Prognosekraft der Mathematiknote für den Studienerfolg, speziell für das Studium der Humanmedizin, konnten Trapman et al. (2007) belegen. Wenn der EMS für den Studienerfolg relevante Fähigkeiten und Kompetenzen prüft, dann muss folglich auch ein Zusammenhang mit der Mathematiknote bestehen. Zudem zielen einige Untertests des EMS wie 'Quantitative und formale Probleme' inhaltlich auf mathematische Fähigkeiten ab. Durch die hier vorliegende Studie konnte die Bedeutung der Mathematiknote als Prädiktor für die Testleistung des EMS bestätigt werden. Der signifikante Einfluss der naturwissenschaftlichen Noten (Biologie, Chemie, Physik) war aufgrund einiger Untertests, die speziell auf naturwissenschaftliche Fähigkeiten abzielen (vgl. Hänsgen & Spicher, 2007) zu erwarten. Die Ergebnisse waren aber nicht so deutlich wie für die Mathematiknote. Wie auch die Einzelfachnoten konnte die Durchschnittsnote (gebildet aus den sechs Einzelfachnoten) die Testleistung teilweise erklären. Der gefundene Zusammenhang ist allerdings etwas kleiner, als zwischen der durchschnittlichen Maturaleistung und dem EMS-Testergebnis in der Literatur (vgl. Hänsgen & Spicher, 2004) berichtet wird.

Kleine Zusammenhänge mit den Schulnoten sehen die Testautoren (Hänsgen & Spicher, 2004) als Hinweis, dass der EMS als auch die Schulleistungen zwar auf einen gemeinsamen Fähigkeitsfaktor zurückzuführen seien, aber der EMS doch in hinreichender Weise eigenständige Fähigkeiten und Kompetenzen prüfe. Grundsätzlich sollen Auswahlverfahren nur Fähigkeiten und Kompetenzen prüfen, die für den Studienerfolg relevant sind und die nicht erst im Studium erworben werden (Spiel et al., 2007). Hierbei stellt sich die Frage, wo und wann diese Fähigkeiten und Kompetenzen erworben werden. Hierbei kommt der Institution Schule wohl die größte Bedeutung für den Erwerb von fach- und bildungsspezifischen Fähigkeiten und Kompetenzen zu. Laut den Testautoren soll der EMS vorwiegend Fähigkeiten erfassen, die über eine langjährige Lern- und Entwicklungsprozesse (vgl. Zentrum für Testentwicklung und Diagnostik, 2007) erworben wurden. Wenn die Schulnoten tatsächlich die erworbenen fach- und bildungsspezifischen Fähigkeiten und Kompetenzen abbilden, dann sollten diese auch ein geeignetes Maß für die im Test geforderten Fähigkeiten und Kompetenzen sein.

Für die mangelnde Prädiktorkraft der Schulnoten bieten sich weitere Erklärungen an:

(1) Die Schulnoten erfassen aufgrund ihrer mangelnden Güte, speziell ihrer mangelnden Vergleichbarkeit, neben der Fachleistung auch nicht-fähigkeitsspezifische Einflussgrößen wie beispielsweise schulangepasstes Sozialverhalten (Schrader & Helmke, 2001) oder Fehler in der Urteilsbildung der Lehrerinnen und Lehrer (vgl. Jäger & Lissmann, 2004). Zusätzlich wird die Prognosekraft durch mangelnde Abgrenzung der Schulfächer untereinander negativ beeinflusst (Wild & Krapp, 2001). Mehrere empirische Befunde stützen die Annahme, dass Mädchen und Jungen mit verschiedenen Maßstäben beurteilt werden (Budde, 2008; Kenney-Benson et al., 2006). Mädchen haben oft die besseren Schulnoten, schneiden aber in objektiven Leistungstests schlechter ab. Dieses Phänomen wird mit der hier vorliegenden Studie bestätigt. Dieser Befund ist sehr bedeutend, spricht er doch für eine unterschiedliche Sozialisierung von Mädchen und Jungen.

(2) Gesteht man den Schulnoten trotz der Mängel in ihrer diagnostischen Güte eine gewisse Validität in der schulischen Leistungsbeurteilung zu (vgl. Unterabschnitt 3.1.1), so sollten sich deutlich größere Effekte zeigen. Möglicherweise sind die vom EMS geprüften Fähigkeiten zu 'eigenständig'. Zur Klärung dieser Frage wäre es notwendig, die Aufgaben des EMS auf ihre inhaltliche Zuordnung zu bestimmten Fähigkeitsdimensionen zu untersuchen. Die nach Hänsgen und Spicher (2007) aufgelisteten Fähigkeiten (vgl. Tabelle 5.1), auf die die einzelnen Untertests abzielen sollen, weisen keine Fähigkeiten auf, die nicht grundsätzlich für einen erfolgreichen Bildungsweg bis hin zur Matura bzw. Abitur gelernt werden oder für diesen Bildungserfolg relevant sind.

(3) Aufgrund der mangelnden Testgüte des EMS (sehr niedrige Trennschärfen; viele nicht verrechnungsfaire Aufgaben nach Testwert; vgl. Spiel et al., 2008) ist die Differenzierungsfähigkeit zwischen 'guten' und 'schlechten' Kandidatinnen und Kandidaten unzureichend. Eine eindeutige Aussage über die Fähigkeit einer Person ist anhand dieser Aufgaben nicht möglich. Für die eindimensionale Erfassung eines Konstrukts müssen die Aufgaben im Sinne des diagnostischen Gütekriteriums Skalierung (siehe Unterabschnitt 2.2.2) verrechnungsfair sein. Die Eindimensionalität eines Untertests ist eine wichtige Voraussetzung für die Interpretation des Messergebnisses. Wenn störende Faktoren das Messergebnis beeinflussen, so können tatsächlich vorhandene Effekte von bestimmten und grundsätzlich geeigneten Einflussgrößen nicht oder nur schwerer in empirischen Untersuchungen gefunden werden (vgl. Bortz & Döring, 2006). Dies führt auch zu einem niedrigeren Zusammenhang zwischen den Schulnoten und der Testleistung. Durch den Ausschluss der nicht verrechnungsfaireren Aufgaben könnten mögliche Störfaktoren minimiert werden. Zudem würde sich die Effizienz des Verfahrens durch eine erhöhte Treffersicherheit deutlich verbessern. Die Anzahl der Geeigneten unter den Ausgewählten könnte so erhöht und gleichzeitig die Anzahl der

Geeigneten unter den Nicht-Ausgewählten verringert werden.

Durch die Einbeziehung der Homogenität des Notenprofils sollte der Informationsgehalt über die schulischen Leistungen erhöht werden. Das Ausmaß der Homogenität zeigte jedoch keinen signifikanten Effekt. Lediglich ein sehr kleiner Interaktionseffekt konnte festgestellt werden. Personen mit nur Einsern hatten tendenziell bessere Testleistungen. Üblicherweise wird nur der Durchschnittswert in die Analysen einbezogen. Der Einbezug des Abweichungsmaßes liefert jedoch grundsätzlich zusätzliche Informationen über die Personen. Obwohl keine Effekte gefunden werden konnten, stellt diese Vorgehensweise eine innovative Herangehensweise dar. Für weiterer Studien sollte zur genaueren Abklärung des Nutzen des Abweichungsmaßes weiterhin untersucht werden.

Als eine Methode die Testgüte des EMS zu erhöhen, wurden die als nicht-verrechnungsfair identifizierter Aufgaben (Spiel et al., 2008) ausgeschlossen. Die daraus entstandene verkürzte Testversion führte jedoch gegenteilig der Erwartungen nicht zu einer höheren Erklärungskraft der Prädiktoren. Die gefundenen Effekte waren in beiden Testversionen praktisch gleich. Tendenziell waren die Effekte in der verkürzten Testversion meist kleiner als in der ungekürzten Testversion. Dies ist darin begründet, weil nach der Testverkürzung weniger Aufgaben in den Untertests waren (für einige Untertests sogar nur die Hälfte der Aufgaben). Mögliche Erklärungen für das Ausbleiben der erwarteten Effekte wären einerseits im Test selbst zu suchen, aufgrund der mangelnden Differenzierungsfähigkeit. Andererseits, dass die eingesetzten Prädiktoren nicht ausreichend mit den im EMS relevanten Aspekten in Zusammenhang stehen. Aufgrund der Befunde aus der Literatur sind Schulnoten (primär die Maturitätsdurchschnitte) jedoch gute Prädiktoren für Studierfähigkeitstests. Zudem stehen, wie bereits erwähnt, die Fähigkeiten, die die Untertests zu messen beanspruchen, in engem Zusammenhang mit einer höheren Schulbildung.

Geschlechtsspezifische Ergebnisse zeigen, dass primär die deutschen Frauen von der Testverkürzung profitieren. Dies könnte dadurch erklärt werden, dass bei der Testverkürzung nur Aufgaben ausgeschieden wurden, die nach dem Testwert als nicht-verrechnungsfair identifiziert wurden. Aufgaben, die sich zusätzlich nach anderen Kriterien wie Geschlecht und Nationalität als nicht-verrechnungsfair zeigten, wurden nicht ausgeschlossen, da sonst in einigen Untertests zu wenige Aufgaben übrig geblieben wären. Mit derart wenigen Aufgaben wäre es nicht mehr sinnvoll eine bestimmte Fähigkeit mit einem Untertest zu messen. Inhaltliche Erklärungen der Leistungsunterschiede durch den Test selbst entziehen sich aufgrund der Geheimhaltung der Aufgaben einer Evaluation und Optimierung.

Gegen die inhaltliche Güte des EMS spricht auch das empirische Ergebnis, dass

die Testleistungen am besten durch das Geschlecht und der Nationalität (Österreich/Deutschland) einer Kandidatin oder eines Kandidaten vorhergesagt werden können. Dies führt zu einem systematischen Ausschließen von bestimmten Personengruppen. Neben den individuellen Konsequenzen kann dies auch auf Dauer mögliche Auswirkungen auf die Gesellschaft haben, wenn systematisch weniger Frauen ein Medizinstudium beginnen. Zur Sicherung der Qualität und der Effizienz des Verfahrens müssen Diskriminierungen, die durch den Test zustande kommen, ausgeschaltet werden. Wie bereits Analysen nach dem Rasch-Modell zeigen sind einige Aufgaben nicht verrechnungsfair hinsichtlich des Kriteriums Geschlecht (vgl. Spiel et al., 2008). Auch wenn das Verfahren selbst die Personen nicht benachteiligt, so besteht dennoch ein bildungswissenschaftliches Interesse die Einflussgrößen auf die Leistung aufzudecken und Erklärungen für deren Wirksamkeit zu finden. Durch unterschiedliche Sozialisation haben die Kandidatinnen und Kandidaten ungleiche Voraussetzungen und somit ungleiche Chancen. Entstehen diese Ungleichheitsdimensionen aufgrund der Sozialisierung, so hat dies für bildungswissenschaftliche Erkenntnisse und bildungspolitische Entscheidungen große Bedeutung. Derartige Geschlechtseffekte können durch gleiche Leistungsrückmeldungen an die Mädchen wie an die Jungen minimiert oder gar ausgeschaltet werden.

Zusätzlich zu den bereits in der Literatur vorhandenen Befunde zur Einschränkung der Testgüte des EMS (vgl. Spiel et al., 2008) machen die Ergebnisse dieser Studie nochmals deutlich, dass im Rahmen der Zulassung zum Medizinstudium nicht nur ein einzelnes Verfahren einzusetzen ist. Es ist unrealistisch und unkritisch zu glauben, dass die Eignung für das Medizinstudium grundsätzlich mit einem einzigen Test festzustellen sei. Für eine valide Prognose des Studiumerfolgs wären auch sicher nicht nur leistungsspezifische Fähigkeiten und Fertigkeiten zu messen nötig.

Ein weiteres zentrales Ergebnis der Studie ist, dass der besuchte Schultyp einen Einfluss auf die Testleistung im EMS hat. Einige Untertests des EMS orientieren sich stark an naturwissenschaftlichen Inhalten. Entsprechend den Unterschieden in den Lehrplänen in den naturwissenschaftlichen Fächern war ein deutlicherer Unterschied zu erwarten gewesen. Insbesondere HTL-Absolventinnen und Absolventen sollten aufgrund der facheinschlägigen Ausbildung eine signifikant höhere Leistung gegenüber den anderen Schultypen erzielen. Einschränkend für diese Studie ist festzustellen, dass die Stichprobenunterschiede zwischen den Schultypen teilweise sehr groß sind. Dies kann zur Folge haben, dass Unterschiede nicht so leicht signifikant werden. Zudem gab es nur wenige Kandidatinnen und Kandidaten, die eine HTL besucht haben. Mögliche Effekte können aufgrund des mathematischen Zusammenhangs zwischen Stichprobengröße

und Signifikanz empirisch nicht gezeigt werden (vgl. Bortz & Döring, 2006).

Hinsichtlich der regionalen Herkunft konnten keine relevanten Effekte zur Erklärung der Leistungsunterschiede im EMS gefunden werden. In der Literatur werden Effekte auf die Leistung in nicht eindeutiger Richtung gefunden. Mit dem hier vorliegenden Ergebnis über den Einfluss der regionalen Herkunft auf objektive Leistungstest wird die Hypothese gestärkt, dass keine Unterschiede zwischen der regionalen Herkunft vorhanden sind. Dieses Ergebnis muss aber relativiert werden. Es könnte auch dadurch erklärt werden, dass es sich bei den Bewerberinnen und Bewerbern für ein Studium bereits um eine spezielle Stichprobe handelt. Die regionale Herkunft gilt als Benachteiligungsfaktor für das Ausschöpfen von Bildungsmöglichkeiten (Unger & Wroblewski, 2006), jedoch die Gruppe der Studienbewerberinnen und Bewerber konnten diesen Ungleichheitsfaktor ausgleichen z.B. durch hohe Schulbildung der Eltern (vgl. Spiel et al., 2008). Bei dieser speziellen Gruppe von Personen ist die regionale Herkunft nicht wirksam.

Zusätzlich zu den bereits vorhandenen Ungleichheitsdimensionen wie beispielsweise regionale Herkunft und Wahl des Schultyps, ungleiche Sozialisierung von Mädchen und Jungen in der Schule und sozioökonomischer Status stellt ein Auswahlverfahren grundsätzlich eine weitere Hürde zur Zulassung zum Hochschulstudium dar. Das Verfahren sollte daher fair sein und nicht weiter die Ungleichheitsdimensionen stützen. Regelmäßige Evaluationen zur Sicherung der Qualität, der Fairness und der Effizienz des Verfahrens sollten durchgeführt werden.

Ein sehr großes Problem, das sich hinsichtlich der empirischen Bewertung der Validität und der Effizienz von Auswahlverfahren stellt ist, dass sie nur zum Einsatz kommen, wenn eine Notwendigkeit für ihre Anwendung besteht und ihr Einsatz rechtlich geregelt wird. Dies hat den entscheidenden Nachteil, dass immer die Gruppe der 'falsch negativ Selektierten' (d.h. jene Personen, die das Erfolgskriterium erreichen aber durch das Verfahren fälschlicherweise ausgeschieden wurden) für Untersuchungen nicht herangezogen werden kann. Derartig zensierte Datensätze verhindern eine eindeutige Bewertung der prognostischen Validität und der Effizienz des Verfahrens (vgl. Kersting, 2003; Kubinger, 2006). So auch im EMS. Für zukünftige Auswahlverfahren wäre es daher ratsam, dieses einer kurzen Probezeit möglichst unter Realbedingungen zu unterziehen, in der keine Selektion durch das Verfahren selbst stattfindet.

Die hier vorgestellten Ergebnisse und deren Erklärungen erwecken vielleicht den Eindruck, dass der EMS zur Auswahl zum Studium der Humanmedizin nicht geeignet ist. Diesem muss entgegen gehalten werden, dass bereits vorhandene Studien dem EMS eine gute prognostische Validität bescheinigen können (vgl. Hänsgen & Spicher, 2000, 2001).

Es ist auch sehr wahrscheinlich, dass der Großteil der gefundenen Geschlechts- und Nationalitätsunterschiede aufgrund unterschiedlicher Sozialisation und unterschiedlichen Bildungssystemen und Bildungskultur zustande kommt und in keinem direkten Zusammenhang mit dem medizinischen Auswahlverfahren steht. Für zukünftige Studien sollten weitere Erklärungsvariablen für die Testleistungen untersucht werden. Die hier vorgelegten Befunde sollen primär dazu beitragen, Wissenschaft und Politik für diese Problematiken zu sensibilisieren und Forschung und Praxis anregen, weiter Aufmerksamkeit der Kontrolle und Optimierung der Qualität und Effizienz von Auswahlverfahren zu widmen.

9. Literatur

- Andres, L., & Looker, E.D. (2001). Rurality and capital: Educational expectations and attainment of rural, urban/rural and metropolitan youth. *Canadian Journal of Higher Education*, 31 (2), 1-46.
- Aronson, E., & Mills, J. (1959). The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 59 (2), 177-181.
- Bacher, J. (2003). Soziale Ungleichheit und Bildungspartizipation im weiterführenden Schulsystem Österreichs. *Österreichische Zeitschrift für Soziologie*, 28 (3), 3-32.
- Bacher, J. (2005). Bildungsungleichheit und Bildungsbenachteiligung im weiterführenden Schulsystem Österreichs – Eine Sekundäranalyse der PISA-2000 Erhebung. *SWS-Rundschau*, 45, 37-62.
- Bacher, J. (2006). Forschungslage zu Bildungsungleichheiten in Österreich. In B. Herzog-Punzenberger (Hrsg.), *Bildungsbe/nach/teiligungen in Österreich und im internationalen Vergleich. KMI Working Paper Series*, 10, 7-26. http://www.oeaw.ac.at/kmi/Bilder/kmi_WP10.pdf
- Baeriswyl, F., Trautwein, U., Wandeler, C. & Lüdtke, O. (2009). Wie gut prognostizieren subjektive Lehrerempfehlungen und schulische Testleistungen beim Übertritt die Mathematik- und Deutschleistung in der Sekundarstufe I? In J. Baumert, K. Maaz, & U. Trautwein (Hrsg.), *Bildungsentscheidungen* (Zeitschrift für Erziehungswissenschaft: Sonderheft 12), (S. 352-372). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bauer, A. (2005). Volkszählungsergebnisse 2001: Soziodemographische Determinanten der Bildungsbeteiligung. *Statistische Nachrichten*, 60 (2), 108-120.
- Baumert, J., Bos, W., Brockmann, J., Gruehn, S., Klieme, E., Köller, O. et al. (2000, November). *TIMSS/III-Deutschland. Der Abschlussbericht. Zusammenfassung ausgewählter Ergebnisse der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur mathematischen und naturwissenschaftlichen Bildung am Ende der Schullaufbahn*. Berlin. Zugriff am 19.07.2010 http://www.timss.mpg.de/TIMSS_im_Ueberblick/TIMSSIII-Broschuere.pdf
- Baumert, J., Stanat, P. & Watermann, R. (Hrsg.). (2006). *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baumert, J., Trautwein, U. & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Zusammenfassung zentraler Befunde* (S. 59-72). Zugriff am 16.07.2010 http://www.mpib-berlin.mpg.de/pisa/PISA-E_Vertief_Zusammenfassung.pdf

- Becker, M., Lüdtke, O., Trautwein, U. & Baumert, J. (2006). Leistungszuwachs in Mathematik. Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem?. *Zeitschrift für Pädagogische Psychologie*, 20 (4), 233-242.
- BGBL. I Nr. 77/2005, *Änderung des Universitätsgesetzes 2002*. Zugriff am 03.01.2010 http://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2005_I_77/BGBLA_2005_I_77.pdf
- BGBL. I Nr. 74/2006, *Änderung des Universitätsgesetzes 2002*. Zugriff am 03.01.2010 http://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2006_I_74/BGBLA_2006_I_74.pdf
- Borland, M. V. & Howsen, R. M. (1999). A Note on Student Academic Performance. In Rural Versus Urban Areas. *American Journal of Economics and Sociology*, 58 (3), 537-546.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5.Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Brambring, M. (1983). Spezielle Eignungsdiagnostik. In N. Birbaumer (Hrsg.), *Enzyklopädie der Psychologie. Grundlagen psychologischer Diagnostik* (S. 414-481). Göttingen: Hogrefe.
- Brügelmann, H. (2006). *Sind Noten nützlich - und nötig? Ziffernzensuren und ihre Alternativen im empirischen Vergleich Eine wissenschaftliche Expertise des Grundschulverbandes erstellt von der Arbeitsgruppe Primarstufe an der Universität Siegen*. Frankfurt/Main: Grundschulverband e.V.
- Budde, J. (2008). *Bildungs(miss)erfolge von Jungen und Berufswahlverhalten bei Jungen/männlichen Jugendlichen* (BMBF Bildungsforschung Band 23). Bonn: Berlin. Zugriff am 14.08.2010 <http://www.bmbf.de/pub/Bildungsmisserfolg.pdf>
- Bundesministerium für Unterricht, Kunst und Kultur – bm:ukk (2010, März). *Allgemeinbildende höhere Schulen (AHS)*. Zugriff am 04.11.2010 <http://www.bmukk.gv.at/schulen/bw/abs/ahs.xml>
- Bundesministerium für Unterricht, Kunst und Kultur – bm:ukk (2010, April). *Bildungswesen in Österreich*. Zugriff am 16.07.2010 <http://www.bmukk.gv.at/schulen/bw/index.xml>
- Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980*. The College Board Research Report, 2001-2. Zugriff am 13.06.2010 http://professionals.collegeboard.com/profdownload/pdf/rdreport200_3919.pdf
- Cakir, H., Delialioğlu, O., Dennis, A., & Duffy, T. (2009). Technology Enhanced Learning Environments for Closing the Gap in Student Achievement Between Regions: Does it Work? *AACE Journal*, 17 (4), 301-315.
- Cohen, J. (1988). *Statistical Power Analyses for the Behavioral Sciences* (2 ed.). Hillsdale: Lawrence Erlbaum Associates.
- Deidesheimer Kreis (Hrsg.). (1997). *Hochschulzulassung und Studieneignungstests. Studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus-clausus – und andere Studiengänge*. Göttingen: Vandenhoeck & Ruprecht.

-
- Dell'mour, R. & Landler, F. (2002). *Akademische Grade zwischen Traum und Wirklichkeit. Einflussfaktoren auf den Studienerfolg*. Schriften des Instituts für Demographie 17. Wien: Institut für Demographie der Österreichischen Akademie der Wissenschaften.
- Demaray, M. K., & Elliott, S. N. (1998). Teachers' Judgments of Students' Academic Functioning: A Comparison of Actual and Predicted Performances. *School Psychology Quarterly*, 13 (1), 8-24.
- Deutsches Institut für Normung e.V. (2002). *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. DIN 33430*. Berlin: Beuth.
- DuBois, P. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.
- Fassmann, H. (2002). Räumliche Disparitäten im österreichischen Schulsystem – Strukturen, Trends und politische Implikationen. *ÖROK-Schriftenreihe*, 162.
- Fischer, G. H. (1995). Derivations of the Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models* (pp. 15-38). New York: Springer.
- Geißler, R. (1990). Schichten in der postindustriellen Gesellschaft. Die Bedeutung des Schichtbegriffs für die Analyse unserer Gesellschaft. In Peter A. Berger & Stefan Hardil (Hrsg.), *Lebenslagen, Lebensläufe, Lebensstile*. Soziale Welt, Sonderband, S. 81-102.
- Haider, G. & Reiter, C. (2004). *PISA 2003-Internationaler Vergleich von Schülerleistungen. Nationaler Bericht*. Graz: Leykam.
- Haidinger, G., Frischenschlager, O., & Mitterauer, L. (2005). Prediction of success in the first-year exam in the study of medicine – a prospective survey. *Wiener Klinische Wochenschrift*, 117 (23-24), 827-838.
- Hänsgen, K.-D. & Spicher, B. (2000). EMS und Chancengleichheit: Zwei Jahre Numerus clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 2. *Schweizerische Ärztezeitung*, 81 (14), 723-730.
- Hänsgen, K.-D. & Spicher, B. (2001). *EMS – Eignungstest für das Medizinstudium in der Schweiz. Bericht über die Durchführung und Ergebnisse 2001. Trends und Schlussfolgerungen aus 4 Jahren Testanwendung*. (Bericht No. 8). Freiburg: Zentrum für Testentwicklung und Diagnostik.
- Hänsgen, K.-D. & Spicher, B. (2004). *EMS – Eignungstest für das Medizinstudium 2004: Bericht 10 über die Durchführung und Ergebnisse 2004* (Bericht No. 10). Fribourg: Zentrum für Testentwicklung und Diagnostik.
- Hänsgen, K.-D. & Spicher, B. (2006). *EMS – Eignungstest für das Medizinstudium 2006: Bericht 12 über die Durchführung und Ergebnisse 2006* (Bericht No. 12). Fribourg: Zentrum für Testentwicklung und Diagnostik.
- Hänsgen, K.-D. & Spicher, B. (2007). *EMS – Eignungstest für das Medizinstudium 2007 – Bericht 13 über die Durchführung und Ergebnisse 2007* (Bericht No. 13). Fribourg: Zentrum für Testentwicklung und Diagnostik.
- Hartung, J., Elpelt, B. & Klösener, K.-H. (2005). *Statistik. Lehr- und Handbuch der angewandten Statistik* (14. Aufl.). München: Oldenbourg.

- Heine, C., Bridis, K., Didi, H.-J., Hasse, K. & Trost, G. (2006). *Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern. Eine Bestandsaufnahme*. HIS-Kurzinformation A 3/2006. Hannover: Hochschulinformationssystem.
- Hell, B., Trapmann, S., Weigand, S., Hirn, J. O. & Schuler, H. (2005, September). *Die Validität von Prädiktoren des Studienerfolgs – eine Metaanalyse*. Vortrag auf der 4. Tagung der Fachgruppe Arbeits- und Organisationspsychologie der Deutschen Gesellschaft für Psychologie in Bonn, 19.-21.09.2005. Zugriff am 27.02.2010 https://www.uni-hohenheim.de/studieneignung/publikationen/metaanalyse_verfahren_ao_2005.pdf
- Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5.Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 7, 77-86.
- Helmke, A. & Schrader, F.-W. (2006). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 83-94). Weinheim: Beltz.
- Herkner, W. (2001). *Lehrbuch Sozialpsychologie* (2. unveränderte Auflage). Bern: Hans Huber.
- Hoge, R. D., & Coladrci, T. (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research*, 59 (3), 297-313.
- Hohner, H.-U. (1997). *Professioneller Wandel in der Medizin als Herausforderung für die psychologische Eignungsdiagnostik* (Habilitationvortrag). Berlin: Institut für Arbeits-, Organisations- und Gesundheitspsychologie an der FU Berlin.
- Hossiep, R. (1995). *Berufseignungsdiagnostische Entscheidungen*. Göttingen: Hogrefe.
- Hradil, S. (1987). *Sozialstrukturanalyse in einer fortgeschrittenen Gesellschaft. Von Klassen und Schichten zu Lagen und Milieus*. Wiesbaden: VS Verlag.
- Ingenkamp, K. (1989). Forschungsstand und „Restauration“ der Zensurengebung. In K. Ingenkamp (Hrsg.), *Diagnostik in der Schule* (S. 55-94). Weinheim: Beltz.
- Ingenkamp, K. & Lissmann, U. (2005). *Lehrbuch der Pädagogischen Diagnostik* (5. Aufl.). Weinheim: Beltz.
- Jäger, R. S. (2005). Zur Auswahl von Studierenden – einige Gedanken und Bedenken. *Psychologische Rundschau*, 56 (2), 144-146.
- Jäger, R. S. & Lissmann, U. (2004). *Von der Beobachtung zur Notengebung*. Landau: Verlag Empirische Pädagogik.
- Kellermann, P. (1991). Studienmotive und Arbeitsperspektiven '90. Ein Forschungsbericht. In H. Guggenberger (Hrsg.), *Hochschulzugang und Studienwahl. Empirische und theoretische Ergebnisse von Hochschulforschung* (S. 169-281). Klagenfurt: Kärntner Druck- und Verlagsgesellschaft m.b.H.
- Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex Differences in Math Performance: The Role of Children's Approach to Schoolwork. *Developmental Psychology*, 42 (1), 11-26.
- Kersting, M. (2003). Grundrate. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 183-186). Weinheim: Beltz, PVU.

-
- Klieme, E. (1991). Problemstellung: Fairness von TMS-Aufgaben. In G. Trost (Hrsg.), *Test für medizinische Studiengänge (TMS): Studien zur Evaluation (15. Arbeitsbericht)*. Bonn: Institut für Test- und Begabungsforschung.
- Klieme, E., Eichler, W., Helmke, A., Lehmann, R. H., Nold, G., Rolff, H.-G. et al. (2006). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Deutsch Englisch Schülerleistungen International (DESI)*. Zugriff am 17.07.2010 http://www.dipf.de/de/pdf-dokumente/projekte-materialien/desi-zentrale-befunde/at_download/file
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Average*. College Board Research Report, 2008-5. New York: The College Board.
- Kolland, F. & Morgeditsch, W. (2007). *Evaluierung der Auswirkung des §124b des Universitätsgesetzes 2002: Bericht an den Nationalrat* [Bericht]. Wien: Bundesministerium für Wissenschaft und Forschung.
- Köller, O. & Baumert, J. (2001). Leistungsgruppierungen in der Sekundarstufe I. Ihre Konsequenzen für Mathematikleistungen und das mathematische Selbstkonzept der Begabung. *Zeitschrift für Pädagogische Psychologie*, 15 (2), 99-110.
- Köller, O. & Baumert, J. (2002). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (5., vollst. überarb. Aufl.) (S. 756-786). Weinheim: Beltz PVU.
- Köller, O., Baumert, J., Cortina, K. S., Trautwein, U. & Watermann, R. (2004). Öffnung von Bildungswegen in der Sekundarstufe II und die Wahrung von Standards: Analysen am Beispiel der Englischleistungen von Oberstufenschülern an integrierten Gesamtschulen, beruflichen und allgemein bildenden Gymnasien. *Zeitschrift für Pädagogik*, 50 (5), 679-700.
- Konegen-Grenier, C. (2002). Studierfähigkeit und Hochschulzugang. Professoren bewerten Studienanfänger. *Forschung und Lehre* 9, 481-483. http://www.forschung-und-lehre.de/wordpress/Archiv/2002/09_2002.pdf
- Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie. Ein Abriss samt neuesten Beiträgen* (S. 19-83). Weinheim: Beltz.
- Kubinger, K. D. (2003). Probabilistische Testtheorie (1. Aufl.). In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Linert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: PVU.
- Machold, I. & Tamme, O. (2009). Land am Rand? Soziale und wirtschaftliche Infrastrukturentwicklung im ländlichen Raum in Österreich. In Claudia Neu (Hrsg.), *Daseinsvorsorge. Eine gesellschaftswissenschaftliche Annäherung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Mitterauer, L., Frischenschlager, O. & Haidinger, G. (2007). *Sex differences in study progress at Medical University of Vienna*. GMS Z Med Ausbild, 24 (2), Doc111.
- Neuhofer, M. (2007). Regionale Verteilung von Bildungschancen in Oberösterreich. *WISO*, 30 (4), 93-108.
- Neumann, M., Schnyder, I., Trautwein, U., Niggli, A., Lüdtke, O. & Cathomas, R. (2007). Schulformen als differenzielle Lernmilieus. Institutionelle und kompositionelle Effekte auf die Leistungsentwicklung im Fach Französisch. *Zeitschrift für Erziehungswissenschaft*, 10 (3), 399-420.
- Österreichisches Normungsinstitut (2005). *Anforderungen an Prozesse und Methoden in der Personalauswahl und -entwicklung - ÖNORM D 4000*. Wien: Österreichisches Normungsinstitut.
- Ortner, T. M. (2003). Anforderungsprofil. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 20-23). Weinheim: Beltz.
- Randhawa, B. S. & Hunt, D. (1987). Sex and Rural-Urban Differences in Standardized Achievement Scores and Mathematics Subskills. *Canadian Journal of Education*, 12 (1), 137-151.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychological and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130 (2), 261-288.
- Roscigno, V. J., Tomaskovic-Devey, D., & Crowley, M. L. (2006). Education and the Inequalities of Place. *Social Forces*, 84 (4), 2121-2145.
- Sauer, J. & Gamsjäger, E. (1996). *Ist Schulerfolg vorhersehbar? Die Determinanten der Grundschulleistung und ihr prognostischer Wert für den Sekundarschulerfolg*. Göttingen: Hogrefe.
- Schäfer, U. & Weiß, M. (2007). Organisation der Schulsysteme und Leitthemen ihrer aktuellen Entwicklung. In Bundesministerium für Bildung und Forschung (Hrsg.), *Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten* (Bd. 2, S. 37-75). Bonn, Berlin: BMBF.
- Schmid, K. (2004). Regionale Bildungsströme in Österreich. *ibw-Bildung & Wirtschaft*, 31.
- Schmidt-Atzert, L. (2004). Abschätzung und Prognosegüte von berufsbezogenen Eignungsbeurteilungen und darauf aufbauenden Entscheidungen. In K. Westhoff, L. J. Hellfritsch, L. F. Hornke, K. D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel & G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430*. Lengerich: Pabst.
- Schmitt, M. (2005). Auswahl von Studierenden: Historie, Zweck und Aufbau dieses Diskussionsforums. *Psychologische Rundschau*, 56 (2), 123-124.
- Schrader, F. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45-58). Weinheim: Beltz.
- Schreiner, C. & Schwantner, U. (Hrsg.). (2009). *PISA 2006. Österreichischer Expertenbericht zum Naturwissenschafts-Schwerpunkt*. Graz: Leykam.

-
- Schuler, H. (1991). Auswahl von Mitarbeitern. In R. v. Rosenstiel, E. Regnet & M. Domsch (Hrsg.), *Führung von Mitarbeitern. Handbuch für erfolgreiches Personalmanagement* (S. 100-125). Stuttgart: Schäffer.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die Berufseignungsdiagnostik*. Göttingen: Verlag für Angewandte Psychologie.
- Schuler, H. (2006). Noten als Prädiktoren von Studien- und Berufserfolg. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 535-541). Weinheim: Beltz.
- Schüpbach, H. & Klauer, C. (2005). Studierendenauswahl: Beiträge der Psychologie zu einem universitären Pilotprojekt. *Psychologische Rundschau*, 56 (2), 137-138.
- Schwabe, M. & Gumpoldsberger, H. (2008). Regionale Disparitäten der Verteilung der Schüler und Schülerinnen auf Schultypen in der Sekundarstufe. *Statistische Nachrichten*, 12, 1122-1131.
- Schwarz, F., Spielauer, M., & Städtner, K. (2003). Gender, Regional and Social Differences at the transition from Lower to Upper Secondary Education. Working Paper 23 (ÖIF).
- Spiel, C., Litzenberger, M. & Haiden, D. (2007). Bildungswissenschaftliche und psychologische Aspekte von Auswahlverfahren. In C. Badelt, W. Wegscheider & H. Wultz (Hrsg.), *Hochschulzugang in Österreich* (S. 479-552). Graz: Grazer Universitätsverlag.
- Spiel, C., Schober, B. & Litzenberger, M. (2008, Jänner). *Evaluation der Eignungstests für das Medizinstudium in Österreich*. Projektbericht für das Bundesministerium für Wissenschaft und Forschung. Wien.
- Spielauer, M., Schwarz, F., & Schmid, K. (2002). *Education and the Importance of the First Educational Choice in the Context of the FAM-SIM+ Family Microsimulation Model for Austria*. Working Paper 15 (ÖIF). http://131.130.67.132/ftp/projekte/famsim/wp_15_famsim_education.pdf
- Statistik Austria (2009, Februar). *Steiniger Weg bis zur Matura*. Zugriff am 10.06.2009 http://www.statistik.at/web_de/dynamic/statistiken/bildung_und_kultur/035560
- Statistik Austria (2010, März). *Endgültige Bevölkerungszahl für die Finanzjahre 2009 und 2010 je Gemeinde*. Zugriff am 05.04.2010 http://www.statistik.at/web_de/static/endgueltige_bevoelkerungszahl_fuer_die_finanzjahre_2009_und_2010_je_gemein_042283.pdf
- Statistik Austria (2010, April). *Formales Bildungswesen*. Zugriff am 05.11.2010 http://www.statistik.at/web_de/statistiken/bildung_und_kultur/formales_bildungswesen/index.html
- Statistisches Bundesamt Deutschland (2010a). *Bevölkerungsstand*. Zugriff am 05.04.2010 <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsstand.psm1>
- Statistisches Bundesamt Deutschland (2010b). *Gemeinden mit 5 000 und mehr Einwohnern nach Fläche und Bevölkerung. Gebietsstand: 31.12.2008*. Zugriff am 05.04.2010 http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Regionales/Gemeindeverzeichnis/Administrativ/Aktuell/07_GemeindenAktuell,property=file.xls

- Stemmler, G. (2005). Studierendenauswahl durch Hochschulen: Ungewisser Nutzen. *Psychologische Rundschau*, 56 (2), 125-127.
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23 (2), pp. 565-578.
- Tent, L. (2006). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 873-880). Weinheim: Beltz.
- Testkuratorium der Föderation deutscher Psychologenverbände (1986). Beschreibung der einzelnen Kriterien für die Testbeurteilung. *Diagnostica*, 32, 358-360.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21 (1), 11-27.
- Trost, G. (1994). *Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht)*. Bonn: Institut für Test- und Begabungsforschung.
- Trost, G. (1996). Eignungskriterien bei der Zulassung zum Medizinstudium in Europa: Ergebnisse einer Erhebung in allen europäischen Ländern. In K.-D. Hänsgen & N. Ischi (Hrsg.), *Eignungsprüfung für das Medizinstudium Kriterien und Testverfahren*. Bericht über das Internationale Symposium in Bern am 8. November 1996. Zentrum für Testentwicklung und Diagnostik am Psychologischen Institut der Universität Fribourg. Zugriff am 04.11.2010 <http://www.unifr.ch/ztd/ems/berichte/b3/eignungskriterien.htm>
- Trost, G. (2005). Studierendenauswahl durch die Hochschulen: Welche Verfahren kommen prinzipiell in Betracht, welche nicht? *Psychologische Rundschau*, 56 (2), 138-140.
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M., et al. (1998). *Evaluation des Tests für medizinische Studiengänge (TMS): Synopse der Ergebnisse* [Bericht]. Bonn: Institut für Test- und Begabungsforschung.
- Unger, M. & Wroblewski, A. (2006). *Studierenden-Sozialerhebung 2006. Bericht zur sozialen Lage der Studierenden*. Wien: Institut für Höhere Studien.
- Wagner, P. & Spiel, C. (2002). Zeitinvestment und Lerneffektivität: Eine Analyse in Hauptschule und Gymnasium hinsichtlich Persönlichkeitsvariablen, Arbeitshaltung und Bedingungsfaktoren. *Empirische Pädagogik*, 16, 357-381.
- Wallner-Paschon, C. (2009). Notengerechtigkeit bei der Risiko- und Spitzengruppe. In B. Suchan, C. Wallner-Paschon & C. Schreiner (Hrsg.), *PIRLS 2006. Die Lesekompetenz am Ende der Volksschule - Österreichischer Expertenbericht*. Graz: Leykam.
- Weishaupt, H. (2009). Bildung und Region. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (2., überarb. u. erw. Aufl.). (S. 217-231). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wild, K.-P. & Krapp, A. (2001). Pädagogisch-psychologische Diagnostik. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch*. München: Beltz.
- Wroblewski, A. & Unger, M. (2003). *Studierenden-Sozialerhebung 2002. Bericht zur sozialen Lage der Studierenden*. Wien: Institut für Höhere Studien.

Young, D. J. (1998). Rural an Urban Differences in Student Achievement in Science an Mathematics: A Multilevel Analysis. *School Effectiveness and School Improvement*, 9 (4), 386-418.

Zentrum für Testentwicklung und Diagnostik. (2007). *TESTINFO '07. Eignungstest für das Medizinstudium (EMS). Version A*. [Broschüre]. Hänsgen, K.-D.: Autor.

Kurzfassung

Die vorliegende Untersuchung geht den Fragen nach ob 1) schulische Aspekte (Schulnoten und Schultyp), 2) die regionale Herkunft sowie 3) Geschlecht und Nationalität die Leistungen im Eignungstest für das Medizinstudium (EMS) erklären können. Zur Steigerung des Informationsgehalts der Schulnoten wird zusätzlich zum Notendurchschnitt das Ausmaß der Homogenität des Notenprofils untersucht. Eine weitere Fragestellung dieser Studie untersucht, ob der Ausschluss von nicht verrechnungsfairer Aufgaben zu einer besseren Prognosekraft der Erklärungsvariablen führt.

Der EMS ist das Kernstück im Auswahlverfahren zum Studium der Medizin an den Medizinischen Universitäten Wien und Innsbruck. Die Daten der hier vorliegenden Studie basieren auf der Evaluationsstudie von Spiel, Schober und Litzenberger (2008) im Auftrag des Bundesministeriums für Wissenschaft und Forschung. Zu diesem Zweck wurde am 06. Juli 2007 eine empirische Erhebung mit allen Studienbewerberinnen und Studienbewerbern durchgeführt. Zusätzlich zu den Testergebnissen und soziodemographischen Daten der Bewerberinnen und Bewerber wurde ein eigens hierfür konstruierter Fragebogen im Anschluss an den EMS vorgegeben.

Die Ergebnisse zeigen, dass die Schulnoten (Einzelfachnoten und Durchschnittsnote) und der Schultyp einen Beitrag zur Aufklärung der Leistungsunterschiede zwischen den Bewerberinnen und Bewerbern leisten können. Dies gilt für den Gesamtestwert als auch für die Untertests. Von den Einzelfachnoten liefert die Mathematiknote den größten Erklärungswert für die Testleistungen. Geschlecht und Nationalität können die Testleistungen am besten erklären. Als Ursachen hierfür sprechen primär Einflüsse aus unterschiedlichen Sozialisierungen. Der Einbezug der Homogenität des Notenprofils sowie die regionale Herkunft liefern keine zusätzlichen Erklärungen für die Leistungsunterschiede zwischen den Personen. Durch den Ausschluss nicht verrechnungsfairer Aufgaben kann die Prognosekraft der Erklärungsvariablen nicht gesteigert werden.

Für zukünftige Studien wird empfohlen weitere Prädiktoren zur Erklärung der Testleistungen in die Analysen aufzunehmen.

Schlüsselwörter: Evaluation, Akademische Eignung, Schulnoten, Geschlechtsunterschiede, Studierendenauswahl, Medizinstudenten

Abstract

The present study investigates if 1) school factors (grades and kind of school), 2) regional provenance, and 3) sex and nationality account for achievements in the aptitude test for medicine (EMS). To enhance the information content of the grades, the measure of the homogeneity of the grades-profile will be included in addition to the grade point average. A further question of this study is to investigate the effect of excluding items not in conformance with scaling fairness. It is expected that the exclusion increases the power of the predictors.

The EMS is the key part of the selection process for the study of human medicine at the Medical University of Vienna and Innsbruck Medical University. The data for this study are based on the evaluation of Spiel, Schober, and Litzenberger (2008) conducted at July 6th, 2007 by order of the Austrian Federal Ministry of Science and Research. For this purpose, an empirical study with all applicants for a university place has been performed. In addition to the test achievements and sociodemographic data, the applicants had to fill in a specifically designed questionnaire.

The results show that grades (grades and grade point average) and kind of school significantly account for the achievement differences between the applicants. These effects are valid for both, Testscore and Subtests. Mathematics shows the highest effect size of all grades. Of all predictors, sex and nationality show the highest effects. The reason for this evidence is probably caused by different socializations. No evidence has been found for the measure of homogeneity of the grades-profile and the regional provenance. The exclusion of scaling unfair items did not increase the power of the predictors.

For future studies, it is recommended to include more predictors for explaining the achievements of the EMS.

Keywords: Evaluation, Academic Aptitude, Grades, Human Sex Differences, Selection of Students, Medical Students

A. Anhang

A.1. Auswahlverfahren für den Hochschulzugang

Die Analyse des Auswahlverfahrens im Studienjahr 2006/07 an den beiden Medizinischen Universitäten (Wien und Innsbruck) zeigt eine deutliche Reduktion der Studienbewerberinnen und Studienbewerber von der Anmeldung bis zur Zulassung siehe Tabelle A.1 (vereinfacht aus Kolland & Morgeditsch, 2007). Gesamt gab es in Wien und Innsbruck 5611 Bewerberinnen und Bewerber von denen letztendlich nach der Voranmeldung und nach dem Test 1190 zugelassen wurden. Dies entspricht einem Anteil (einer Selektionsrate) von rund 21 Prozent. Zur Berechnung der prognostischen Vali-

Tabelle A.1.: Reduktion der Bewerberinnen und Bewerber im Verlauf der Anforderungsschritte an den Medizinischen Universitäten Wien und Innsbruck 2006/07 (vereinfacht aus Kolland & Morgeditsch, 2007)

	Wien	Innsbruck	Wien+Innsbruck
Persönliche Anmeldung	3429	2182	5611
Teilnahme am EMS	2230	1415	3645
Zulassung	829	361	1190
Erfolgreiche (zugelassen)	24.2%	16.5%	21.2%

dität ist die Basisrate notwendig. Diese ist in der Praxis nur sehr schwierig zu bestimmen, es genügt jedoch eine Abschätzung ihrer Höhe (Spiel et al., 2007). In Anlehnung an Stemmler (2005) wurde ebenfalls eine Basisrate von 60% bei der Hochschulauswahl angenommen. Mit steigender Basisrate steigt die Wahrscheinlichkeit der Auswahl einer geeigneten Person. Mit Hilfe der Taylor-Russell-Tafeln (Taylor & Russell, 1939; Lienert & Raatz, 1998, S. 421) lässt sich der Prozentsatz der Geeigneten nach der Selektion ermitteln.

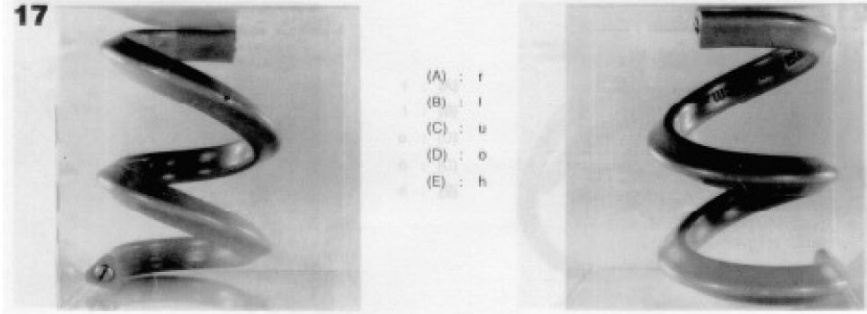
Für eine ermittelte Selektionsrate von 21% und einer Basisrate von 60% ergeben sich für verschiedene Validaten des Selektionsverfahren, die in Tabelle A.2 dargestellten Werte Geeigneter und Ungeeigneter unter den Selektierten. Will man maximal 20% Ungeeigneter nach dem Auswahlverfahren unter den Selektierten haben, so müsste das Verfahren bei 21% Selektionsrate eine prognostische Validität von mindestens $r = .4$ aufweisen.

Tabelle A.2.: Prozentsatz Geeigneter und Ungeeigneter für variierende Validitäten bei einer Basisrate von 60% und einer Selektionsrate von 21%.

Validität des Selektionsverfahrens	Unter den Selektierten sind	
	Geeignete	Ungeeignete
0.10	65.3%	34.7%
0.20	70.4%	29.6%
0.30	75.5%	24.5%
0.40	80.4%	19.6%
0.50	85.3%	14.7%
0.60	89.9%	10.1%
0.70	94.1%	5.9%

A.1.1. Überblick über die Untertests des EMS nach Hänsgen und Spicher, 2007

Vorgegebene Untertests am Vormittag
1. Quantitative und formale Probleme
Dieser Untertest prüft die Fähigkeit, im Rahmen medizinischer und naturwissenschaftlicher Fragestellungen mit Zahlen, Größen, Einheiten und Formeln richtig umzugehen.
<p>Beispielaufgabe: Eine Broteinheit (BE) ist definiert als diejenige Nahrungsmenge in Gramm, die 12 Gramm Kohlenhydrate enthält. Bei der Verbrennung von 1 g Kohlenhydraten im Organismus werden 16 Kilojoule (kJ) an Energie frei. Ein Patient, der auf Diät gesetzt ist, soll pro Tag 4800 kJ zu sich nehmen, ein Fünftel davon in Kohlehydraten.</p> <p>Wie viele sind dies täglich? (A) 60 BE (B) 25 BE (C) 6 BE (D) 5 BE (E) 0,5 BE</p>

2. Schlauchfiguren
Die Aufgaben dieses Untertests prüfen das räumliche Vorstellungsvermögen.
<p>Beispielaufgabe: Jede Aufgabe besteht aus zwei Abbildungen eines durchsichtigen Würfels, in dem sich ein, zwei oder drei Kabel befinden. Die erste Abbildung (links) zeigt stets die Vorderansicht des Würfels; auf dem rechten Bild daneben, in welchem derselbe Würfel noch einmal abgebildet ist, soll die Testteilnehmerin oder der Testteilnehmer heraus finden, ob die Abbildung die Ansicht von rechts (r), links (l), unten (u), oben (o) oder von hinten (h) zeigt.</p>
<p>17</p>  <p style="margin-left: 100px;">(A) : r (B) : l (C) : u (D) : o (E) : h</p>
<div style="text-align: center;"> <p>Hier sehen Sie den Würfel von vorne!</p> </div> <div style="text-align: center;"> <p>Hier sehen Sie den Würfel von ? (hinten!)</p> </div>

3. Textverständnis

In diesem Untertest wird die Fähigkeit geprüft, umfangreiches und komplexes Textmaterial aufzunehmen und zu verarbeiten.

Beispielaufgabe:

Zu den Aufgaben der Schilddrüse gehören Bildung, Speicherung und Freisetzung der jodhaltigen Hormone Trijodthyronin (T3) und Thyroxin (T4). In der Schilddrüse befinden sich zahlreiche Hohlräume, Follikel genannt, deren Wände von einer Schicht so genannter Epithelzellen gebildet werden. Diese Follikel sind mit einer Substanz gefüllt, in der die Hormone T3 und T4 als inaktive Speicherformen enthalten sind. Beim Menschen ist in den Follikeln so viel T3 und T4 gespeichert, dass der Organismus damit für etwa 10 Monate versorgt werden kann. Das für die Hormonbildung erforderliche Jod entstammt der Nahrung und wird von den Epithelzellen als Jodid aus dem Blut aufgenommen. Die Jodidaufnahme erfolgt an der äußeren Zellmembran der Epithelzellen durch eine sogenannte Jodpumpe. Diese wird durch ein Hormon aus der Hirnanhangsdrüse, das TSH, stimuliert und kann pharmakologisch durch die Gabe von Perchlorat gehemmt werden. Ferner gibt es erbliche Schilddrüsenerkrankungen, bei deren Vorliegen die Jodpumpe nicht funktioniert. Bei Gesunden wird das in die Epithelzellen aufgenommene Jodid im nächsten Schritt unter dem Einfluss eines Enzyms in freies Jod umgewandelt und in die Follikel abgegeben. Die Aktivität dieses Enzyms kann ebenfalls pharmakologisch gehemmt werden. Die letzten Schritte der Hormonbildung finden in den Follikeln, also außerhalb der einzelnen Epithelzellen, statt. In dort vorhandene so genannte Tyrosin-Reste (des Thyreoglobulins) wird zunächst ein Jodatom eingebaut. So entstehen Monojodtyrosin-Reste (MIT), von denen ein Teil durch die Bindung je eines weiteren Jodatoms in Dijodtyrosin-Reste (DIT) umgewandelt wird. Durch die Verknüpfung von je zwei DIT-Resten entsteht schliesslich T4, während aus der Verbindung je eines MIT-Restes mit einem DIT-Rest T3 hervorgeht. T3 und T4 werden dann in den Follikeln gespeichert und bei Bedarf über die Epithelzellen ins Blut freigesetzt. Diese Freisetzung von T3 und T4 ins Blut (Sekretion) wird über die Hirnanhangsdrüse und den Hypothalamus, einen Teil des Zwischenhirns, gesteuert: Das erwähnte Hormon TSH stimuliert ausser der Bildung auch die Sekretion von T3 und T4; es ist hinsichtlich seiner eigenen Sekretionsrate jedoch abhängig von der Stimulation durch das hypothalamische Hormon TRH. Die TRH-Sekretion wiederum wird z.B. durch Kälte stimuliert, während Wärme hemmend wirken kann. Neben diesen übergeordneten Steuerungsmechanismen existiert noch ein sogenannter Rückkoppelungsmechanismus: Eine hohe Konzentration von T3 und T4 im Blut hemmt die TSH- und die TRH-Sekretion, eine niedrige Konzentration stimuliert sie. Bei den an der Steuerung der Schilddrüsenhormon-Sekretion beteiligten Arealen von Hirnanhangsdrüse und Hypothalamus können krankheitsbedingte Störungen auftreten, die zu einer Über- oder Unterfunktion der Schilddrüse führen. Eine der Hauptwirkungen von T3 und T4 ist die Beeinflussung des Energieumsatzes durch eine Steigerung des Sauerstoffverbrauchs in stoffwechselaktiven Organen. Entsprechend senkt eine zu niedrige Konzentration der beiden Hormone im Blut (Hypothyreose) den Energieumsatz bzw. die Stoffwechselaktivität unter den normalen Wert, während bei einer zu hohen Konzentration (Hyperthyreose) die Stoffwechselaktivität gesteigert wird. Die Hormone T3 und T4 können ebenso wie TSH und TRH für diagnostische und therapeutische Zwecke synthetisch hergestellt werden.

Auf einen solchen Text folgen Fragen, die sich ausschließlich auf im Text vorhandene Inhalte beziehen; eine Frage mit niedrigem Schwierigkeitsgrad ist zum Beispiel so formuliert:

Welche der folgenden Aussagen über die TSH-Sekretion lässt bzw. lassen sich aus dem Text ableiten?

Die TSH-Sekretion kann unter bestimmten Bedingungen verändert werden durch...

1. Kälte.
2. T3- und T4- Gaben.
3. Über Jahre hinweg erfolgende Perchlorat-Gaben.

- Nur Aussage I lässt sich ableiten.
- Nur die Aussagen I und II lassen sich ableiten.
- Nur die Aussagen I und III lassen sich ableiten.
- Nur die Aussagen II und III lassen sich ableiten.
- Alle drei Aussagen lassen sich ableiten.

4. Planen und Organisieren

Dieser Untertest prüft anhand verschiedener Szenarien jene Fähigkeiten, die für eine effiziente Selbstorganisation im Studium wichtig sind.

Beispielszenario: Semesterarbeit

Sie haben die Aufgabe, im kommenden Wintersemester eine Semesterarbeit zu einem vorgegebenen Thema zu schreiben. Nach gründlichem Studium der einschlägigen Literatur sollen Sie in Ihrer Arbeit zunächst einen Überblick über die wichtigsten vorliegenden Ergebnisse zum Thema geben und dann einen eigenen Lösungsansatz entwickeln.

Die folgenden Termine liegen bereits fest:

- 11.10., 17.00 Uhr: Sie erhalten das Thema Ihrer Semesterarbeit, Literaturhinweise und weitere Instruktionen.
- 17.10.: Beginn der Vorlesungszeit
- 19.12. bis 1.1.: Weihnachtsferien (vorlesungsfreie Zeit)
- 11.2.: Ende der Vorlesungszeit
- 7.3., 8.00 Uhr: Abgabe der Semesterarbeit
- Während der Vorlesungswochen: Jede Woche von Montag bis Donnerstag Lehrveranstaltungen; an diesen Tagen können Sie sich nicht der Semesterarbeit widmen.
- In der ersten Hälfte der Weihnachtsferien (16. bis 24.12.): Urlaubsreise
- Zwei volle Wochen im unmittelbaren Anschluss an die Vorlesungszeit: Prüfungsvorbereitung (1 Woche) und Teilnahme an mündlichen Prüfungen (1 Woche)

Die folgenden Arbeitsschritte bzw. -phasen müssen Sie einplanen (sie sind hier nicht in chronologischer Abfolge aufgeführt):

- Zwei Treffen mit dem Betreuer Ihrer Arbeit: Eines unmittelbar vor dem Schreiben der Semesterarbeit, eines vor der abschließenden Überarbeitung. Dauer: Je ein voller Arbeitstag. (Der Betreuer ist nur dienstags und freitags verfügbar, und zwar von der dritten bis zur neunten Vorlesungswoche sowie nach dem Ende der Vorlesungszeit.)
- Literatursuche und -beschaffung (Dauer: 5 Tage)
- Erstellung des Konzepts der Semesterarbeit; die Literatur muss hier bereits verarbeitet sein (Dauer: 3 Tage)
- Korrekturlesen durch einen Freund (Überprüfung auf Tippfehler, unklare Formulierungen etc.) (Dauer: 3 Tage)
- Lesen und Zusammenfassen der Literatur (Dauer: 10 Tage)
- Schreiben der Semesterarbeit (Dauer: 18 Tage)
- Abschließendes Überarbeiten der Semesterarbeit (Dauer: 4 Tage)
- Kopieren der Arbeit (unmittelbar vor der Abgabe) (Dauer: 1 Tag)
- Ausführen der Korrekturvorschläge Ihres Freundes, bevor Sie die Arbeit mit dem Betreuer durchsprechen (Dauer: 1 Tag)
- Reserve für Unvorhergesehenes (Dauer: 1 Tag)

Eine Arbeitsphase muss stets beendet sein, bevor die nächste beginnt (Ausnahme: Korrekturlesen).

Ferner müssen Sie einplanen:

- An den Wochenenden arbeiten Sie – wie auch Ihr Freund – nur samstags (im Urlaub natürlich gar nicht).
- Am zweiten Weihnachtstag arbeiten Sie nicht.

Welche der folgenden Aussagen über die Semesterarbeit trifft bzw. treffen zu?

- I. Noch vor den Weihnachtsferien können Sie mit dem Schreiben der Semesterarbeit beginnen.
- II. Insgesamt werden Sie sich in 18 Wochen jeweils mindestens einen Tag lang mit der Semesterarbeit beschäftigen.

- (A) Nur Aussage I trifft zu.
- (B) Nur Aussage II trifft zu.
- (C) Beide Aussagen treffen zu.
- (D) Keine der beiden Aussagen trifft zu.

Welche der folgenden Aussagen über die Einbeziehung Ihres Freundes trifft bzw. treffen zu?

- I. Der günstigste Zeitraum, in dem Ihr Freund Ihre Arbeit Korrektur lesen kann, ist die erste Hälfte der letzten Vorlesungswoche.
 II. Angenommen, Ihr Freund teilt Ihnen mit, er habe leider nur in der Zeit vom 25. bis zum 27. Februar Gelegenheit zum Korrekturlesen; dann nützt Ihnen sein Angebot für den ursprünglich geplanten Zweck dieses Korrekturvorgangs nicht mehr.

- (A) Nur Aussage I trifft zu.
 (B) Nur Aussage II trifft zu.
 (C) Beide Aussagen treffen zu.
 (D) Keine der beiden Aussagen trifft zu.

5. Konzentriertes und Sorgfältiges Arbeiten

Mit diesem Test soll Ihre Fähigkeit, rasch, sorgfältig und konzentriert zu arbeiten erfasst werden.

Beispielaufgabe:

Sie sehen nachfolgend ein Blatt mit 40 Zeilen, die aus je 40 Buchstaben u und m gebildet werden.

Ihre Aufgabe ist es, zeilenweise jedes u zu markieren, VOR dem in der Zeile unmittelbar ein m steht:

m m

Sie dürfen kein u markieren, vor dem kein m steht. Sie dürfen natürlich auch kein m markieren. Beides wären Fehler. Nachfolgend sehen Sie ein richtig bearbeitetes Beispiel:

m m m m u u m m m m u u u m m m

Vorgegebene Untertests am Nachmittag

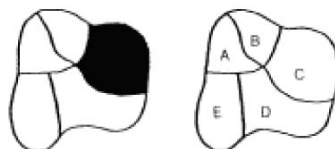
6. Figuren lernen - Einprägephase

Die Aufgaben dieses Tests prüfen, wie gut man sich Einzelheiten von Gegenständen einprägen und merken kann.

Beispielaufgabe:

Gezeigte Figur zum Einprägen

Gezeigte Figur beim Abfragen



Die Testperson hat vier Minuten Zeit, um sich 20 solcher Figuren einschließlich der Lage der schwarzen Flächen einzuprägen. Nach ca. einer Stunde muss sie angeben können, welcher Teil der Abbildung geschwärzt war, und dies direkt auf dem Antwortbogen eintragen.

7. Fakten lernen - Einprägephase

Die Aufgaben dieses Tests prüfen, wie gut man sich Fakten einprägen und merken kann.

Beispielaufgabe:

15 Patienten vorgestellt, von denen jeweils der Name, die Altersgruppe, Beruf und Geschlecht, ein weiteres Beschreibungsmerkmal (z.B. Familienstand) sowie die Diagnose erfahren wird.

Ein Beispiel für eine derartige Fallbeschreibung ist: Lemke, 30 Jahre, Dachdecker, ledig, Schädelbasisbruch

Eine Frage zum obigen Beispiel könnte z.B. lauten: Der Patient mit dem Schädelbasisbruch ist von Beruf

...

- (A) Installateur
 (B) Lehrer
 (C) Dachdecker
 (D) Handelsvertreter
 (E) Physiker

8. Medizinisch-naturwissenschaftliches Grundverständnis

In diesem Untertest wird das Verständnis für Fragen der Medizin und der Naturwissenschaften geprüft.

Beispielaufgabe:

Im Kindesalter kann das Zentrum für Sprache, Spracherwerb und Sprachverständnis noch in der linken oder in der rechten Hälfte (Hemisphäre) des Gehirns in einem umschriebenen Hirnrindengebiet (sog. Sprachregion) angelegt werden. Spätestens im zwölften Lebensjahr sind die sprachlichen Fähigkeiten jedoch fest in einer der beiden Hemisphären verankert, und zwar bei den Rechtshändern in der Regel links, bei den Linkshändern in der Mehrzahl ebenfalls links, zum Teil aber auch rechts; die korrespondierende Region der Gegenseite hat zu diesem Zeitpunkt bereits andere Funktionen fest übernommen. Welche der nachfolgenden Aussagen lässt bzw. lassen sich aus diesen Informationen ableiten?

Bei irreversiblen Hirnrindenverletzungen im Bereich der so genannten Sprachregion der linken Hemisphäre ...

1. kommt es bei erwachsenen Linkshändern in der Regel zu keinen wesentlichen Sprachstörungen.
2. kommt es bei einem Vorschulkind in der Regel zu einer bleibenden Unfähigkeit, die Muttersprache wieder zu erlernen.
3. ist bei zwanzigjährigen Rechtshändern die Fähigkeit, eine Sprache zu erlernen, in der Regel verloren gegangen.

- (A) Nur Ausfall I ist zu erwarten.
 (B) Nur Ausfall II ist zu erwarten.
 (C) Nur Ausfall III ist zu erwarten.
 (D) Nur die Ausfälle I und III sind zu erwarten.
 (E) Nur die Ausfälle II und III sind zu erwarten.

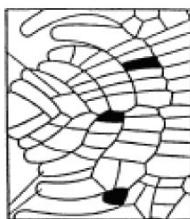
Untertests *Figuren Lernen & Fakten Lernen - Abrufphase*

9. Muster zuordnen

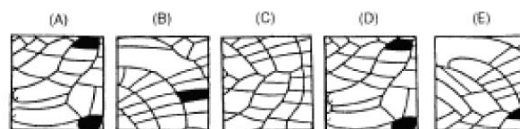
Im vorletzten Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wieder zu erkennen.

Beispielaufgabe:

Muster



Musterausschnitte



10. Diagramme und Tabellen

Mit dieser Aufgabengruppe wird die Fähigkeit geprüft, Diagramme und Tabellen richtig zu analysieren und zu interpretieren. In dieser Form werden während des Studiums zahlreiche Zusammenhänge vermittelt.

Beispielaufgabe:

Die folgende Tabelle beschreibt die Zusammensetzung und den Energiegehalt von vier verschiedenen Milcharten. Unter Energiegehalt der Milch verstehen wir dabei die Energiemenge, gemessen in Kilojoule (kJ), welche 100 Gramm (g) Milch dem Organismus ihres Konsumenten liefern können.

Milchart	Eiweiss	Fett	Milchzucker	Salze	Energiegehalt
menschliche Muttermilch	1,2 g	4,0 g	7,0 g	0,25 g	294 kJ
Vollmilch	3,5 g	3,5 g	4,5 g	0,75 g	273 kJ
Magermilch	3,3 g	0,5 g	4,5 g	0,75 g	160 kJ
Buttermilch	3,0 g	0,5 g	3,0 g	0,55 g	110 kJ

Welche Aussage lässt sich aus den gegebenen Informationen nicht ableiten?

- (A) Menschliche Muttermilch enthält mehr als doppelt soviel Milchzucker wie Buttermilch.
- (B) Vollmilch enthält im Vergleich zur menschlichen Muttermilch etwa die dreifache Menge an Salzen und Eiweiß.
- (C) Zur Aufnahme der gleichen Energiemenge muss ein Säugling fast dreimal soviel Buttermilch wie Muttermilch trinken.
- (D) Der Unterschied zwischen Magermilch und Vollmilch ist bei der Mehrzahl der aufgeführten Merkmale geringer als der Unterschied zwischen Magermilch und Buttermilch.
- (E) Der Eiweissgehalt der Milch ist für den Energiegehalt von entscheidender Bedeutung.

A.2. Struktur des deutschen Bildungssystems

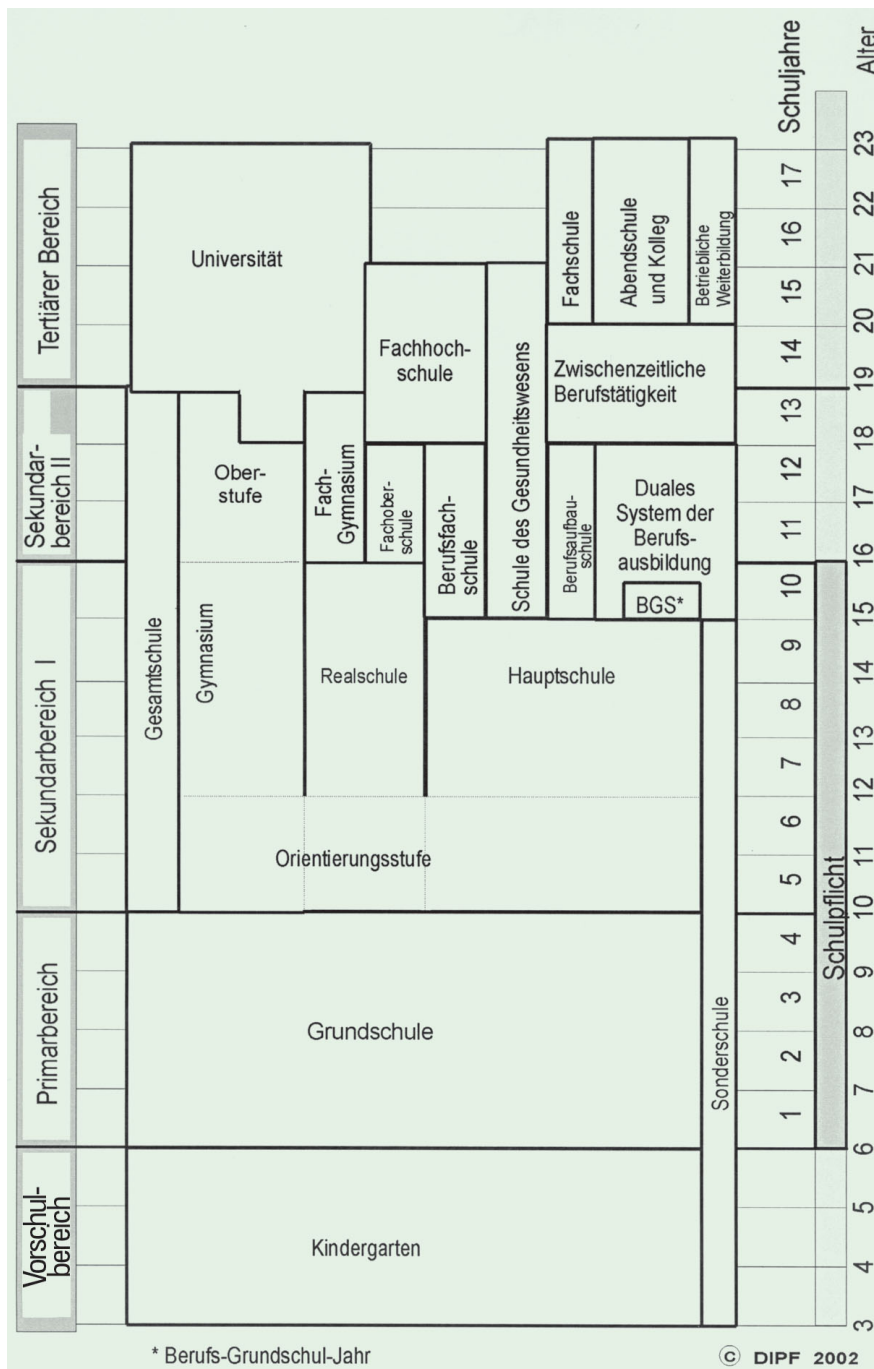


Abbildung A.1.: Struktur des deutschen Bildungssystems (Schäfer & Weiß, 2007).

A.3. Fragebogen

+

Liebe Kandidatinnen und Kandidaten,
Ziel dieses Fragebogens ist es, die Ergebnisse des Auswahlverfahrens genauer zu analysieren und daraus Optimierungsmöglichkeiten abzuleiten. Wir bitten Sie daher um ehrliche und vollständige Antworten. Der Fragebogen wurde von einer Arbeitsgruppe für Evaluation entwickelt. Es wird gewährleistet, dass die Auswertung völlig anonym erfolgt und unabhängig von der Entscheidung über Ihre Aufnahme ins Medizinstudium ist.

In welchem **Schultyp** haben Sie Ihre Matura/ Ihr Abitur gemacht?
(Bitte keine Mehrfachnennungen)

- Naturwissenschaftliches Gymnasium
- Gymnasium mit anderem Schwerpunkt (z.B. humanistisch, sprachlich)
- Gymnasium f. Berufstätige (Studienberechtigungsprüfung)
- Berufsbildende höhere Schule mit technischem Schwerpunkt (z.B. HTL)
- Berufsbildende höhere Schule ohne techn. Schwerpunkt (z.B. HAK, HBLA)
- Fach-/Berufsoberschule/Kollegschule (gilt nur für Deutschland)
- Berufliches Gymnasium (gilt nur für Deutschland)
- Gesamtschule
- anderer Schultyp mit mathemat.-naturwissenschaftl. Schwerpunkt
- anderer Schultyp ohne mathemat.-naturwissenschaftl. Schwerpunkt

Wo haben Sie den Großteil Ihrer **Schulzeit** im Alter zwischen 14 und 19 Jahren verbracht?

- Großstadt (> 50.000 Einwohner)
- mittel große Stadt (10.000-50.000 Einwohner)
- Kleinstadt bzw. ländliche Region (< 10.000 Einwohner)

Für Wien und Innsbruck:
EMS-Ettikett bitte HIER aufkleben!
Für Graz, Wien und Innsbruck:
Bearbeitungsnummer im Feld unten eintragen und darunter nochmals die entsprechenden Ziffern ankreuzen!

1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1
2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3
4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4
5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7
8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8
9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9
0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0

Welche Sprache wird/wurde in Ihrem Elternhaus gesprochen (**Muttersprache**)?

- deutsch
- türkisch
- ex-jugoslawische Sprachen
- sonstige Sprachen

Bitte geben Sie den **höchsten Schulabschluss Ihrer Eltern** bzw. der ersten Bezugspersonen an!

VATER:

- ohne abgeschlossene Schulausbildung
- Pflichtschule (Deutschland: Hauptschulabschluss)
- berufsbildende mittlere Schule, Lehre, mittlere Reife
- Matura/Abitur
- Matura/Abitur plus 2-3jährige Ausbildung
- Hochschulstudium/Universität

MUTTER:

- ohne abgeschlossene Schulausbildung
- Pflichtschule (Deutschland: Hauptschulabschluss)
- berufsbildende mittlere Schule, Lehre, mittlere Reife
- Matura/Abitur
- Matura/Abitur plus 2-3jährige Ausbildung
- Hochschulstudium/Universität

Wann haben Sie Ihre **Matura** bzw. Ihr **Abitur** absolviert?

- 1999 oder früher
- zwischen 2000 und 2004
- im Jahr 2005 oder 2006
- im Jahr 2007

Bitte geben Sie Ihre **letzte Schulnote** in den untenstehenden Fächern an. Falls Sie ein anderes Notensystem hatten, ordnen Sie Ihre Leistung bitte entsprechend zu (1 = sehr gut; 2 = gut; 3 = befriedigend; 4 = genügend; 5 = nicht genügend) !

	1	2	3	4	5		1	2	3	4	5
Deutsch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Biologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Englisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Chemie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mathematik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Physik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

+

+

+

+

Wie hoch war Ihr **Interesse** für die folgenden Bereiche in der Schulzeit? (1 = *sehr gering* bis 5 = *sehr hoch*)

	1	2	3	4	5	
Mathematik und Naturwissenschaften (z.B. Physik, Mathematik, Chemie, Biologie)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Sprachen (z.B. Englisch, Französisch, Russisch, Spanisch, Italienisch)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Wirtschafts-, Sozial- und Geisteswissenschaften (z.B. Geschichte, Wirtschaftskunde)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Der **wieviele Antritt** zu einem Zulassungstest fürs Medizinstudium (inkl. Antritte in anderen Ländern) war es **heute** für Sie?

erster Antritt
 zweiter Antritt
 dritter Antritt
 vierter Antritt
 fünfter Antritt

Wie gut haben Sie sich auf den Test **vorbereitet**? (*gar nicht* bis *sehr intensiv*)

gar nicht
 eher wenig
 mittelmäßig
 intensiv
 sehr intensiv

Was haben Sie zur **Vorbereitung** genutzt? Geben Sie bitte jeweils das **Ausmaß** an! (1 = *sehr wenig* bis 5 = *sehr intensiv*)

	1	2	3	4	5		1	2	3	4	5	
Universitäts-Homepage	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Studium zusätzlicher Literatur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
andere Testforen im Internet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Recherche zu Studienanforderungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Teilnahme an Kursen/Trainings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Sonstiges	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie **wichtig** ist es für Sie Medizin zu studieren? (*wenig wichtig* bis *sehr wichtig*)

wenig wichtig
 eher wenig
 mittelmäßig
 wichtig
 sehr wichtig

Was würden Sie tun, **wenn Sie nicht aufgenommen werden**? (1 = *keinesfalls* bis 5 = *sicher*)

+

	1	2	3	4	5
neuerlicher Testantritt (Akzeptanz der Wartezeit)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
andere Ausbildung im medizinischen Bereich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
naturwissenschaftliche Studienrichtung inskribieren (z.B. Pharmazie)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
human- od. sozialwissenschaftliche Studienrichtung inskribieren (z.B. Psychologie)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bewerbung für einen Medizin-Studienplatz in einem anderen Land	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie hoch schätzen Sie Ihre **Fähigkeiten** in den folgenden Bereichen ein? (1 = *sehr niedrig* bis 5 = *sehr hoch*)

	1	2	3	4	5		1	2	3	4	5	
math.-naturwissenschaftl. Bereich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		sprachliche Fähigkeiten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
selbstorganisiertes Arbeiten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Lern- und Merkfähigkeit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stressbewältigung bei Prüfungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		konzentriertes Arbeiten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie **zuversichtlich** sind Sie, dass Sie in diesem Test erfolgreich waren? (*sehr wenig* bis *sehr zuversichtlich*)

sehr wenig
 eher wenig
 mittelmäßig
 eher zuversichtlich
 sehr zuversichtlich

Wie **angespannt/gestresst** fühlen bzw. fühlten Sie sich....? (1 = *sehr wenig* bis 5 = *sehr stark*)

	1	2	3	4	5		1	2	3	4	5	
vor dem Test	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		im Moment (nach dem Test)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Für wie **fair** erachten Sie dieses Auswahlverfahren? (*gar nicht fair* bis *sehr fair*)

gar nicht fair
 eher wenig
 mittelmäßig
 eher fair
 sehr fair

Herzlichen Dank für Ihre Mitarbeit!

+

© Spiel, Schober & Litzenberger (Universität Wien)

+

A.4. Beschreibung der Stichprobe

A.4.1. Verteilung der Nationalitäten

Tabelle A.3.: Häufigkeiten der Nationalitäten

Nationalität	Häufigkeit	Nationalität	Häufigkeit
Österreich	1936	Luxemburg	23
Deutschland	1366	Mazedonien	5
Italien (Südtirol)	86	Mexiko	1
Afghanistan	4	Moldau	1
Ägypten	1	Niederlande	1
Albanien	2	Nigeria	1
Armenien	1	Pakistan	1
Bosnien und Herzegowina	15	Palau	1
Bulgarien	4	Philippinen	1
China (Volksrepublik)	2	Polen	11
Dänemark	1	Portugal	1
Finnland	1	Rumänien	9
Frankreich	4	Russland	3
Georgien	2	Schweden	2
Griechenland	4	Schweiz	3
Grossbrit. u. Nordirland	1	Serbien	13
Guatemala	1	Slowakei	12
Indien	1	Spanien	1
Irak	1	Südafrika	1
Iran, Islamische Republik	3	Syrien, Arabische Rep.	1
Israel	2	Tschechien	4
Italien	16	Türkei	26
Korea (Republik/Süd)	1	Ukraine	4
Kroatien	13	Ungarn	20
Lettland	1	Vereinigte St. v. Amerika	1
Libyen	2	Vietnam	1
Liechtenstein	2	Weißrussland	1
Litauen	1	Total	3623

A.4.2. Regionale Herkunft

Tabelle A.4.: Bevölkerung nach regionaler Herkunft für Österreich und Deutschland, Stand: 31.12.2008.

Regionale Herkunft	ÖSTERREICH		DEUTSCHLAND	
	Einwohner	Prozent	Einwohner	Prozent
Großstadt > 50.000	2.651.622	31.79	32.636.681	39.80
mittelgroße Stadt 10. – 50.000	1.046.496	12.55	27.179.734	33.15
Kleinstadt bzw. ländliche Region < 10.000	4.643.206	55.67	22.185.985	27.06
Gesamt	8.341.324	100.00	82.002.400	100.00

Tabelle A.5.: Prüfgröße Chi-quadrat (χ^2) sowie beobachtete (f_o) und erwartete (f_e) Häufigkeiten getrennt für Österreich und Deutschland.

Regionale Herkunft	ÖSTERREICH			DEUTSCHLAND		
	f_o	f_e	$\frac{(f_o-f_e)^2}{f_e^2}$	f_o	f_e	$\frac{(f_o-f_e)^2}{f_e^2}$
Großstadt > 50.000	868	611	107.8	506	540	2.20
mittelgroße Stadt 10. – 50.000	418	241	129.5	457	450	0.11
Kleinstadt bzw. ländliche Region < 10.000	637	1070	175.5	395	367	2.07
Summe	1923		$\chi^2 = 412.8^{**}$	1358		$\chi^2 = 4.38$
f_o ... beobachtete Häufigkeit	** ... $p < .01$					
f_e ... erwartete Häufigkeit						

Entsprechend der drei erhobenen Kategorien (siehe 5.2.2) wurden alle Gemeinden des Herkunftslandes (Österreich, Deutschland) der Bevölkerungszahl (Stand: 31.12.2008) nach zugeordnet und die Summe der Einwohnerzahl für jede der drei Kategorien gebildet (Tabelle A.4). Die Daten hierfür stammen für Österreich aus der Registerzählung „Endgültige Bevölkerungszahl für die Finanzjahre 2009 und 2010 je Gemeinde“ (Statistik Austria, 2010) und für Deutschland aus dem Auszug des Gemeindeverzeichnisses „Gemeinden mit 5.000 und mehr Einwohnern nach Fläche und Bevölkerung“ (Statistisches Bundesamt Deutschland, 2010b). Für Deutschland wurde die Kategorie 'Kleinstadt bzw. ländliche Region < 10.000' als Rest auf den „Bevölkerungsstand am 31.08.2010“ (Statistisches Bundesamt Deutschland, 2010a) berechnet. Die Berechnung der Abweichungen von der nationalen Verteilung erfolgten mittels eindimensionalen χ^2 -Test (vgl. Bortz, 1999, S. 159). Tabelle A.5 zeigt die beobachteten und erwarteten¹ Häufigkeiten sowie den χ^2 -Wert für Österreich und Deutschland. Für Österreich

¹Die erwarteten Häufigkeiten wurden entsprechend der Prozentwerte aus Tabelle A.4 geschätzt

zeigt sich auf dem 1%-Niveau eine signifikante Abweichung von der erwarteten Verteilung ($\chi^2_{(2;99\%)} = 9.21$), für Deutschland hingegen ist die Abweichung nicht signifikant ($p = .112^2$).

A.4.3. Schultyp

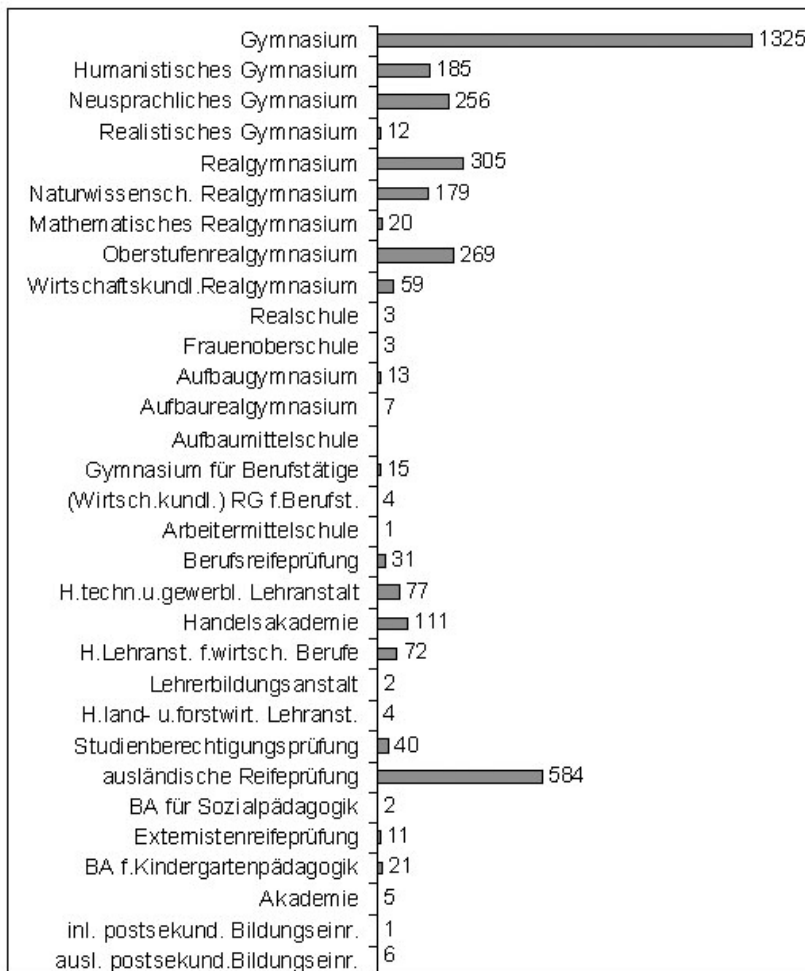


Abbildung A.2.: Absolute Häufigkeiten der Kandidatinnen und Kandidaten der einzelnen Schultypen (Quelle: Medizinische Universitäten Wien und Innsbruck).

²Die Signifikanz p wurde bei gegebenen χ^2 -Wert in MS-Excel mit der Funktion CHIVERT berechnet. Diese Funktion gibt Werte der Verteilungsfunktion $(1-\text{Alpha})$ einer Chi-Quadrat-verteilten Zufallsgröße zurück.

A.5. Bildung des Notenprofils

Die im Fragebogen in einer 5-stufigen Skala erhobenen letzten Schulnote (Matura/Abitur) für die sechs Fächer Deutsch, Englisch, Mathematik, Biologie, Chemie und Physik wurden für weitere Berechnungen zur Analyse des Notenprofils wie folgt zusammengefasst. Zunächst wurde das arithmetische Mittel aus allen Noten berechnet. Diese Werte wurden in das übliche Schulnotensystem (1 bis 5) wie folgt kategorisiert:

- Werte von 1 bis 1.49 wurden zu 1
- Werte von 1.51 bis 2.49 wurden zu 2
- Werte von 2.51 bis 3.49 wurden zu 3
- Werte von 3.51 bis 4.49 wurden zu 4
- Werte von 4.51 bis 5 wurden zu 5

Noten-Mittelwerte die genau zwischen zwei Noten lagen (Bsp. 1.5; 2.5 usw.) wurden, um einen systematischen Rundungsfehler zu umgehen, basierend auf einer Zufallsfunktion³ abgerundet, wenn die Zufallszahl 0 war und aufgerundet, wenn die Zufallszahl 1 war.

³RND(RV.UNIFORM(0,1)); RV.UNIFORM(min, max). Numerisch. Erzeugt eine Reihe von gleichverteilten Zufallszahlen im Bereich min bis max. RND() rundet auf einen ganzzahligen Wert.

A.6. Deskriptive Ergebnisse

Tabelle A.6.: Bivariate (Inter-)Korrelationen der Schulnoten, Durchschnittsnote und der Testwerte.

	Korrelation r							
	2	3	4	5	6	7	8	9
SCHULNOTEN								
1. Deutsch	0.45	0.26	0.32	0.27	0.25	0.57	-0.11	-0.11
2. Englisch	1.00	0.27	0.25	0.22	0.25	0.56	-0.15	-0.15
3. Mathematik		1.00	0.33	0.40	0.45	0.63	-0.29	-0.27
4. Biologie			1.00	0.47	0.41	0.60	-0.10	-0.09
5. Chemie				1.00	0.52	0.67	-0.12	-0.11
6. Physik					1.00	0.65	-0.11	-0.09
7. Durchschnittsnote						1.00	-0.21	-0.20
TESTWERTE								
8. ungekürzte Version							1.00	0.95
9. verkürzte Version								1.00

alle Korrelationen signifikant: $p < .01$

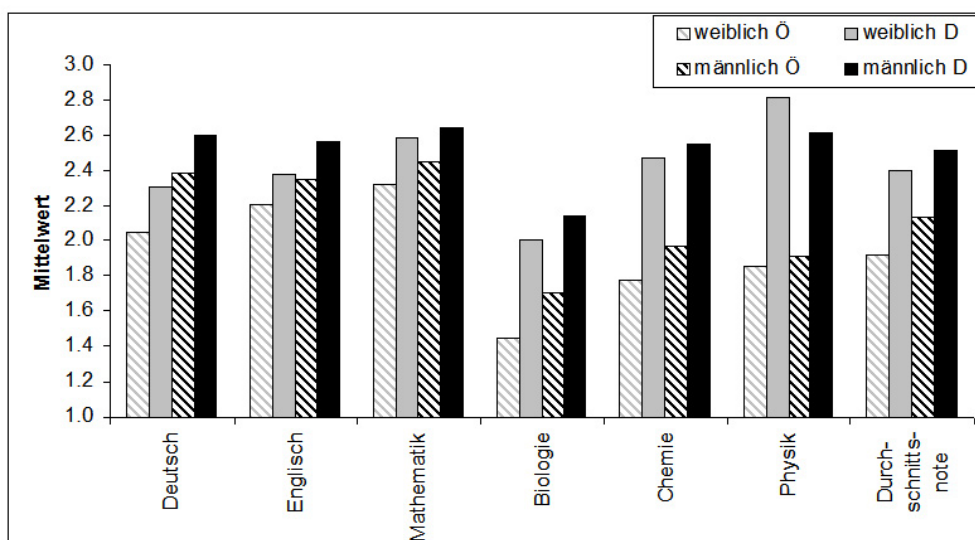


Abbildung A.3.: Mittelwerte der Einzelfachnoten und der Durchschnittsnote; getrennt nach Geschlecht und Nationalität.

Tabelle A.7.: Mittelwerte der Einzelfachnoten und der Durchschnittsnote,
Stichprobenumfang (n); getrennt nach Geschlecht und Nationalität.

Geschlecht	Nationalität	Schulnote	n	Mittelwert
weiblich	Österreich	Deutsch	1081	2.05
		Englisch	1085	2.21
		Mathematik	1083	2.33
		Biologie	1076	1.45
		Chemie	1067	1.78
		Physik	1065	1.86
		Durchschnittsnote	1090	1.92
	Deutschland	Deutsch	752	2.31
		Englisch	751	2.38
		Mathematik	753	2.59
		Biologie	749	2.01
		Chemie	683	2.47
		Physik	677	2.82
		Durchschnittsnote	757	2.40
männlich	Österreich	Deutsch	825	2.39
		Englisch	820	2.35
		Mathematik	817	2.46
		Biologie	805	1.70
		Chemie	805	1.97
		Physik	807	1.91
		Durchschnittsnote	827	2.13
	Deutschland	Deutsch	589	2.60
		Englisch	587	2.56
		Mathematik	591	2.64
		Biologie	572	2.14
		Chemie	540	2.56
		Physik	533	2.62
		Durchschnittsnote	592	2.51

A.7. Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität

A.7.1. Testwert (Einzelfachnoten)

listenweiser Ausschluss von 675 Personen (N=2948).

Verletzung der Voraussetzungen:

Levene-Test: sign. ($p = .018$; ungekürzte Testversion)

A.7 Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität

Tabelle A.8.: Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Testwerts nach regionale Herkunft, Geschlecht und Nationalität; ungekürzte und verkürzte Testversion.

regionale Herkunft	Geschlecht	Nationalität	n	ungekürzte Version		verkürzte Version		d	t	sign.
				M	SD	M	SD			
Großstadt	weiblich	Ö	430	96.68	9.43	96.94	9.50	-0.25	1.204	
		D	225	102.34	9.43	103.08	9.19	-0.75	2.584	
		Gesamt	655	98.63	9.80	99.05	9.83	-0.42	2.436	
	männlich	Ö	386	101.64	10.49	101.58	10.28	0.06	0.247	
		D	192	104.73	9.04	105.12	9.06	-0.39	1.269	
		Gesamt	578	102.67	10.13	102.76	10.02	-0.09	0.490	
	Gesamt	Ö	816	99.03	10.24	99.13	10.14	-0.11	0.671	
		D	417	103.44	9.32	104.02	9.18	-0.58	2.760	
		Gesamt	1233	100.52	10.15	100.79	10.09	-0.27	2.083	
mittelgroße Stadt	weiblich	Ö	218	95.60	8.03	96.10	8.21	-0.50	1.815	
		D	203	102.61	8.36	103.30	8.46	-0.69	2.391	
		Gesamt	421	98.98	8.90	99.57	9.07	-0.59	2.852	
	männlich	Ö	177	101.73	9.39	101.70	9.38	0.03	0.095	
		D	175	105.01	9.51	105.21	9.27	-0.20	0.623	
		Gesamt	352	103.36	9.58	103.44	9.48	-0.09	0.370	
	Gesamt	Ö	395	98.35	9.17	98.61	9.18	-0.26	1.205	
		D	378	103.72	8.98	104.18	8.88	-0.47	2.137	
		Gesamt	773	100.97	9.46	101.33	9.45	-0.36	2.303	
Kleinstadt/ ländliche Region	weiblich	Ö	393	97.14	8.96	97.34	9.21	-0.20	0.907	
		D	208	102.62	9.87	103.24	9.66	-0.62	2.009	
		Gesamt	601	99.04	9.64	99.38	9.77	-0.34	1.897	
	männlich	Ö	219	101.71	9.67	101.90	9.72	-0.18	0.612	
		D	122	104.32	9.28	105.10	8.81	-0.78	2.028	
		Gesamt	341	102.64	9.60	103.04	9.52	-0.40	1.677	
	Gesamt	Ö	612	98.78	9.47	98.97	9.64	-0.19	1.078	
		D	330	103.25	9.68	103.93	9.38	-0.68	2.820	
		Gesamt	942	100.34	9.77	100.71	9.84	-0.36	2.506	
Gesamt	weiblich	Ö	1041	96.63	8.98	96.91	9.14	-0.28	2.138	
		D	636	102.52	9.24	103.20	9.11	-0.69	4.033	sign.
		Gesamt	1677	98.86	9.52	99.30	9.62	-0.44	4.072	sign.
	männlich	Ö	782	101.68	10.01	101.70	9.92	-0.02	0.098	
		D	489	104.72	9.26	105.15	9.06	-0.42	2.180	
		Gesamt	1271	102.85	9.84	103.02	9.74	-0.17	1.387	
	Gesamt	Ö	1823	98.80	9.76	98.96	9.77	-0.17	1.622	
		D	1125	103.48	9.31	104.05	9.13	-0.57	4.459	sign.
		Gesamt	2948	100.58	9.85	100.90	9.84	-0.32	3.936	sign.

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha')$; α' ... alpha Korrektur (Bonferoni)

A.7.2. Untertests (Einzelfachnoten)

Tabelle A.9.: Effekte der Einzelfachnoten, Geschlecht und Nationalität in den einzelnen Untertests (nur signifikante Ergebnisse); F-Wert, Signifikanz (p) und Effektstärke (η_p^2) der ungekürzten Testversion.

	abhängige Variable	F(1,2930)	p	η_p^2
Deutsch	SFIG	5.21	.023	.002
	TV	10.01	.002	.003
	PO	11.40	.001	.004
	MNGV	5.39	.020	.002
	Dut	12.03	.001	.004
Englisch	QFP	5.05	.025	.002
	TV	23.08	.000	.008
	MNGV	27.89	.000	.009
	Fakten	27.77	.000	.009
	Dut	17.87	.000	.006
Mathe	QFP	191.61	.000	.061
	SFIG	34.10	.000	.012
	TV	40.95	.000	.014
	PO	111.56	.000	.037
	MNGV	28.22	.000	.010
	Figuren	38.01	.000	.013
	Fakten	15.33	.000	.005
	Muster	20.14	.000	.007
	Dut	87.68	.000	.029
	Konz	42.68	.000	.014
Biologie	TV	18.84	.000	.006
	MNGV	10.07	.002	.003
	Fakten	6.91	.009	.002
	Muster	6.40	.011	.002
	Dut	4.49	.034	.002
Chemie	SFIG	8.29	.004	.003
	TV	12.56	.000	.004
	MNGV	21.80	.000	.007
	Figuren	5.39	.020	.002
	Muster	5.88	.015	.002
Physik	QFP	8.23	.004	.003
	Dut	6.51	.011	.002
	Konz	4.05	.044	.001

	abhängige Variable	F(1,2930)	p	η_p^2
Nationalität	QFP	157.28	.000	.051
	SFIG	74.98	.000	.025
	TV	340.55	.000	.104
	PO	112.08	.000	.037
	MNGV	288.66	.000	.090
	Figuren	47.86	.000	.016
	Fakten	45.25	.000	.015
	Muster	80.34	.000	.027
	Dut	196.31	.000	.063
	Konz	10.93	.001	.004
Geschlecht	QFP	232.11	.000	.073
	SFIG	189.78	.000	.061
	TV	89.84	.000	.030
	PO	39.24	.000	.013
	MNGV	103.18	.000	.034
	Fakten	8.90	.003	.003
	Dut	332.62	.000	.102
	Konz	17.10	.000	.006
regionale Herkunft	MNGV	4.87 ¹⁾	.008	.003
Nationalität* Geschlecht	QFP	5.01	.025	.002
	SFIG	8.22	.004	.003
	TV	18.35	.000	.006
	PO	21.66	.000	.007
	MNGV	14.67	.000	.005
	Figuren	6.30	.012	.002
	Fakten	9.90	.002	.003
Nationalität* regionale Herkunft	Dut	4.92	.027	.002
	Konz	4.77	.029	.002
	Muster	3.48 ¹⁾	.031	.002

¹⁾ ... F(2,2930)

A.7 Leistungsunterschiede in Abhängigkeit von Einzelfachnoten, regionaler Herkunft, Geschlecht und Nationalität

Tabelle A.10.: Effekte der Einzelfachnoten, Geschlecht und Nationalität in den einzelnen Untertests (nur signifikante Ergebnisse); F-Wert, Signifikanz (p) und Effektstärke (η_p^2) der verkürzten Testversion.

	abhängige Variable	F(1,2930)	p	η_p^2
Deutsch	SFIG	5.01	.025	.002
	TV	13.98	.000	.005
	PO	6.76	.009	.002
	MNGV	6.44	.011	.002
	Fakten	4.68	.031	.002
	Dut	13.94	.000	.005
Englisch	TV	19.75	.000	.007
	MNGV	14.09	.000	.005
	Fakten	20.29	.000	.007
	Dut	14.54	.000	.005
Mathe	QFP	152.83	.000	.050
	SFIG	29.37	.000	.010
	TV	23.43	.000	.008
	PO	88.53	.000	.029
	MNGV	17.12	.000	.006
	Figuren	34.60	.000	.012
	Fakten	12.79	.000	.004
	Muster	17.95	.000	.006
	Dut	79.33	.000	.026
Biologie	TV	12.42	.000	.004
	MNGV	4.71	.030	.002
	Fakten	4.36	.037	.001
	Muster	12.00	.001	.004
	Dut	4.12	.042	.001
Chemie	QFP	5.83	.016	.002
	SFIG	6.76	.009	.002
	TV	4.42	.036	.002
	MNGV	23.94	.000	.008
	Figuren	5.48	.019	.002
Physik	QFP	8.58	.003	.003
	Dut	4.17	.041	.001

	abhängige Variable	F(1,2930)	p	η_p^2
Nationalität	QFP	129.18	.000	.042
	SFIG	82.65	.000	.027
	TV	268.50	.000	.084
	PO	102.97	.000	.034
	MNGV	266.91	.000	.083
	Figuren	40.12	.000	.014
	Fakten	38.50	.000	.013
	Muster	50.34	.000	.017
	Dut	154.63	.000	.050
Geschlecht	QFP	228.84	.000	.072
	SFIG	170.47	.000	.055
	TV	53.69	.000	.018
	PO	15.25	.000	.005
	MNGV	68.48	.000	.023
	Fakten	9.13	.003	.003
	Dut	294.77	.000	.091
Nationalität* Geschlecht	SFIG	5.83	.016	.002
	TV	14.08	.000	.005
	PO	12.63	.000	.004
	MNGV	11.68	.001	.004
	Figuren	8.57	.003	.003
	Fakten	9.36	.002	.003
	Muster	4.81	.028	.002
Dut	4.15	.042	.001	

Tabelle A.11.: Signifikanz (p), Effektstärken (η_p^2), Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Faktors Geschlecht (w/m) für alle Untertests; ungekürzte und verkürzte Testversion.

Untertest	n	ungekürzte Testversion				verkürzte Testversion				d	t	sign.
		p	η_p^2	M	SD	p	η_p^2	M	SD			
QFP	w 1688	<.001	.073	98.18	9.14	<.001	.072	98.34	9.41	-0.16	1.535	
	m 1275			103.28	10.15			103.50	9.80			
SFIG	w 1688	<.001	.061	98.12	9.63	<.001	.055	98.40	9.87	-0.28	2.599	
	m 1275			103.07	9.48			103.03	9.09			
TV	w 1688	<.001	.030	99.12	9.70	<.001	.018	99.81	9.51	-0.69	6.463	sign.
	m 1275			102.15	10.09			102.04	10.00			
PO	w 1688	<.001	.013	99.64	9.82	<.001	<.01	100.12	9.97	-0.47	4.385	sign.
	m 1275			101.90	10.09			101.54	10.18			
MNGV	w 1688	<.001	.034	99.03	9.73	<.001	.023	99.67	9.79	-0.63	5.868	sign.
	m 1275			102.26	9.91			102.22	9.63			
Figuren	w 1688	.129	<.01	99.90	9.86	.209	<.01	100.03	9.96	-0.13	1.216	
	m 1275			100.43	10.11			100.51	10.02			
Fakten	w 1688	.003	<.01	101.03	9.70	.003	<.01	101.23	9.73	-0.20	1.826	
	m 1275			99.63	10.11			99.83	10.14			
Muster	w 1688	.247	<.01	100.04	10.16	.465	<.01	100.16	9.98	-0.12	1.105	
	m 1275			100.31	9.82			100.31	9.95			
DuT	w 1688	<.001	.102	97.89	9.06	<.001	.091	98.29	9.09	-0.40	3.898	sign.
	m 1275			103.87	10.26			103.97	10.37			
Konz	w 1688	<.001	<.01	99.77	9.60	--	--	--	--	--	--	--
	m 1275			101.21	9.93			--	--			

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... t>t(df,α'); α' ... alpha Korrektur (Bonferoni)

Tabelle A.12.: Signifikanz (p), Effektstärken (η_p^2), Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Faktors Nationalität (Ö/D) für alle Untertests; ungekürzte und verkürzte Testversion.

Untertest	n	ungekürzte Testversion				verkürzte Testversion				d	t	sign.
		p	η_p^2	M	SD	p	η_p^2	M	SD			
QFP	Ö 1833	<.001	.051	99.12	9.88	<.001	.042	99.45	9.89	-0.325	3.128	
	D 1130			102.40	9.63			102.36	9.69			
SFIG	Ö 1833	<.001	.025	99.27	9.91	<.001	.027	99.34	9.96	-0.073	0.700	
	D 1130			101.84	9.61			102.10	9.33			
TV	Ö 1833	<.001	.104	98.22	9.58	<.001	.084	98.79	9.70	-0.570	5.552	sign.
	D 1130			103.99	9.58			103.97	9.05			
PO	Ö 1833	<.001	.037	99.28	9.91	<.001	.034	99.37	9.98	-0.091	0.877	
	D 1130			102.77	9.77			102.93	9.87			
MNGV	Ö 1833	<.001	.090	98.44	9.76	<.001	.083	98.84	9.84	-0.401	3.873	sign.
	D 1130			103.63	9.37			103.88	8.89			
Figuren	Ö 1833	<.001	.016	99.35	9.85	<.001	.014	99.51	9.87	-0.161	1.548	
	D 1130			101.40	10.04			101.42	10.08			
Fakten	Ö 1833	<.001	.015	99.74	9.87	<.001	.013	99.99	9.95	-0.253	2.434	
	D 1130			101.54	9.85			101.66	9.82			
Muster	Ö 1833	<.001	.027	99.04	9.80	<.001	.017	99.34	9.96	-0.306	2.943	
	D 1130			101.98	10.10			101.66	9.81			
DuT	Ö 1833	<.001	.063	98.97	9.88	<.001	.050	99.40	9.96	-0.429	4.127	sign.
	D 1130			102.89	9.83			102.90	9.86			
Konz	Ö 1833	.001	<.01	100.27	9.83	--	--	--	--	--	--	--
	D 1130			100.59	9.67			--	--			

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... t>t(df,α'); α' ... alpha Korrektur (Bonferoni)

A.8. Leistungsunterschiede in Abhängigkeit von der Durchschnittsnote, regionaler Herkunft, Geschlecht und Nationalität

A.8.1. Testwert (Durchschnittsnote)

abhängige Variable: Testwert

unabhängige Variablen: Durchschnittsnote (1 bis 4), Geschlecht (weiblich, männlich), Nationalität (Österreich, Deutschland), Ort (Großstadt, mittelgroße Stadt, Kleinstadt/ländliche Region)

Verletzung der Voraussetzungen:

Levene-Test:

ungekürzte Testversion: sign. ($p = .003$); verkürzte Testversion: sign. ($p = .028$)

Tabelle A.13.: Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Testwerts nach Durchschnittsnote, Geschlecht und Nationalität; ungekürzte und verkürzte Testversion.

Durchschnittsnote	Geschlecht	Nationalität	n	ungekürzte Version		verkürzte Version		d	t	sign.
				M	SD	M	SD			
Note 1	weiblich	Ö	322	101.08	9.06	101.21	9.26	0.13	0.543	
		D	27	106.13	7.10	106.52	6.90	0.39	0.526	
		Gesamt	349	101.47	9.02	101.62	9.20	0.15	0.654	
	männlich	Ö	178	107.42	9.70	107.48	9.70	0.06	0.186	
		D	13	112.57	9.70	112.27	8.11	-0.30	0.245	
		Gesamt	191	107.77	9.77	107.81	9.65	0.04	0.116	
Note 2	weiblich	Ö	546	96.00	8.23	96.47	8.36	0.47	2.720	
		D	416	103.77	9.28	104.39	9.10	0.62	2.956	
		Gesamt	962	99.36	9.51	99.89	9.53	0.54	3.825	sign.
	männlich	Ö	381	101.86	9.59	101.73	9.51	-0.13	0.578	
		D	284	106.09	8.75	106.46	8.74	0.37	1.469	
		Gesamt	665	103.66	9.47	103.75	9.47	0.08	0.484	
Note 3	weiblich	Ö	204	91.51	7.52	91.58	7.70	0.07	0.270	
		D	299	99.67	8.97	100.55	8.98	0.88	3.579	sign.
		Gesamt	503	96.36	9.32	96.91	9.55	0.55	2.848	
	männlich	Ö	252	97.04	8.75	97.22	8.62	0.18	0.692	
		D	274	103.08	9.08	103.44	8.83	0.36	1.400	
		Gesamt	526	100.19	9.41	100.46	9.26	0.27	1.453	
Note 4	weiblich	Ö	18	92.22	7.07	92.40	7.31	0.18	0.197	
		D	15	94.19	9.43	96.70	9.88	2.51	2.138	
		Gesamt	33	93.11	8.15	94.35	8.71	1.24	1.709	
	männlich	Ö	16	94.78	8.38	94.92	8.21	0.13	0.126	
		D	21	96.94	7.74	97.51	7.72	0.57	0.651	
		Gesamt	37	96.01	7.99	96.39	7.93	0.38	0.575	
Gesamt	weiblich	Ö	1090	96.60	8.99	96.89	9.13	0.29	2.271	
		D	757	102.04	9.39	102.80	9.23	0.75	4.794	sign.
		Gesamt	1847	98.83	9.53	99.31	9.62	0.48	4.724	sign.
	männlich	Ö	827	101.45	10.08	101.46	9.98	0.01	0.074	
		D	592	104.51	9.18	104.87	9.01	0.35	2.023	
		Gesamt	1419	102.73	9.83	102.88	9.73	0.15	1.318	
	Gesamt	Ö	1917	98.69	9.77	98.86	9.77	0.17	1.698	
		D	1349	103.13	9.37	103.71	9.19	0.58	4.924	sign.
		Gesamt	3266	100.52	9.85	100.86	9.83	0.34	4.371	sign.

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha)$; α ... alpha Korrektur (Bonferoni)

A.8.2. Untertests (Durchschnittsnote)

AVn: Untertests (M = 100; SD = 10)

QFP, SFIG, TV, PO, MNGV, Figuren, Fakten, Muster, DuT, Konz

Verletzung der Voraussetzungen:

ungekürzte Testversion: Box-M-Test: sig. ($p = .004$)

Levene-Test: sign. ($p < .05$) für TV, PO und DuT

verkürzte Testversion: Box-M-Test: sign. ($p = .007$)

Levene-Test: sign. ($p < .05$) für SFIG, PO und DuT

Tabelle A.14.: Signifikanz (p), Effektstärken (η_p^2), Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Faktors Durchschnittsnote (1/2/3/4) für alle Untertests; ungekürzte und verkürzte Testversion.

Untertest	ungekürzte Testversion				verkürzte Testversion				d	t	sign.	
	p	η_p^2	M	SD	n	p	η_p^2	M				SD
QFP	<.001	.039	1	103.46	10.30	540	<.001	.033	103.37	9.71	0.09	0.456
			2	100.63	9.75	1627			100.88	9.87	-0.26	2.329
			3	98.37	9.62	1029			98.64	9.85	-0.27	1.980
			4	96.46	8.60	70			97.37	9.98	-0.91	1.756
SFIG	<.001	.012	1	101.74	9.29	540	<.001	.013	101.80	9.03	-0.06	0.342
			2	100.35	9.85	1627			100.45	9.76	-0.10	0.865
			3	99.63	10.11	1029			99.83	10.15	-0.21	1.478
			4	95.97	9.89	70			96.39	10.22	-0.41	0.760
TV	<.001	.039	1	102.09	10.41	540	<.001	.029	102.56	9.93	-0.20	1.019
			2	101.12	10.02	1627			101.33	9.77	-0.21	1.862
			3	98.69	9.39	1029			99.31	9.71	-0.62	4.583
			4	95.28	7.39	70			96.53	8.29	-1.25	2.617
PO	<.001	.018	1	102.65	10.40	540	<.001	.014	101.67	10.72	0.43	2.151
			2	100.97	9.89	1627			101.15	9.89	-0.18	1.646
			3	99.05	9.61	1029			99.34	9.70	-0.29	2.093
			4	97.74	7.94	70			98.24	8.07	-0.50	1.037
MNGV	<.001	.033	1	102.65	10.24	540	<.001	.029	102.71	9.57	-0.06	0.325
			2	100.87	9.98	1627			101.19	9.77	-0.33	2.950
			3	98.78	9.58	1029			99.27	9.79	-0.50	3.608
			4	98.04	7.69	70			98.47	8.61	-0.43	0.882
Figuren	<.001	.015	1	101.22	10.26	540	<.001	.013	101.26	10.38	-0.04	0.197
			2	100.75	10.01	1627			100.76	10.03	-0.01	0.088
			3	99.00	9.75	1029			99.25	9.76	-0.25	1.795
			4	95.91	8.41	70			96.08	8.62	-0.17	0.352
Fakten	<.001	.017	1	103.12	9.90	540	<.001	.014	103.32	9.85	-0.20	1.048
			2	100.71	9.98	1627			100.85	10.04	-0.14	1.299
			3	98.55	9.47	1029			98.83	9.59	-0.28	2.029
			4	96.24	9.21	70			97.03	9.28	-0.78	1.515
Muster	<.001	.012	1	100.98	9.71	540	<.001	<.01	101.17	9.93	-0.19	0.997
			2	100.65	9.89	1627			100.66	9.93	-0.01	0.134
			3	99.20	10.20	1029			99.36	10.02	-0.16	1.114
			4	96.49	9.54	70			96.65	9.63	-0.16	0.299
DuT	<.001	.038	1	103.37	10.39	540	<.001	.034	103.62	10.29	-0.25	1.274
			2	100.73	9.92	1627			101.01	9.99	-0.27	2.452
			3	98.72	9.47	1029			98.98	9.51	-0.26	1.902
			4	96.86	9.18	70			97.48	9.41	-0.62	1.190
Konz	<.001	.012	1	102.44	9.44	540	--	--	--	--	--	
			2	100.56	9.81	1627	--	--	--	--	--	
			3	99.08	10.01	1029	--	--	--	--	--	
			4	96.31	8.89	70	--	--	--	--	--	

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... t>t(df,α'); α' ... alpha Korrektur (Bonferoni)

A.8 Leistungsunterschiede in Abhängigkeit von der Durchschnittsnote, regionaler Herkunft, Geschlecht und Nationalität

Tabelle A.15.: Signifikanz (p), Effektstärken (η_p^2), Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Faktors Geschlecht (w/m) für alle Untertests; ungekürzte und verkürzte Testversion.

Untertest	ungekürzte Testversion					verkürzte Testversion					d	t	sign.	
	p	η_p^2	M	SD	n	p	η_p^2	M	SD	n				
QFP	<.001	.014	w	98.10	9.18	1847	<.001	.013	98.28	9.46	1847	-0.18	1.801	
			m	103.15	10.16	1419			103.42	9.87	1419			
SFIG	<.001	.019	w	98.12	9.65	1847	<.001	.019	98.39	9.88	1847	-0.27	2.648	
			m	103.05	9.49	1419			102.99	9.09	1419			
TV	.002	<.01	w	99.16	9.73	1847	.003	<.01	99.89	9.60	1847	-0.73	7.096	sign.
			m	102.09	10.02	1419			101.97	10.01	1419			
PO	.004	<.01	w	99.54	9.75	1847	.119	<.01	100.00	9.90	1847	-0.46	4.430	sign.
			m	101.71	9.99	1419			101.39	10.04	1419			
MNGV	<.001	<.01	w	99.07	9.74	1847	.001	<.01	99.69	9.80	1847	-0.62	5.981	sign.
			m	102.22	9.93	1419			102.20	9.61	1419			
Figuren	.118	<.01	w	99.95	9.89	1847	.331	<.01	100.09	10.01	1847	-0.14	1.342	
			m	100.47	10.12	1419			100.50	10.04	1419			
Fakten	.020	<.01	w	100.99	9.73	1847	.013	<.01	101.19	9.75	1847	-0.20	1.931	
			m	99.47	10.13	1419			99.70	10.20	1419			
Muster	.084	<.01	w	100.10	10.16	1847	.548	<.01	100.26	10.04	1847	-0.15	1.456	
			m	100.23	9.76	1419			100.24	9.92	1419			
DuT	<.001	.017	w	97.92	9.01	1847	<.001	.011	98.33	9.06	1847	-0.41	4.113	
			m	103.75	10.22	1419			103.84	10.33	1419			
Konz	.067	<.01	w	99.60	9.64	1847	--	--	--	--	1847	--	--	
			m	101.24	10.10	1419			--	--	1419			

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha)$; α ... alpha Korrektur (Bonferoni)

Tabelle A.16.: Signifikanz (p), Effektstärken (η_p^2), Mittelwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) des Faktors Nationalität (Ö/D) für alle Untertests; ungekürzte und verkürzte Testversion.

Untertest	ungekürzte Testversion					verkürzte Testversion					d	t	sign.	
	p	η_p^2	M	SD	n	p	η_p^2	M	SD	n				
QFP	<.001	<.01	Ö	99.05	9.92	1917	<.001	<.01	99.38	9.95	1917	-0.32	3.176	
			D	102.06	9.69	1349			102.13	9.79	1349			
SFIG	.006	<.01	Ö	99.29	9.91	1917	.001	<.01	99.36	9.95	1917	-0.07	0.699	
			D	101.64	9.69	1349			101.85	9.43	1349			
TV	<.001	.016	Ö	98.17	9.56	1917	<.001	.016	98.73	9.72	1917	-0.56	5.547	sign.
			D	103.66	9.63	1349			103.73	9.22	1349			
PO	.001	<.01	Ö	99.18	9.87	1917	.010	<.01	99.27	9.92	1917	-0.09	0.896	
			D	102.34	9.68	1349			102.50	9.76	1349			
MNGV	<.001	.014	Ö	98.35	9.78	1917	<.001	.014	98.72	9.87	1917	-0.37	3.691	
			D	103.42	9.42	1349			103.71	8.91	1349			
Figuren	.011	<.01	Ö	99.37	9.84	1917	.002	<.01	99.54	9.85	1917	-0.17	1.651	
			D	101.32	10.10	1349			101.31	10.17	1349			
Fakten	.028	<.01	Ö	99.68	9.87	1917	.072	<.01	99.96	9.95	1917	-0.28	2.717	
			D	101.25	9.95	1349			101.37	9.95	1349			
Muster	<.001	<.01	Ö	99.04	9.81	1917	.001	<.01	99.32	10.02	1917	-0.29	2.811	
			D	101.75	10.02	1349			101.57	9.79	1349			
DuT	<.001	.011	Ö	98.88	9.86	1917	<.001	.010	99.31	9.95	1917	-0.42	4.157	sign.
			D	102.68	9.72	1349			102.74	9.75	1349			
Konz	.706	<.01	Ö	100.16	9.93	1917	--	--	--	--	1917	--	--	
			D	100.53	9.79	1349			--	--	1349			

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha)$; α ... alpha Korrektur (Bonferoni)

A.9. Einfluss des Schulnotenprofils

Tabelle A.17.: Durchschnittliche Testwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) in Abhängigkeit der Durchschnittsnote und des Ranges; ungekürzte und verkürzte Testversion.

Durchschnittsnote	Range	n	ungekürzte Version		verkürzte Version		d	t	sign.
			M	SD	M	SD			
Note 1	0	129	107.12	10.35	107.13	9.93	0.00	0.010	
	1	346	103.18	9.35	103.22	9.59	-0.04	0.182	
	2	62	99.91	8.59	100.58	9.04	-0.67	1.239	
Note 2	0	23	100.65	9.89	101.52	9.38	-0.86	0.922	
	1	400	102.17	9.64	102.49	9.56	-0.32	1.469	
	2	875	101.26	9.74	101.57	9.63	-0.31	2.091	
	3	324	99.50	9.62	99.93	9.93	-0.43	1.763	
Note 3	0	12	96.66	7.89	95.76	8.48	0.89	0.732	
	1	186	99.34	9.82	99.79	9.76	-0.45	1.378	
	2	505	98.37	9.60	98.65	9.65	-0.28	1.448	
	3	295	97.66	9.44	98.27	9.39	-0.62	2.435	
	4	31	98.24	8.82	99.02	8.84	-0.79	1.028	
Note 4	1	35	92.57	7.75	93.73	8.17	-1.16	1.697	
	2	21	96.89	8.21	97.32	8.56	-0.43	0.466	
	3	9	93.10	7.33	94.49	7.77	-1.39	1.012	

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha')$;

α' ... alpha Korrektur (Bonferoni)

A.10. Leistungsunterschiede in Abhängigkeit vom Schultyp, Durchschnittsnote und Geschlecht

Tabelle A.18.: Durchschnittliche Testwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) in Abhängigkeit des Schultyps und der Durchschnittsnote (Note); ungekürzte und verkürzte Testversion.

Schultyp	Note	n	ungekürzte Version		verkürzte Version		d	t	sign.
			M	SD	M	SD			
HTL	1	19	109.39	8.02	109.93	8.12	-0.55	0.578	
	2	34	100.95	7.52	100.83	7.42	0.12	0.172	
	3	16	97.97	7.56	99.15	8.36	-1.18	1.148	
	Gesamt	69	102.58	8.74	102.95	8.88	-0.37	0.723	
Hum+Neu+ And.Gym	1	110	104.61	11.23	104.97	11.07	-0.36	0.803	
	2	154	101.08	8.51	101.21	8.15	-0.13	0.401	
	3	69	96.25	7.85	96.37	7.90	-0.12	0.249	
	4	6	91.99	7.25	92.51	7.55	-0.52	0.302	
	Gesamt	339	101.08	9.84	101.29	9.68	-0.21	0.880	
Gym. ohne spez. Schwerpunkt	1	150	104.23	8.65	104.03	8.75	0.19	0.563	
	2	216	99.48	9.11	99.75	9.18	-0.26	0.902	
	3	93	94.97	7.77	94.91	7.84	0.06	0.149	
	4	6	95.16	3.93	94.57	3.60	0.59	0.483	
	Gesamt	465	100.05	9.27	100.09	9.32	-0.04	0.201	
Nawi+Mathe+ Realgym.	1	88	104.12	9.63	104.50	9.64	-0.39	0.818	
	2	197	99.69	9.41	99.81	9.25	-0.12	0.405	
	3	123	94.53	8.86	94.46	8.70	0.07	0.191	
	4	7	94.09	10.54	94.01	9.64	0.08	0.044	
	Gesamt	415	99.00	9.91	99.12	9.85	-0.12	0.540	
Oberstufen- realgym.	1	49	100.67	8.81	101.32	9.27	-0.64	1.045	
	2	125	95.81	9.62	96.58	9.56	-0.77	1.954	
	3	75	93.23	9.83	93.33	9.84	-0.10	0.186	
	4	7	91.80	6.61	92.17	7.69	-0.37	0.239	
	Gesamt	256	95.88	9.78	96.41	9.90	-0.54	1.930	
wirtschaftl. Realgym.+ HAK+HLW	1	49	99.57	8.91	99.20	9.13	0.38	0.614	
	2	132	95.87	8.54	96.10	9.06	-0.23	0.621	
	3	41	92.95	9.84	93.62	9.38	-0.66	0.959	
	Gesamt	222	96.15	9.09	96.32	9.27	-0.18	0.607	
Gesamt	1	465	103.64	9.70	103.79	9.78	-0.15	0.717	
	2	858	98.78	9.21	99.04	9.16	-0.26	1.783	
	3	417	94.66	8.75	94.77	8.72	-0.11	0.550	
	4	26	93.24	7.27	93.30	7.18	-0.06	0.080	
	Gesamt	1766	99.01	9.77	99.20	9.76	-0.19	1.835	

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha')$; α' ... alpha Korrektur (Bonferoni)

Tabelle A.19.: Post-hoc Paarvergleiche (nach Scheffé) der Schultypen; ungekürzte und verkürzte Testversion.

(i) Schultyp	(j) Schultyp	ungekürzte Version		verkürzte Version	
		Mean Difference (i-j)	p	Mean Difference (i-j)	p
1) Gymn. ohne spez. Schwerpunkt	2)	-1.065	.762	-1.243	.622
	3)	0.965	.795	0.887	.851
	4)	4.146	.000	3.649	.000
	5)	3.824	.000	3.719	.000
	6)	-2.530	.478	-2.832	.347
2) Hum.+Neu+ And.Gym	1)	1.065	.762	1.243	.622
	3)	2.030	.112	2.130	.083
	4)	5.212	.000	4.893	.000
	5)	4.889	.000	4.962	.000
	6)	-1.465	.919	-1.589	.891
3) Nawi+Mathe+ Realgymn	1)	-0.965	.795	-0.887	.851
	2)	-2.030	.112	-2.130	.083
	4)	3.182	.002	2.763	.016
	5)	2.860	.016	2.832	.019
	6)	-3.494	.133	-3.719	.091
4) Oberstufen- realgymnasium	1)	-4.146	.000	-3.649	.000
	2)	-5.212	.000	-4.893	.000
	3)	-3.182	.002	-2.763	.016
	5)	-0.322	1.000	0.069	1.000
	6)	-6.676	.000	-6.482	.000
5) wirtschaftl. Gym+HAK+HLW	1)	-3.824	.000	-3.719	.000
	2)	-4.889	.000	-4.962	.000
	3)	-2.860	.016	-2.832	.019
	4)	0.322	1.000	-0.069	1.000
	6)	-6.354	.000	-6.551	.000
6) HTL	1)	2.530	.478	2.832	.347
	2)	1.465	.919	1.589	.891
	3)	3.494	.133	3.719	.091
	4)	6.676	.000	6.482	.000
	5)	6.354	.000	6.551	.000

Tabelle A.20.: Durchschnittliche Testwerte (M), Standardabweichungen (SD) und Stichprobenumfänge (n) in Abhängigkeit des Schultyps und des Geschlechts; ungekürzte und verkürzte Testversion.

Schultyp	Geschlecht	n	ungekürzte Version		verkürzte Version		d	t	sign.
			M	SD	M	SD			
HTL	weiblich	12	101.49	7.44	103.19	7.87	-1.70	1.442	
	männlich	58	102.70	8.99	102.77	9.12	-0.06	0.108	
	Gesamt	70	102.50	8.70	102.84	8.86	-0.34	0.678	
Hum+Neu+ And.Gym.	weiblich	185	98.04	8.66	98.51	8.75	-0.46	1.509	
	männlich	157	104.55	9.96	104.48	9.71	0.07	0.206	
	Gesamt	342	101.03	9.82	101.25	9.66	-0.22	0.910	
Gym. ohne spez. Schwerpunkt	weiblich	279	98.10	8.65	98.22	8.84	-0.13	0.507	
	männlich	190	102.71	9.56	102.62	9.50	0.09	0.281	
	Gesamt	469	99.97	9.30	100.01	9.35	-0.04	0.198	
Nawi+Mathe+ Realgym.	weiblich	207	96.88	9.61	97.03	9.58	-0.15	0.494	
	männlich	209	101.10	9.76	101.19	9.67	-0.08	0.276	
	Gesamt	416	99.00	9.90	99.12	9.84	-0.12	0.539	
wirtschaftl. Gym.+HAK+HLW	weiblich	166	94.92	8.47	95.01	8.76	-0.09	0.270	
	männlich	61	99.46	9.58	99.76	9.53	-0.30	0.534	
	Gesamt	227	96.14	8.99	96.29	9.20	-0.14	0.510	
Oberstufen- realgym.	weiblich	171	94.87	9.21	95.60	9.40	-0.73	2.204	
	männlich	88	97.66	10.59	97.82	10.66	-0.16	0.327	
	Gesamt	259	95.82	9.77	96.36	9.88	-0.54	1.943	
Gesamt	weiblich	1020	96.82	9.01	97.13	9.15	-0.31	2.292	
	männlich	763	101.81	9.97	101.84	9.87	-0.03	0.206	
	Gesamt	1783	98.96	9.75	99.14	9.75	-0.19	1.809	

d ... Mittelwertsdifferenz; t ... Prüfgröße (t-Test); sign. ... $t > t(df, \alpha')$; α' ... alpha Korrektur (Bonferoni)

„Hiermit bestätige ich, dass die vorliegende Arbeit in allen relevanten Teilen
selbstständig durchgeführt wurde.“

Wien, im November 2010

Ing. Andreas Pfaffel

B. Lebenslauf

Persönliche Daten

Name	Ing. Andreas PFAFFEL
Adresse	A-1160 Wien, Roseggergasse 37/17
Geburtsdatum	16. November 1977, Wien
e-mail Adresse	andreas.pfaffel@gmx.at
Familienstand	ledig
Nationalität	Österreich

Ausbildung

seit Okt.2002	Diplomstudium der Psychologie an der Fakultät für Psychologie (Universität Wien)
Mai 1997	Reif- und Diplomprüfung mit ausgezeichnetem Erfolg an der HBLVA für chemische Industrie
1992-1997	Höhere Bundeslehr- und Versuchsanstalt für chemische Industrie, Wien, Rosensteingasse; Ausbildungszweig: Technische Chemie
1984-1992	Pflichtschulen, Wien

Berufliche Tätigkeiten

seit SS 2008	Studienassistent am Institut für Wirtschaftspsychologie, Bildungspsychologie und Evaluation, Universität Wien <ul style="list-style-type: none">•Betreuung von Lehrveranstaltungen:<ul style="list-style-type: none">-Forschungsmethoden und Evaluation (Prüfungsmanagement, Prüfungsentwicklung, Prüfungskorrektur, Prüfungsaufsicht, e-Learning Plattform, Prüfungseinsichten)-Fachliteraturseminar u. Diplomandenseminare•Unterstützung der Forschung (Literaturrecherche, Gestaltung von Texten, Grafiken und Tabellen)•Unterstützung bei administrativen Tätigkeiten
1998-2002	Österreichisches Getränkeinstitut ÖGI, Wien <i>Analytik</i> <ul style="list-style-type: none">•Selbständige Durchführung von nasschemischen- sowie instrumentellen Analysenverfahren:<ul style="list-style-type: none">-Atomabsorptionsspektroskopie (Flammen-, Graphitrohr- und Hydridtechnik), Gaschromatographie (GC), Massenspektroskopie (GC-MS), Flüssigchromatographie (HPLC), Photometrie, Isotachophorese <i>Entwicklung & Methodik</i> <ul style="list-style-type: none">-Entwicklung neuer Analysenmethoden im Team-Mitarbeit in der Planung, Durchführung und Auswertung von Ringanalysen

Qualitätssicherung

Validierung neuer Analysemethoden sowie Revalidierung von Routineverfahren
 Kalibrierung und Wartung von Analysengeräten
 Erstellen von Standardarbeitsanweisungen (SOP)
 1997 - 1998 Medizinisches Labor des Heeresspitals, Wien (Van-Swieten-Kaserne)
 Durchführung von Blutbildanalysen, Photometrische Harnanalysen, Elektrophoretische Proteinauftrennung, Drogenschnelltests

Projekte

Okt. 2009 Transfer und Nachhaltigkeit im E-Learning, Universität Wien
 -statistische Datenanalyse
 Mai 2008 - März 09 Nationaler Bildungsbericht, Universität Wien
 Literaturrecherche, Layouting von Text, Abbildungen und Tabellen
 2007 - 2008 Evaluation der Eignungstests für das Medizinstudium in Österreich, Universität Wien
 -Einlesen und Erstellen des Datensatzes, Bereinigung der Daten
 -Statistische Auswertungen (ANOVA, MANOVA, multiple Regressionen, Clusteranalysen, Diskriminanzanalysen, Konfigurationsfrequenzanalysen KFA, Häufigkeitsanalysen, Korrelationen)
 -Erstellen von Grafiken, Tabellen, Gestaltung des Evaluationsberichtes

Praktika

Sommer 2007 Institut für Wirtschaftspsychologie, Bildungspsychologie und Evaluation
 Praktikum im Projekt TALK (Training zum Aufbau von LehrerInnen-Kompetenzen zur Förderung von Bildungsmotivation und Lebenslangem Lernen)
 Sommer 1996 Rembrantin Lack GesmbH, Wien
 Praktikum im Entwicklungslabor (aktiver Korrosionsschutz)
 Qualitätskontrolle verschiedener Lacke, Herstellung von Lacken (Mischen, Mahlen), Ausfärben von Lacken
 Sommer 1994 Nalco Chemical Ltd., Wien
 Praktikum im Qualitätslabor
 Qualitätskontrolle von Polymeren (Dichte, Viskosität), Schwermetallanalyse mittels Atomabsorptionsspektroskopie (AAS), Herstellung mikrobiologischer Schnelltests

Weitere Kenntnisse

Englisch
 SPSS, Word, Excel, Latex, Reference Manager (Endnote, jabref, Citavi), Adobe GoLive (html)
 Literaturrecherche: DB-Service & EZB der Uni Wien
 e-Learning Plattform Moodle
 Führerschein B