

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Institut und Poliklinik für Allgemeinmedizin
Direktor: Prof. Dr. med. Martin Scherer

Die Qualifikation zur Allgemeinen Ärztlichen Reife. Ein Vergleich zweier Studiengangsmodelle.

Dissertation

zur Erlangung des Grades eines Doktors der Humanbiologie
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Rebecca Katharina Machnik (geb. Keim), Dipl.-Psych.
aus Hamburg

Hamburg 2022

Angenommen von der Medizinischen Fakultät der Universität Hamburg am: 09.01.2023
Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der Vorsitzende: Prof. Dr. Martin Scherer

Prüfungsausschuss, zweite Gutachterin: Prof. Dr. Sigrid Harendza

Abstract

Introduction: At the University Hospital Hamburg-Eppendorf, the integrated medical model curriculum (iMED, „*integrierte Modellstudiengang Medizin Hamburg*“) replaced the standard curriculum (RSG, „*Regelstudiengang*“) in 2012. Based on new regulations for medical curricula, it is one of Germany's firsts to focus on the most modern concepts for training medical students, with the aim of interlacing the theoretical basic subjects in medicine with clinical practice right from the start of university attendance. In particular, a focus was placed on the integrated teaching of communicative and practical skills in the new curriculum.

The study aimed to analyse the differences in the performance of students in a clinical-practical OSCE (= objective structured clinical examination), trying to establish the level of general medical maturity („*Allgemeine ärztliche Reife*“) among the participants, and to further investigate whether potential differences in performance were caused by the type of medical curriculum completed or more so through individual (personality) differences.

Methods: Between the years 2016 to 2018, four cross-sectional test times were carried out. Each time, voluntary graduates from both curricula groups (RSG vs. iMED) were recruited for participation in a self-designed comparative OSCE that, unbeknownst to the participants, measured their general medical maturity. Additionally, they had to complete a psychometric and sociodemographic questionnaire containing the following aspects: personality profile (NEO-FFI 30), life satisfaction (Satisfaction with Life Scale (SWLS)), social support (F-SoZu K14), anxiety and depression values (Hospital Anxiety and Depression Scale (HADS-D)), socio-demographics and aspired medical specialisation after completing their studies.

Result: On average, the participants (N = 100, 50% female) showed satisfactory performance in the test for general medical maturity, regardless of their completed curriculum. The overall sample achieved extraordinarily positive psychometric values and surprisingly showed a "personality norm profile" with similar characteristics that applied to all participants.

Discussion: The change in the Hamburg medical curriculum did not lead to a significant difference in the performance in terms of medical maturity of the examined graduates of both curricula. An in-depth investigation is recommended, especially concerning the particular personality patterns and outstanding psychometric results found among the participants.

Inhaltsverzeichnis

Abstract	I
Inhaltsverzeichnis	II
Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
Abkürzungsverzeichnis	VII
1. Einleitung	1
2. Stand der Forschung	4
2.1. Kapitelüberblick	4
2.2. Der Begriff der Arztreife	4
2.3. Persönlichkeit	5
2.4. Subjektive Lebenszufriedenheit	7
2.5. Soziale Unterstützung	8
2.6. Angst und Depression	9
2.7. Zusammenfassung	10
3. Fragestellung und Arbeitshypothesen	12
3.1. Vorbemerkung	12
3.2. Konkrete Fragestellungen	12
3.3. Arbeitshypothesen	13
4. Material und Methoden	15
4.1. Methodische Vorarbeiten und Lehrkontext	15
4.2. Outcome Evaluation KUMplusKOM: der Vergleichs-OSCE	15
4.2.1. Literaturrecherche	16
4.2.2. Extraktion der KUMplusKOM-Kompetenzen	16
4.2.3. Konzeption der Vergleichs-OSCE Teststationen	17
4.2.4. Die Vergleichs-OSCE Teststationen	17
4.3. Teilnehmer:innen	21
4.4. Konkrete Durchführung des Vergleichs-OSCE und Fragebogenerhebung	21
4.5. Benotung	22
4.6. Psychometrische Instrumente	23
4.6.1. NEO-FFI 30 (NEO-Fünf-Faktoren-Inventar)	24
4.6.2. Satisfaction with Life Scale (SWLS)	25
4.6.3. F-SoZu K14: Fragebogen zur Erfassung sozialer Unterstützung	26
4.6.4. Hospital Anxiety and Depression Scale (HADS-D)	28
4.7. Soziodemografie und angestrebte Facharztrichtung	29
4.8. Eingabe und Auswertung	29
4.9. Analyseplan und tatsächliches statistisches Vorgehen	30

4.9.1. Kategorisierung der KUMplusKOM-Komponenten	30
4.9.2. Umgang mit geringen und ungleichen Gruppengrößen	30
4.10. Datenschutz und Genehmigung durch die Ethik-Kommission	31
5. Ergebnisse	32
5.1. Allgemeine Beschreibung der Stichprobe	32
5.2. Die Performanz im Vergleichs-OSCE	33
5.3. Vergleich zwischen den Studiengangsformen.....	35
5.4. Angestrebte Facharztrichtung	36
5.5. Psychometrie.....	38
5.5.1. NEO-FFI 30: Erfassung der Persönlichkeitseigenschaften	38
5.5.2. Satisfaction with Life Scale (SWLS): Globalrating Lebenszufriedenheit.....	41
5.5.3. F-SoZu K14: Fragebogen zur Erfassung sozialer Unterstützung	42
5.5.4. Hospital Anxiety and Depression Scale (HADS-D).....	43
6. Diskussion	45
6.1. Überblick: Die wichtigsten Ergebnisse	45
6.2. Methodische Stärken und Schwächen der Arbeit.....	45
6.2.1. Allgemeine methodische Herausforderungen	45
6.2.2. Effekte der Testkonstruktion.....	46
6.2.3. Selektionseffekte	47
6.3. Allgemeine ärztliche Reife und Vergleichs-OSCE	48
6.3.1. Ergebnisse der einzelnen Vergleichs-OSCE-Stationen	49
6.3.1. Wunschfacharztweiterbildung.....	51
6.4. Psychometrische und soziodemografische Ergebnisse	52
6.4.1. Persönlichkeitsmessung.....	52
6.4.2. Lebenszufriedenheit, soziale Unterstützung und Angst und Depression	53
6.4.3. Soziodemografie	55
6.5. Schlussfolgerung	56
7. Zusammenfassung	57
8. Erklärung des Eigenanteils an dieser Arbeit	58
9. Literaturverzeichnis	60
Anhang	VII
Vergleichs-OSCE: Prüfungsanweisungen für die Studierenden, chronologische Abfolge	VIII
Fragebogen, 2. Fassung.....	XII
Einverständniserklärung, Teilnehmer:innenexemplar.....	XVI
Teilnehmer:inneninformation	XVII
Danksagung	XVIII
Lebenslauf	XIX

Abbildungsverzeichnis

Abbildung 1. Reanimationspuppe	18
Abbildung 2. Armmodell zum Üben von Injektionen	19
Abbildung 3. Auszug aus dem auszufüllenden Totenschein in Station 7	20
Abbildung 4. Beispielbild zur Diagnosestellung, Station 7	20
Abbildung 5. OSCE-Schulnotenverteilung unter allen Teilnehmer:innen	34
Abbildung 6. Zusammenhang zwischen Wunschfachrichtung und OSCE-Note	37
Abbildung 7: NEO-FFI 30-Ergebnisse der Teilnehmer:innen, (N=97, N=99)	39
Abbildung 8. Big-5-Ausprägung und Wunschfachrichtung der Stichprobe.....	40

Tabellenverzeichnis

Tabelle 1. Vergleichs-OSCE Teststationen	19
Tabelle 2. Benotung des Vergleichs-OSCEs, angelehnt an Approbationsordnung.....	23
Tabelle 3. Übersicht der verwendeten psychometrischen Instrumente*	24
Tabelle 4. Beschreibung der Stichprobe, N = 100	32
Tabelle 5. Übersicht der OSCE-Leistung der Teilnehmer:innen	35
Tabelle 6. Angestrebte Fachweiterbildungsrichtung der Teilnehmer:innen, N = 98*	36
Tabelle 7. Ergebnisse des NEO-FFI-30, N = 97	38
Tabelle 8. SWLS-Auswertungskategorien und Ergebnisse der Teilnehmer:innen, n = 98	41
Tabelle 9. Gemessene soziale Unterstützung der Stichprobe, n = 74	42
Tabelle 10. Ergebnisse Auswertung HADS-D, n = 74	43

Abkürzungsverzeichnis

EBM.....	Evidenzbasierte Medizin
F-SoZu K14	Fragebogen zur Erfassung sozialer Unterstützung, Kurzform mit 14 Items
HADS-D	Hospital Anxiety and Depression Scale, deutsche Version
HADS-D/A	Hospital Anxiety and Depression Scale, deutsche Fassung, Subskala Angst
HADS-D/D	Hospital Anxiety and Depression Scale, deutsche Fassung, Subskala Depression
KUMplusKOM.....	Klinische Untersuchung Medizin plus Kommunikation
M	Mean, = arithmetisches Mittel
NEO-FFI 30	NEO-Fünf-Faktoren-Inventar, 30 Item Kurzversion
PJ	Praktisches Jahr
PR.....	Prozentrang
RSG.....	Regelstudiengang
SD.....	Standard Deviation, = Standardabweichung
SWB	Subjective Well-Being, subjektives Wohlbefinden
SWLS	Satisfaction with Life Scale (Globalrating Lebenszufriedenheit)

1. Einleitung

Aufbau und konkrete Inhalte der medizinischen Ausbildung sind seit jeher bedeutsame Themen, einerseits, da hochqualifiziertes, kompetentes Fachpersonal von höchster Wichtigkeit im Berufsfeld ist – geht es in diesem Arbeitsfeld doch oft genug buchstäblich um Leben und Tod. Andererseits stehen Ärzt:innen mit den schnell anwachsenden Wissensständen durch den medizinischen Fortschritt vor der besonderen Aufgabe, den Anforderungen moderner Medizin standzuhalten. Zudem zeichnet sich durch den demografischen Wandel in Deutschland ab, dass in den kommenden Jahrzehnten nicht nur Therapeutika und Therapien, sondern auch die häufigsten Krankheitsbilder und Patiententypen weitere Veränderungen erfahren werden – und so auch der Berufsalltag der Behandler:innen.

In den letzten Jahrzehnten wuchs zudem das gesellschaftliche Bewusstsein für die Relevanz kommunikativer Fertigkeiten zur gelungenen Berufsausübung im medizinischen Sektor. Der Erfolg medizinischer Behandlung wird auch maßgeblich durch kommunikative Prozesse, und die Beziehung zwischen Behandler:in und Patient:in beeinflusst. Die Persönlichkeit und der Kommunikationsstil des Arztes oder der Ärztin können wie ein potentes Medikament wirken, Symptome lindern, Adhärenz erhöhen und selbst biophysiological Marker positiv beeinflussen. Kommunikationstechniken wie partizipative Entscheidungsfindung kann man lernen, dennoch gibt es Hinweise darauf, dass auch Persönlichkeitsaspekte und individuelle Belastung eine wichtige Rolle für die Ausprägung von kommunikativen Fähigkeiten spielen.

In den letzten Jahren wurde vielerorts die allgemeine Erschöpfung spürbar, die medizinisches Personal durch seinen körperlich, psychisch und auch kognitiv fordernden Beruf in einem Hochleistungsfeld unter schwierigen Bedingungen erlebt. Streiks, Kündigungswellen und Personalmangel im Gesundheitssystem sind vieldiskutierte Tatsachen in der deutschen und internationalen Nachrichtenlandschaft. Medizinisch tätige Personen mit guten Kommunikationsfertigkeiten und resilienter Persönlichkeit verbleiben länger und zufriedener in Ihrem Beruf.

Über das reine Krankheitswissen hinaus wurde daher zunehmend verstärkt auf die curriculare Vermittlung von Kommunikationsfertigkeiten und Empathieempfinden gegenüber Patient:innen und innerhalb klinischer Teams gelegt; es besteht mittlerweile ein allgemeiner Konsens darüber, dass ein guter Arzt oder eine gute Ärztin mehr als nur aktuelles, evidenzbasiertes klinisches Wissen mitbringen muss. Der zentrale Begriff zur Beschreibung der Befähigung zur erfolgreichen medizinischen Berufsausübung mit allen diesen Facetten ist dabei der sogenannte „Arztreife“.

Am Universitätsklinikum Hamburg-Eppendorf löste 2012 der integrierte Modellstudiengang Medizin Hamburg (iMED) den Regelstudiengang Medizin ab. Basierend auf der neuen Approbationsordnung, orientiert er sich als einer der ersten in Deutschland bei der Vermittlung der

medizinischen Lehrinhalte an modernsten Konzepten für die Ausbildung von Ärzt:innen und setzt unter anderem Empfehlungen des Wissenschaftsrates für die Weiterentwicklung des Medizinstudiums um (Wissenschaftsrat, 2014). Studiendauer und Zulassungskriterien blieben (zunächst) gleich. An der interdisziplinären mehrjährigen Entwicklung des Curriculums waren sowohl Lehrpersonen als auch Student:innen beteiligt, mit dem Ziel, bereits zu Studienbeginn die theoretischen Grundlagenfächer mit der klinischen Praxis zu verschränken. Insbesondere wurde unter anderem ein besonderer Schwerpunkt auf die integrierte Vermittlung von kommunikativen und praktischen Fertigkeiten gelegt.

Dem Gegenüber stand das Vorgehen im bisherigen Regelstudiengangsystem (=“RSG“): beginnend mit einem stark naturwissenschaftlich ausgerichteten Lehrplan, der nach 4 Semestern mit dem sogenannten Physikum abgeschlossen wurde, einer umfangreichen Prüfung mit Schwerpunkten auf Biophysik und anderen naturwissenschaftlichen Themen, an der viele Studierende scheiterten. Viele angehende Mediziner:innen hatten durch den frühen naturwissenschaftlichen Schwerpunkt, der sich erst in der zweiten Studienhälfte zu angewandter Medizin am Menschen wandelte, sehr spät ihren ersten Patientenkontakt in ihrer Rolle als angehende Ärzt:innen. Kommunikation wurde in Hamburg wie an den meisten deutschen Universitäten zunächst nur unsystematisch und vereinzelt unterrichtet, da dieses Fach noch nicht lange an Bedeutung gewonnen hatte und in Konkurrenz zu den zahlreichen anderen zu vermittelnden Inhalten stand.

Mit der Revision des Lehrplanes für das iMED-Curriculum sollte ein zeitgemäßerer, praktischerer Studiengang entstehen, in dem eine frühe Synthese aus Theorie, Klinik und Praxis mit besonderer Beachtung modernerer Unterrichts- und Prüfungsformen besonders qualifizierte Ärztinnen und Ärzte hervorbringen sollte. Seit Winter 2017 wird ausschließlich nach dem iMED Curriculum gelehrt. Nach einer mehrjährigen Übergangszeit, in der beide Systeme noch parallel bestanden, ist das Gros der letzten regulären RSG-Absolvent:innen zum Jahr 2017 erfolgreich verabschiedet worden.

In dieser Arbeit werden die durch das medizinische Curriculum am Universitätsklinikum Hamburg-Eppendorf erzeugten Unterschiede in der Arztreife von Absolvent:innen beider Studiengangsformen erstmals genauer untersucht, mit dem Ziel, belastbare Aussagen über die Arztreife der Testpersonen und Empfehlungen für die weitere curriculare Ausgestaltung geben zu können. Dazu wurde eine vergleichende Testung von Absolvent:innen beider Studiengänge in einer standardisierten Prüfungssituation durchgeführt. Auch werden die relevanten psychometrischen und soziodemografischen Eigenschaften der Studierenden beleuchtet. Die Arztreife, also eben die Synthese aus klinisch-praktischem Wissen und kommunikativen Fertigkeiten ist zentral für das Wohlergehen der Patient:innen und daher von elementarer Wichtigkeit für angehende Ärzt:innen. Aus diesen Ergebnissen kann abgeleitet werden, inwieweit die aktuelle medizinisch-universitäre Ausbildung in Hamburg Absolvent:innen hervorbringt, die dem

Ideal optimal klinisch ausgebildeter und sensibel kommunizierender Ärzt:innen entsprechen. Weiter können die Ergebnisse Kriterien für die Auswahl zukünftiger geeigneter Kandidat:innen an die Hand geben. Damit ist die Relevanz nicht nur für Patient:innen und Gesundheitssystem, sondern auch für den Bildungssektor groß.

Nach einer umfassenden Schilderung zunächst des Forschungsstandes in den relevanten Themenfeldern, werden die konkreten Fragestellungen und Arbeitshypothesen vorgestellt, die dieser Arbeit zugrunde lagen. Im Anschluss folgt die eingehende Beschreibung des verwendeten Materials und der Methoden. Darauf aufbauend werden die Ergebnisse erläutert, und anschließend intensiv diskutiert. Die Arbeit endet mit einer kurzen Zusammenfassung aller relevanten Inhalte.

2. Stand der Forschung

2.1. Kapitelüberblick

Nach der einleitenden Erläuterung wird in den direkt folgenden Unterkapiteln nun unter Bezugnahme der relevanten aktuellen Forschungsliteratur der Rahmen für die Methodik dieser Arbeit erläutert, die nach den Fragestellungen vorgestellt wird. Jedes Unterkapitel beschreibt ein Kriterium der Testung. Dabei richtet sich die Reihenfolge der Darstellung hier zunächst nach der Bedeutsamkeit für die Fragestellungen und anknüpfend nach der zeitlichen Reihung der Testungsinhalte. Es werden, wo notwendig, sowohl die theoretischen Grundkonzepte näher beleuchtet, als auch ein erster Überblick über die für die Arbeit erhobenen konkreten psychometrischen Skalen und ihre besondere Notwendigkeit für die empirische Untersuchung des Themenkomplexes geboten.

Im Anschluss an die spätere Vorstellung der Ergebnisse wird wieder auf das hier geschilderte empirische Ausgangsmaterial eingegangen und, wenn bedeutsam, die besonderen Herausforderungen jedes ebenfalls kapitelweise diskutiert.

2.2. Der Begriff der Arztreife

Der eigentliche Erfolg medizinischer Ausbildung wird oftmals anhand akademischer Leistungen bestimmt, obwohl dies tatsächlich ein unzulänglicher Prädiktor für die spätere ärztliche Kompetenz zu sein scheint (Ferguson et al., 2007). Will man also sinnvollere und belastbarere Operationalisierungen für Studienerfolg von Absolvent:innen wählen, braucht es zunächst einmal ganz allgemein objektive, reliable und valide Kriterien, um diesen sinnvoll messen zu können. Im besonderen Falle des Medizinstudienabschlusses bietet sich dafür der Begriff der „(allgemeinen) Arztreife“ oder „Allgemeinen Ärztlichen Reife“ an. Dabei handelt es sich um ein Konzept, das im deutschsprachigen Raum zuletzt 2003 als Antwort auf die neue Approbationsordnung für Ärzte (ÄAppO, Güntert, 2003) für den Hamburger Lernzielkatalog KliniCum eingeführt wurde, um einen schwierig zu definierenden Begriff in untersuchbare, lehrbare und damit letztlich auch abprüfbare Worte zu fassen (Kuhnigk et al., 2006). Mit Arztreife gemeint ist damit zunächst einmal eine nach erfolgreichem Abschluss des Studiums fundierte Befähigung der Absolvent:innen zu medizinischen „Generalist:innen“, die dazu in der Lage sind, qualifizierte Ersteinschätzungen zu den häufigsten Erkrankungsbildern und Symptomen abgeben und die Behandlung dieser einleiten zu können.

Die bei den meisten Studierenden an den Studienabschluss anschließende mehrjährige Facharztweiterbildung vertieft dann die Kenntnisse in den 34 spezialisierteren Grunddisziplinen wie Chirurgie, Urologie oder zum Beispiel Psychosomatischer Medizin und Psychotherapie, oder

bereitet mit dem, wenn man so will, generalistischen Allgemeinmedizinischen Facharzt umfänglich auf eine spätere hausärztlichen Tätigkeit in Theorie und Praxis vor (siehe Muster Weiterbildungsordnung der Bundesärztekammer, 2018).

In der Definition des Arztreifebegriffs wurde bewusst von der Forderung nach überwiegend konkreten Detailkenntnissen Abstand genommen, um den sich mit dem medizinischen Fortschritt ständig wandelnden und anwachsenden Anforderungen gerecht zu werden (Kuhnigk et al., 2006, Seite 5). Dies hat natürlich zur Folge, dass die konkrete Operationalisierung dieses Kriteriums stark nach Universität und Lehrplan variiert. Im Folgenden beschränkt sich die nähere Operationalisierung in diesem Dokument daher auf die KUMplusKOM-Vorarbeiten am Hamburger Vorbild, da diese Arbeit an Hamburger Studierenden erhoben wurde und, wenn man so will, der Begriff der Hamburger Ärztlichen Reife im Zentrum der Erhebung steht (für eine ganz konkrete Operationalisierung an geeigneterer Stelle siehe Methodenkapitel 4.2.2ff). Am Universitätsklinikum Hamburg Eppendorf bekam der Begriff der Allgemeinen Arztreife mit der Konzeption des Curriculums des Modellstudiengangs eine besondere Bedeutung, da er eben dort als zentrales Zielkriterium des Studienerfolgs benannt wird (ohne besonders definiert zu werden): *„Ziele des Studiums der Medizin in Hamburg: Die Studierenden sollen eine allgemeine Arztreife erreichen. Die Studierenden sollen während des Studiums individuelle Schwerpunkte entwickeln.“* (Hamburger Lernzielkatalog, 2009). Aus eben diesem Grund erscheint dieses Kriterium als Grundlage unserer über Abschlussnoten hinausgehende Messung des Studienerfolgs an Hamburger Medizinstudierenden besonders geeignet. Da zum Zeitpunkt der Materialsammlung und –Erstellung für diese Arbeit keine passenden Instrumente zur Erhebung der Arztreife von Medizinstudiengangsabsolvent:innen vorlagen, wurden diese eigens dafür von den Hamburger Lehrexpert:innen konzipiert und erstmals verwendet (siehe Konzeption der Vergleichs-OSCE-Stationen).

2.3. Persönlichkeit

2009 konnte von Poropat et al. in einer umfangreichen Metaanalyse mit kumulativen Stichprobengrößen von über 70000 Personen zum wiederholten Male ein bedeutender Einfluss der Persönlichkeit (insbesondere einzelner Facetten, siehe unten) als Determinante für (universitären) Bildungserfolg nachgewiesen werden (Bergold & Steinmayr, 2018; Mayer & Skimmyhorn, 2017). Auch Zusammenhänge zwischen Persönlichkeit und Entscheidungen zu Studien-, oder späterer Facharzttrichtung wurden belegt (zum Beispiel (Barbosa-Camacho et al., 2020; Jafrani et al., 2017; Nawaiseh et al., 2020). Empathievermögen und kommunikative Fertigkeiten, beides wichtige Eigenschaften für kompetente angehende Ärzt:innen, korrelieren ebenfalls stark mit verschiedenen Persönlichkeitsmaßen (vergleiche zum Beispiel(Guilera,

Batalla, Forne, & Soler-Gonzalez, 2019). Damit scheint ein deutlicher Zusammenhang zwischen Persönlichkeit und Bildungs- sowie Berufserfolg sicher empirisch belegt.

Dabei werden vielfältige, teilweise sehr unterschiedliche und vor allem vieldiskutierte Definitionen und Konzepte zur Operationalisierung des Persönlichkeitsbegriffes zugrunde gelegt. Einigkeit herrscht darüber, dass sich der psychologische Persönlichkeitsbegriff grundsätzlich auf zeitstabile individuelle Eigenschaften bezieht, sogenannte „*traits*“ – Eigenschaften also, die sich im Verlauf eines Menschenlebens nur minimal und langsam bis gar nicht verändern (Mischel, 2004). Diese traits müssen von Individuen nicht nur überwiegend stabil über lange biografische Zeiträume, sondern auch über viele verschiedene Situationen gezeigt und berichtet werden, um die Kriterien für Persönlichkeitseigenschaften erfüllen zu können. Dabei manifestiert sich ihr Einfluss in einer Person sowohl auf der emotionalen, als auch auf Kognitions- und der Verhaltensebene (Montag, 2016).

In der Differentiellen Psychologie hat sich besonders das sogenannte Fünf-Faktoren-Modell der Persönlichkeit (auch „*Big Five*“) nach Costa & McCrae (1992) als gängigstes, gut überprüfbares und vielfach empirisch bestätigtes Modell zur Abbildung individueller Persönlichkeitsunterschiede etablieren können. Die Entwicklung dieses Konzept begann bereits in den 1930er Jahren durch Thurstone et al. und wurde schließlich durch Costa & McCrae faktorenanalytisch in seine heutige Form überführt. Dabei wird davon ausgegangen, dass sich die wichtigsten Persönlichkeitsaspekte jedes Menschen hinreichend mittels fünf verschiedener unabhängiger (kultur-)stabiler Facetten oder Hauptdimensionen abbilden lassen: Neurotizismus (emotionale Labilität), Extraversion (Geselligkeit), Offenheit für neue Erfahrungen (Aufgeschlossenheit), Gewissenhaftigkeit (Perfektionismus) und Verträglichkeit (Kooperationsbereitschaft, Empathie und Rücksichtnahme) (Asendorpf & Neyer, 2012).

Es gibt vielerlei Belege für eine besonders große Bedeutsamkeit der individuellen Ausprägungen insbesondere der Persönlichkeitsfacetten Gewissenhaftigkeit als auch Neurotizismus für akademischen Erfolg (Hayat, 2020; Poropat 2009 und Vedel & Poropat 2017), wobei insbesondere in der medizinischen Ausbildung hohe Gewissenhaftigkeitswerte und niedrige Neurotizismusausprägungen von Vorteil zu sein scheinen (Sobowale, Ham, Curlin, & Yoon, 2018). Eine Fülle gängiger, gut evaluierter und beständig weiterentwickelter moderner Persönlichkeitstests basieren auf der Big-Five-Grundannahme. Aus eben diesem Grund haben wir uns bei der Methodik dieser Arbeit für dieses Konstrukt zur Messung von Persönlichkeitseigenschaften entschieden.

2.4. Subjektive Lebenszufriedenheit

Das Konzept der Subjektiven Lebenszufriedenheit (oder im englischsprachigen Raum auch „*happiness*“, also Glücklichkeit) hat in den letzten Jahrzehnten im Zusammenhang mit Maßnahmen des öffentlichen Gesundheitswesens als potentiell wichtiges Kriterium für Bevölkerungsgesundheit erheblich an empirischer Bedeutung gewonnen (Cooke, Melchert, & Connor, 2016).

Subjektive Lebenszufriedenheit als empirisch untersuchbares psychologisches Konzept wird definiert über die Annahme, dass Individuen (zumindest in großen Teilen) unabhängig von ihren tatsächlichen Lebensumständen, sondern vielmehr durch ihre eigenen subjektive Einschätzungen, Haltungen und Gefühle ihre Lebensqualität und Zufriedenheit damit bestimmen (Kusier & Folker, 2021). Dabei werden also selbst so sachlich bedeutsame Aspekte wie Einkommen, Gesundheit oder Berufstätigkeit als Entscheidungsgrundlage vernachlässigt, und einseitig verzerrt emotionale oder idealistische Kriterien unwillkürlich überbetont. Schimmack et al. (2002) konnten der Annahme folgend tatsächlich an amerikanischen College Studierenden belegen, dass diese unwillkürlich und stabil an ihrer Lebenswahrnehmung in den Bereichen Familie, Beziehungen (romantisch wie freundschaftlich) und akademischen Erfolg fest machten, wie zufrieden sie mit ihrem Leben waren – und bei ihrer Beurteilung der eigenen Lebenszufriedenheit eben Gesundheit und finanzieller Situation weitaus weniger Bedeutung zumaßen. Welche Aspekte die Lebenszufriedenheit maßgeblich prägen, verändert sich nach Lebensalter, Umfeld und sogar Primingeinflüssen¹ (Oishi et al., 2003). Luhmann et al. (2013) konnten nachweisen, dass die allgemeine subjektive Lebenszufriedenheit in jüngeren Lebensalter ein starker Prädiktor für die Lebenszufriedenheit in den folgenden Jahren und mit der Wahrscheinlichkeit für das Eingehen einer Ehe und der Familiengründung assoziiert ist – unabhängig von Geschlecht, Alter, sozioökonomischen Status und Persönlichkeit. Familienstand und soziale Eingebundenheit wiederum sind vielfach als bedeutsame Vorhersagekriterien für akademischen Erfolg nachgewiesen worden (s.u., Harandi et al, 2017).

Messungen Subjektiver Lebenszufriedenheit haben sich in Querschnitts- und Langzeitstudien als relevante Einflussgröße auf verschiedenste Lebensbereiche erwiesen, mit positiven Korrelationen zwischen höherer Zufriedenheit und allgemein besserer Gesundheit (Diener et al., 2011), verzögerter Mortalität (Oishi, Diener, & Lucas, 2007) und höherem Berufserfolg (Lyubomirsky, S., King, L., Diener, E. , 2005). Die Bedeutung des Konzeptes als stabile,

¹ Priming, seltener zu Deutsch auch Bahnungseffekt, ist eine psychologisch-neurophysiologische Begrifflichkeit. Diese beschreibt das Phänomen, dass die Reizverarbeitung von Individuen durch vorangegangene, meist unbewusste Konfrontation mit ähnlichen oder gleichen Reizen gebahnt und beschleunigt wird. Dadurch ergibt sich eine bessere oder schnellere Verarbeitung der Reize, mit teils umfangreichen Verhaltenskonsequenzen, angeblich durch die Verknüpfung der Reize mit entsprechenden Assoziationen (Myers, 2008). Für eine nähere Diskussion des nicht ganz unumstrittenen Konzeptes siehe zum Beispiel Kahnemanns, 2012.

psychometrisch messbare Einflussgröße (neben der individuellen Persönlichkeit und demografischen Aspekten, s.u.) auf biografische Meilensteine betont seine Relevanz bei der Erhebung von Moderator- und Mediatorvariablen auf akademischen Erfolg und erklärt die Entscheidung, es als wichtiges Kriterium in die Erhebung einzuschließen.

2.5. Soziale Unterstützung

Dem sozialen Netzwerk einer Person wird zunehmend eine beachtliche Bedeutung als Einflussgröße auf psychische (Harandi et al, 2017) und somatische Verfassung (H.-H. Wang, Wu, & Liu, 2003) des Individuums zugeschrieben, wobei sowohl die protektive Wirkung guter sozialer Netzwerke und damit hohe soziale Unterstützung belegt werden konnte, als auch der adverse Einfluss unzureichender sozialer Netzwerke auf verschiedenste Gesundheitsoutcomes wie endokrinologische oder kardiovaskuläre Gesundheit, Immunresonanz bis hin zu Mortalitätsrisiko (Callaghan & Morrissey, 1993). Dabei sind die Mechanismen, nach denen der Einfluss sozialer Unterstützung auf Individuen einwirken könnte, um diese Outcomes zu moderieren oder gar zu verursachen, seit Jahren Quelle anhaltender Diskussion (vergleiche zum Beispiel Thoits, 1985).

Die Begrifflichkeit „soziale Unterstützung“ betont heute vor allem qualitative Aspekte menschlicher Interaktion. Dabei kann die Hilfeleistung in instrumentelle (zum Beispiel materielle Hilfen), emotionale (Trost und Aufmerksamkeit) und informationale Unterstützung, zum Beispiel durch Lösungsvorschläge, weiter untergliedert werden und umfasst so eine Vielzahl an Facetten hilfreicher sozialer Interaktion (Klauer, 2009).

Eine populäre mögliche Erklärung für den Einfluss sozialer Unterstützung auf Wohlbefinden und Gesundheit ist die „Buffer-Theorie“ (Antonovsky, 1974) nach der soziale Netzwerke Stressoren gegenüber gut eingebundenen Personen buchstäblich abfedern (englisch: „to buffer“) und so negative Auswirkungen verhindern. Dem gegenüber steht die „Attachment²-Hypothese“, die nach Bowlby & Ainsworth (1978) die oben aufgeführten Effekte so erklärt: in der Kindheit sicher an Eltern oder andere ähnlich bedeutsame Bezugspersonen gebundene Individuen können auch im späteren Lebensalter stabile, sozial supportive Bindungen eingehen und von ihnen profitieren.

In der hier vorgelegten Arbeit liegt ein Fokus auf medizinischen Weiterbildungserfolg, gemessen an unserer Operationalisierung der Arztreife durch einen umfangreichen praktisch-theoretischen Fertigkeitstest, deshalb liegt eine besondere Berücksichtigung der Studienlage zu den Zusammenhängen sozialer Unterstützung und akademischen Erfolgs nahe. So konnten etwa

² Attachment wird in diesem Zusammenhang mit Bindung übersetzt.

(Stadtfeld, Vörös, Elmer, Boda, & Raabe, 2019) zeigen, dass sozial isolierte Ingenieursstudierende signifikant niedrigere Noten aufwiesen und mit höherer Wahrscheinlichkeit ihr Studium abbrachen, als sozial gut eingebundene Kommiliton:innen. Zahlreiche Studien konnten die Beziehung zwischen sozialer Unterstützung und Stressbelastung (und damit in der Konsequenz sogar akademischen Erfolges und psychologischer Gesundheit) unterschiedlichster Studierendengruppen verschiedenster Studienbereiche belegen (Wilks, 2008).

Auch der Einfluss von risikoreichem Verhalten auf akademischen Erfolg konnten mit sozialer Unterstützung in Beziehung gesetzt werden (Wong, 2017) genauso wie der problematische Konsum von Drogen oder delinquentes Verhalten von Jugendlichen mit geringer sozialer Unterstützung (McGuire et al, 2021).

Zusammenfassend muss die große Bedeutsamkeit hoher sozialer Unterstützung für akademischen Erfolg unweigerlich zu einem Einschluss dieses Kriteriums als Einflussgröße bei jedweder Form der Untersuchung medizinischen Bildungserfolgs führen – so auch in dieser Arbeit.

2.6. Angst und Depression

Medizinisches Personal ist in seiner Berufsausübung sowohl durch die eigentliche professionelle Funktion als auch durch systemimmanente Herausforderungen einer beständig hohen Arbeitslast ausgesetzt (Fishbein et al., 2019). Dabei werden nicht nur die kognitiven Kapazitäten, sondern oft emotionale und zuweilen auch körperliche Ressourcen gefordert und überanspruchert (Grochowska, Gawron & Bodys-Cupak, 2022). Daher erfahren medizinisch tätige Personen absolut und verglichen mit vielen anderen Berufsgruppen enorm viel andauernden negativen Stress, also eine erhöhte physiologische und mentale Anspannung und Belastung (LeBlanc, 2009). Dies kann schwerwiegende Konsequenzen für die psychische Gesundheit der Betroffenen haben und mitunter sogar im Auftreten von psychiatrischen Diagnosen wie Depressionen und Angststörungen gipfeln (Dyrbye et al., 2014).

Verschiedene Arbeiten erbrachten Hinweise zur Bedeutung von Angst und Depression für akademische Leistung und allgemeine mentale Gesundheit schon bei Schüler:innen und Studierenden: Medizinstudierende, die schon zu Studienbeginn von hohen Ängstlichkeits- und Depressionsausprägungen berichteten, wiesen im Fortlauf ihrer Ausbildung stark erhöhte Risiken für die Entwicklung noch stärkerer Symptome bis hin zu Depressions-, Angststörungs- oder Burnouterkrankungen auf (Walkiewicz et al., 2012). Depression und Ängstlichkeit korrelieren mit schlechteren Noten und einer höheren Wahrscheinlichkeit eines vorzeitigen Studien- oder Schulabbruchs (Eisenberg et al., 2009). Nyer et al. (2013) konnten in einer Untersuchung an 287 amerikanischen Studierenden wenig überraschend zeigen, dass etwa Schlafstörungen mit dem Auftreten von Depressions- und Angstsymptomen und einer Reduktion von Funktio-

nalität und akademischer Leistung einhergehen, auch wenn keine klinische Depressionsdiagnose gegeben war. Schlafstörungen treten in unterschiedlichen Formen häufig unter Studierenden auf, mit Spannweiten von 13 – über 60 % in verschiedenen Untersuchungen (Nadorff, Nazem & Fiske, 2013; Lund et al., 2010).

Die nationalen Auswahlverfahren zur Medizinstudienplatzvergabe konzentrieren sich traditionell überwiegend auf hohe Performanz im Sinne hervorragender Schulabschlussnoten und ähnlicher Maße für kognitive Leistungsfähigkeit, in deutlich geringerem Umfang auf Resilienz oder vergleichbare „*soft skills*“ und psychische Schutzfaktoren (Hissbach et al., 2012; Meyer et al., 2019). Da, wie oben ausgeführt, der spätere Berufserfolg und die Zufriedenheit maßgeblich aber insbesondere mit der Widerstandskraft gegenüber den Herausforderungen des Berufsbildes zusammenzuhängen scheinen (Epstein & Krasner, 2013; (Van Hoek, Portzky, & Franck, 2019); Walkiewicz et al. 2012), liegt eine genauere Untersuchung dieses Themenfeldes zu Beginn der Berufskarriere besonders nahe und führte zum Einbezug entsprechender Erhebungsinstrumente in den psychometrischen Studienteil dieser Arbeit.

2.7. Zusammenfassung

Zum Zeitpunkt der Veröffentlichung dieser Arbeit ist das im Folgenden weiter erläuterte Vorgehen des umfassenden querschnittlichen Vergleichs der Absolvent:innen zweier moderner, sorgsam konzipierter Hamburger Medizinstudiengangsformen einzigartig. Dies begünstigt einerseits die regionale Qualitätssicherung in der Lehr- und Prüfungserstellung am Universitätsklinikum Hamburg-Eppendorf.

Andererseits ist insbesondere der Schwerpunkt auf eine umfangreiche und fundierte praktische und theoretische Abbildung der berufsrelevanten Kompetenzen angehender Ärzt:innen durch erfahrene Expert:innen in Hinsicht auf moderne Curriculumsentwicklung von hoher Bedeutung für die zukünftige Gestaltung und den Erfolg medizinischer universitärer Lehre im Allgemeinen. Die in Deutschland in dieser Form erstmalige Überführung des Begriffes der Arztreife in messbare konkrete Fertigkeiten und die anschließende empirische Überprüfung an geeigneten Testpersonen hat das Potenzial für weitreichende Implikationen für die zukünftige nationale Lehrforschung im medizinischen Sektor.

Ein besonderes Augenmerk ist in diesem Sinne auf die aufwendige Konzeption des praktischen OSCE`s zu legen, dessen daraus resultierende hohe Inhaltsvalidität, Objektivität und Reliabilität insbesondere im deutschen Sprachraum als recht einzigartig zu beurteilen ist und sich einer Prüfungsform bedient, die sich auch international seit Jahrzehnten wachsender Beliebtheit erfreut und gut etabliert ist – und die Möglichkeit zur ökonomischen gleichzeitigen

Untersuchung auch kommender großer Studierendengruppen bietet. Damit ist die Replizierbarkeit der Methodik besonders erleichtert und dem empirischen Wissenszuwachs eine gute Plattform geboten.

Die weitere Erhebung und In-Bezug-Setzung relevanter Stör-, Mediator- und Moderatorvariablen wie den oben ausgeführten psychometrischen Aspekten Persönlichkeit, subjektiver Lebenszufriedenheit, sozialer Unterstützung, Depressivität und Ängstlichkeit sowie umfassender soziodemografischer Daten ermöglicht eine besonders intensive und forschungsrelevante Auswertung unter Berücksichtigung des aktuellen Forschungsstandes. Wir wissen aus der Literatur, dass soziodemografischen Unterschiede zwischen Personen in Deutschland einen Einfluss auf das erlangte Bildungsniveau haben können und als wichtige Kontrollvariable mit erhoben werden sollten (Geissler, 2006; Pollak, 2010).

Damit ist die Vorstellung der Hauptkonstrukte abgeschlossen. Es ist, wie bereits erläutert, von allen anzunehmen, dass sie einen Einfluss auf die spätere Performanz von Ärztinnen und Ärzten haben, daher werden sie in dieser Arbeit als weitere Testkriterien fungieren.

3. Fragestellung und Arbeitshypothesen

3.1. Vorbemerkung

Mit der Einführung des iMED Modellstudiengangs veränderte sich 2012 das Studium der Humanmedizin in Hamburg umfassend mit dem Ziel, nach modernsten Ausbildungsstandards zukunftsfähige Ärzt:innen heranzubilden: nun lernen die Student:innen schon ab dem ersten Semester sowohl theoretische, als auch klinisch-praktische Lehrinhalte integriert, also zeitgleich und ineinander verschränkt kennen. Besonders betont wird dabei die frühe Auseinandersetzung und Übung der Lernenden von kommunikativen und praktischen Fertigkeiten im Rahmen des KUMplusKOM Lehrkonzeptes. Bisher waren die Inhalte im Regelstudiengang chronologisch vermittelt worden, zunächst die naturwissenschaftlichen Grundlagen, dann erst sehr viel später klinisch-praktische und kommunikative Themen. Ein potentieller Mangel, da diese Schwerpunktsetzung möglicherweise auf Kosten entsprechender Fertigkeiten der Studierenden ging, und der daher mit der Überarbeitung des Lehrplans abgestellt werden sollte. Die vorliegende Arbeit dient der Outcome-Evaluation der beiden Curricula unter besonderer Beachtung des KUMplusKOM Lehrkonzeptes: wir untersuchen den Ausbildungserfolg anhand der Performanz zweier Studierendenabschlussgruppen sowohl des neuen Lehrkonzeptes als auch des etablierten Regelstudiengangs im direkten Vergleich.

Es soll geprüft werden, ob die Leistungen der Student:innen sich direkt vor Eintritt ins Praktische Jahr hinsichtlich ihrer Arztreife und insbesondere in Bezug auf die Integration der Kompetenzen, ihr fachliches Wissen, sowie kommunikative und praktisch-klinischen Fertigkeiten zwischen dem Modell- und dem Regelstudiengang unterscheiden und welche ihrer Eigenschaften die Performanz maßgeblich beeinflussen.

3.2. Konkrete Fragestellungen

Es werden die folgenden Fragestellungen untersucht:

1. Unterscheiden sich die Absolvent:innen der beiden Studiengangsformen iMED und RSG hinsichtlich Ihrer Gesamtperformanz, gemessen an der Gesamtpunktzahl, in einem Test zur Arztreife (also dem hier vorgestellten Vergleichs-OSCE³)?

- a. Erreichen die Absolvent:innen des iMED-Curriculums bessere Noten in den Stationen, in denen kommunikative und praktische Fertigkeiten integriert abgeprüft werden

³ Das Akronym „OSCE“ steht für „*objective structured clinical examination*“ und beschreibt eine strukturierte klinisch-praktische Prüfung, die ähnlich einem Zirkeltraining in verschiedene Stationen aufgebaut ist (Harden 1988).

(insbesondere in der Station „Somatisierungsstörung“, aber auch bei den Stationen „Blutdruckberatung“, und „Untersuchung Abdomen“)?

b. Unterscheiden sich die Absolvent:innen beider Studiengänge (iMED und RSG) hinsichtlich ihrer in der Prüfungssituation gezeigten kommunikativen und praktischen Fertigkeiten?

2. Inwieweit ist die Persönlichkeit der Student:innen mit der OSCE-Performance assoziiert?

a. Zeigen Personen abhängig von ihren Ausprägungen einzelner erhobener Persönlichkeitsvariablen (Neurotizismus, Extraversion, Offenheit für neue Erfahrungen, Gewissenhaftigkeit sowie Verträglichkeit, ermittelt mit dem NEO-FFI 30) bessere Ergebnisse, also höhere Gesamtpunktzahlen, im Test zur Arztreife? Oder höhere Punktzahlen in einer oder mehreren der acht einzelnen Teststationen (also „Blutdruckberatung“, „Somatisierungsstörung“, „Venöse Blutentnahme“, „Schwindel“, „Reanimation“, „Untersuchung Abdomen“, „Totenschein“ oder „Evidenzbasierte Medizin“)?

b. Wie ist die individuelle Performanz der Teilnehmer:innen, gemessen an den erreichten Punktzahlen auf den Einzelstationen, abhängig von den durch den NEO-FFI30 ermittelten gesamten Persönlichkeitsprofilen (also den Ausprägungen für alle 5 psychometrische Skalen gleichzeitig)?

3. Präferieren unterschiedliche Persönlichkeitsprofile unterschiedliche Weiterbildungsfächer?

a. Gibt es einen Zusammenhang zwischen Weiterbildungspräferenz und OSCE-Performanz?

4. Inwiefern sind Unterschiede in der OSCE-Performance mit soziodemografischen Charakteristika (also Geschlecht, Alter, Familienstand, Bildungsstand, und Migrationshintergrund) und den erhobenen weiteren psychometrischen Skalen (Subjektive Lebenszufriedenheit, gemessen mit dem SWLS; Soziale Unterstützung, gemessen mit dem F-SoZu K14, sowie Angst und Depression, gemessen mit dem HADS) assoziiert?

3.3. Arbeitshypothesen

1. Der Ausbildungserfolg der Absolvent:innen des iMED-Studiengangs, gemessen an der Gesamtpunktzahl im Vergleichs-OSCE sollte höher ausfallen, als der der RSG-Absolvent:innen, da durch das oben ausgeführte integrierte KUMplusKOM-Curriculum die iMED-Absolvent:innen besser zielgerichtet auf eine höhere ärztliche Reife hin ausgebildet wurden.

2. Es wird entsprechend dem aktuellen Forschungsstand erwartet, dass insbesondere Personen mit hohen Ausprägungen der Persönlichkeitsvariable Gewissenhaftigkeit bessere Ergebnisse, also höhere Gesamtpunktzahlen, im Test zur Arztreife (also dem Vergleichs-OSCE)

erzielen, als solche Personen, die niedrigere Ausprägungen aufweisen. Ebenso sind bessere Ergebnisse in einer oder mehreren der acht einzelnen Teststationen (also „Blutdruckberatung“, „Somatisierungsstörung“, „Venöse Blutentnahme“, „Schwindel“, „Reanimation“, „Untersuchung Abdomen“, „Totenschein“ oder „Evidenzbasierte Medizin“) erwartbar.

3. Analog dazu wird erwartet, dass Personen mit hohen Neurotizismuswerten gegenüber solchen mit niedrigeren Ausprägungen insgesamt schlechter im Test zur Arztreife abschneiden.

4. Es wird davon ausgegangen, dass Testpersonen mit hohen Extraversionsausprägungen sich für andere Weiterbildungsfächer entscheiden (also solche mit mehr Patient:innenkontakt und Teamarbeit), als Personen mit geringen Ausprägungen des Persönlichkeitsaspektes.

5. Es wird ebenso erwartet, dass der soziodemografische Hintergrund der Testpersonen Einfluss auf die Performanz, gemessen sowohl an der Gesamtpunktzahl im Vergleichs-OSCE als auch der erreichten Punktzahl auf den Einzelstationen, haben wird.

6. Personen, die von hoher subjektiver Lebenszufriedenheit im SWLS berichten, solche, die über viel soziale Unterstützung gemessen mit dem F-SoZu K14 verfügen und Personen, die geringe Werte in der Testung auf Angst- und Depressionssymptome mittels HADS erzielen, sollten höhere Gesamtpunktzahlen im Vergleichs-OSCE erreichen, als diejenigen, die in diesen Bereichen wenig Ressourcen aufweisen.

4. Material und Methoden

4.1. Methodische Vorarbeiten und Lehrkontext

Das im Folgenden erläuterte empirische Vorgehen und die anschließend dargestellte eigentliche Datenerhebung knüpfen eng an theoretische Vorarbeiten zur Entwicklung eines Prüfungs- und Messinstruments zur Outcome-bezogenen Evaluation des KUMplusKOM Lehrkonzepts im iMED Modellstudiengang an.

KUMplusKOM steht für „Klinische Untersuchung Medizin plus Kommunikation“ und betont im Curriculum des Modellstudiengangs iMED am Universitätsklinikum Hamburg-Eppendorf den besonderen Schwerpunkt auf einen integrierten Erwerb kommunikativer und praktischer Fertigkeiten.

Um den Erfolg des neuen KUMplusKOM-Curriculums wissenschaftlich zu überprüfen, wurde 2014, finanziert durch eine Förderung des Förderfonds Lehre, das theoretische Fundament zur praktischen Überprüfung der ärztlichen Fertigkeiten konzipiert. Auf diesem baut diese Arbeit auf, und schließen die Datenerhebung und Analyse direkt an.

4.2. Outcome Evaluation KUMplusKOM: der Vergleichs-OSCE

2014 begannen die Vorarbeiten und letztlich die konzeptionelle Planung zum sogenannten Vergleichs-OSCE. Das einem Zirkeltraining ähnliche Prüfungsformat OSCE ist Medizinstudent:innen aus beiden Hamburger Studiengangsformen sehr vertraut, da es bereits früh in verschiedenen Fächern Anwendung findet, seit Jahrzehnten international bewährt und zum typischen Prüfungsformat im Medizinstudium geworden ist. Dabei besteht ein OSCE klassischerweise aus verschiedenen aufeinander folgenden Teststationen, auf denen in der Regel unterschiedliche Themenbereiche zumeist praktisch, seltener auch schriftlich oder in Interviewform standardisiert abgeprüft werden. Zumeist wird mit einem OSCE zum Semesterende der Wissensstand zu einzelnen erfolgten Lehrmodulen oder Fächern ermittelt, seltener auch als bedeutsamere Zwischenprüfung wie im iMED-Studiengang mit der Physikumsäquivalenzprüfung, die die tradierte Physikumsprüfung nach dem 4. Semester im Medizinstudium ablöst und auch in der erneuerten Form eine wichtige Zäsur im Studium darstellt.

Bei dem Vergleichs-OSCE sollten die Absolvent:innen des Regelstudiengangs mit denen des iMED-Studiengangs hinsichtlich ihrer Allgemeinen Ärztlichen Reife (siehe 2.2) und insbesondere ihrer KUMplusKOM-Kompetenzen geprüft und verglichen werden.

Die Vorarbeiten zur Durchführung des Vergleichs-OSCE wurden auf vier konsekutive Abschnitte hin angelegt: (1) eine systematische Literaturrecherche zum Begriff der „Arztreife“ und ihrer Operationalisierung und edumetrischen Anwendung, darauffolgend (2) die Identifikation

und Kategorisierung der im iMED-Curriculum verankerten KUMplusKOM-Kompetenzen. Im Anschluss (3) die Entwicklung eines KUMplusKOM-Modells aus dem Abgleich der Literaturrechercheergebnisse und den Ableitungen der relevanten Curriculumsinhalte und abschließend (4) die Konzeption der praktischen OSCE-Prüfung mit Ihren verschiedenen Teststationen (Scherer, Ehrhardt & Lühmann, 2014)⁴. Im Folgenden werden diese Vorarbeiten genauer beschrieben.

4.2.1. Literaturrecherche

Initial wurde eine systematische Literaturrecherche sowohl medizinischer als auch pädagogischer bibliographischer Datenbanken durchgeführt. Diese bezog sich international sowohl auf englisch- als auch deutschsprachige, möglichst aktuelle, Primärarbeiten, die Evidenzen, oder die Validierung von Kompetenzmodellen zur ärztlichen Reife behandeln, beziehungsweise sich der Validierung passender Outcome-Instrumente widmen. Im Anschluss wurden die Abstracts und Titel aller 50 aufgefundenen relevanten Arbeiten durch zwei unabhängige Expert:innen hinsichtlich ihrer Relevanz gesichtet. Die Sichtungen ergaben, dass in der internationalen Literatur zum damaligen Zeitpunkt kein Konstrukt, welches den Kriterien der „allgemeinen Arztreife“ entspricht, in Gänze operationalisiert oder in Prüfungsformate umgesetzt worden war (Redaktionsschluss für die Recherchen: Februar 2015). Es fanden sich einerseits Operationalisierungen von Einzelfertigkeiten beziehungsweise konzeptionelle Papiere, Curricula und Positionspapiere zu integrativen Ausbildungsmodellen. Überlegungen aus diesen Materialien flossen punktuell in die Konstruktion des KUMplusKOM-Kompetenzmodells ein (s. 4.2.2).

4.2.2. Extraktion der KUMplusKOM-Kompetenzen

Für die Extraktion der im iMED-Studiengang vermittelten KUMplusKOM-Kompetenzen mussten diese zunächst operationalisiert, beziehungsweise inhaltsanalytisch aufgearbeitet werden. Dazu wurden in einem induktiven Prozess unter Zuhilfenahme qualitativer Forschungsmethoden diese systematisch aus der zentralen Lehrdatenbank des Studiengangs extrahiert. Die digitale Lehrdatenbank bildet die Lernziele aller Unterrichtsveranstaltungen und Prüfungen des Curriculums standardisiert ab und wurde bei der Entwicklung des iMED-Curriculums von den an der medizinischen Ausbildung beteiligten Fachrichtungen gemeinsam erstellt. Dort sind alle KUMplusKOM-Lernziele entsprechend vom Lehrpersonal besonders gekennzeichnet.

⁴ Förderfonds Lehre Antrag, nicht öffentlich zugänglich.

Zwei Beurteiler:innen identifizierten und kategorisierten diese Eintragungen unabhängig voneinander qualitativ nach Ähnlichkeitsgesichtspunkten. Die so entstandene Kategorienstruktur wurde anschließend im gemeinsamen Diskurs konsentiert.

Bei der Bestandsaufnahme wurde ein besonderes Augenmerk auf die Exhaustivität der Lehrinhaltsdaten gelegt und Sorge für die Disjunktheit der einzelnen identifizierten Kompetenzen getragen. Abschließend wurden die gefundenen Kompetenzen hierarchisch gegliedert, so dass eine Priorisierung der für die KUMplusKOM-Kompetenzen relevanten Teilkompetenzen möglich wurde. Zusätzlich zu den so entstandenen Hierarchieebenen wurden die Beziehungen verschiedener Teilelemente sichtbar gemacht. Dadurch entstand ein vorläufiges KUMplusKOM-Kompetenzmodell, das die vermittelten Inhalte des iMED-Studiengangs erschöpfend abbildet und als Grundlage für die Konzeption des Vergleichs-OSCEs dienen kann.

4.2.3. Konzeption der Vergleichs-OSCE Teststationen

Das vorläufige Kompetenzmodell wurde in einem nächsten Schritt in einer Fokusgruppe, die sich aus Expert:innen für das medizinische Curriculum, Testkonzeption, OSCE-Prüfungen und medizinische Kommunikation zusammensetzte, validiert. Dabei entstand in gegenseitiger Ergänzung der identifizierten Kompetenzen ein konsentiertes endgültiges Modell, das das ideale Kompetenzspektrum der Medizinabsolventen möglichst umfassend und aus der Theorie sowie aus der Praxis informiert, repräsentieren soll.

Basierend auf diesem Abgleich wurden die 8 thematisch überwiegend disjunkten Teststationen des Vergleichs-OSCEs (siehe Tabelle 1. Vergleichs-OSCE Teststationen*) in enger interdisziplinärer Zusammenarbeit mit den Fachvertreter:innen konzipiert. Die Teststationen sollten dabei einerseits repräsentativ und erschöpfend für die (wenigen) in der internationalen Evidenz genannten Kompetenzen zur Arztreife sein, andererseits die KUMplusKOM-Kompetenzen und deren Integration und Ausprägung abbilden. Eine Besonderheit der OSCE-Konzeption ist dabei die Möglichkeit der Vergabe einer Gesamtbeurteilung der Testteilnehmer:innen über alle Stationen hinweg, im Gegensatz zu den regulären OSCE-Prüfungen in iMED. So kann neben den einzelnen Kompetenzen auch die Stimmigkeit einer übergreifenden Handlungskette eingeschätzt werden.

4.2.4. Die Vergleichs-OSCE Teststationen

Wie eingangs beschrieben, ergaben sich 8 praktische Teststationen aus dem konsentierten Kompetenzmodell. Zu jedem Themenfeld wurden von erfahrenen Vertreter:innen des jeweiligen Fachgebietes in Zusammenarbeit mit Expert:innen für Prüfungskonzeption und –Organisation typische Behandlungsanlässe identifiziert und zu Prüfungsinhalten aufgearbeitet. Zusammen mit den Prüfungsfragen wurden standardisierte Beurteilungsbögen entwickelt, die

nach einer Pilotierung im Vorfeld an freiwilligen Medizinstudent:innen mehrfach angepasst wurden, bis ein optimales Ergebnis erreicht wurde. Leider können die konkreten Beurteilungsbögen in dieser Arbeit aus Datenschutzgründen nicht en détail vorgestellt oder angehängt werden, dennoch wird sich um eine umfassende und verständliche inhaltliche Darstellung im Folgenden bemüht. Die Aufgabenstellungen für die Studierenden finden sich im Anhang. Die Prüfung bediente sich verschiedenster Formate, so wurden unter anderem Filmmaterialien für Station 1 bereitgestellt, die einen rauchenden Risikopatienten für Blutdruckprobleme zeigten, aber auch psychosomatische Patientenrollen für Simulationspatient:innen⁵ entwickelt und nach einer Pilotierung und Anpassung geschult.

Abbildung 1. Reanimationspuppe



Quelle: <https://www.uke.de/studium-lehre/learnumgebung/meditreff/index.html>

Auf der dazugehörigen Station trafen die Testungsteilnehmer:innen dann auf Patient:innen, die über Fußschmerzen ohne somatisches Korrelat klagten (Station 2).

Inhaltlich wurden entsprechend dem KUMplusKOM-Konzept sowohl rein praktische Fertigkeiten abgeprüft, wie zum Beispiel Blutabnahme (Station 3), als auch kommunikative Fähigkeiten gemessen (zum Beispiel die psychosomatische Gesprächssituation in Station 2) (s.

Tabelle 1. Vergleichs-OSCE Teststationen). Inhaltlich sind diese Fertigkeiten durch ihre gewollte Integration in den meisten anderen Stationen nicht komplett voneinander zu trennen, da beispielsweise sowohl kommunikative Fertigkeiten für eine abdominale Untersuchung eines Probanden, der über Bauchschmerzen klagt (Station 6), vonnöten sind, als auch die praktische Kompetenz der eigentlich Tastuntersuchung. Erst die Integration beider Fertigkeiten bedingt eine optimale Prüfungserfüllung.

⁵ Simulationspatient:innen sind speziell geschulte (Laien-)darsteller:innen, die anhand vorgefertigter Rollenskripte Patient:innen mit verschiedensten Erkrankungen und Anliegen in medizinischen Lehr- und Prüfungskontexten simulieren (Cleland et al, 2006).

Tabelle 1. Vergleichs-OSCE Teststationen*

Nr.	Thema	Fach	Inhalt
1	Blutdruckberatung	Allgemeinmedizin	Patientenfilm, Prüfergespräch zur optimalen Weiterbehandlung/ Risiken
2	Somatisierungsstörung	Psychosomatik	Anamnesegespräch SimPat mit Fußschmerz
3	Venöse Blutentnahme	Allgemeinmedizin	am Armmodell
4	Schwindel	Neurologie	Neurologische Anamnese bei Schwindel, mit Untersuchungsproband, Frenzelbrille
5	Reanimation	Notfallmedizin	Notfallsimulation an der Reanimationspuppe
6	Untersuchung Abdomen	Allgemeinmedizin	Körperliche Untersuchung mit Untersuchungsproband
7	Totenschein	Allgemeinmedizin	Fallvignette Todesfall im Seniorenheim, inkl. Bild, Todesbescheinigung und Diagnose ausstellen,
8	Evidenzbasierte Medizin (EBM)	Allgemeinmedizin	Fallvignette Raucher, digitale Recherche Leitlinien-gerechter Behandlung und Fragen zum Fall

*eigene Darstellung

Abbildung 2. Armmodell zum Üben von InjektionenQuelle: <https://www.uke.de/studium-lehre/lernumgebung/meditreff/index.html>

Jede Teststation wurde von einer Lehrperson überwacht, die anhand des eigens entwickelten standardisierten Beurteilungsbogens die Leistungen der Testungsteilnehmer:innen in Echtzeit schriftlich bewertete.

Zusätzlich lagen für jede Station gesonderte standardisierte Anweisungen für die Prüflinge als auch die Beurteilenden vor. Den Probandinnen wurden alle benötigten Hilfsmittel zur Verfügung gestellt. So wurde zum Beispiel auf der Reanimationsstation an einer entsprechenden Übungspuppe simuliert (Station 5, siehe Abbildung 1. Reanimationspuppe), oder bei der neu-

rologischen Untersuchung nicht nur ein Proband mit vorgeblichen (eingeübten) Schwindelsymptomen befragt, sondern auch die zur optimalen Untersuchung benötigte Frenzelbrille⁶ angeboten (Station 4). Für die Ausstellung eines Totenscheines (siehe Abbildung 3) wurde den Teilnehmer:innen in Station 7 ein Foto der Einblutungen am Arm einer verstorbenen Patientin (Abbildung 4) zusammen mit einer Fallvignette vorgelegt, aus der eine Diagnose abgeleitet werden sollte.

Abbildung 3. Auszug aus dem auszufüllenden Totenschein in Station 7

Todesursache / Klinischer Befund <small>(Nicht Endzustände wie Atemstillstand, Herz-Kreislaufversagen, Kachexie, Verbluten)</small>		Zeildauer zwischen Beginn der Krankheit und Tod	Erläuterungen
I.a) Unmittelbare Todesursache (z. B. Pneumonie)			Unmittelbar zum Tode führende Krankheit
b) Dies ist eine Folge von (z. B. Lungenembolie)			Vorangegangene Ursachen Krankheiten, die die unmittelbare Todesursache unmittelbar herbeigeführt haben, mit der ursprünglichen Ursache (Grundleiden) an letzter Stelle
c) Hierfür ursächl. Grundleiden (z. B. Thrombose)			Andere wesentliche Krankheiten, die zum Tode beigetragen haben, ohne mit der unmittelbaren Todesursache oder dem Grundleiden im Zusammenhang zu stehen
II. Mit zum Tode führende Krankheiten, die nicht mit dem Grundleiden in Zusammenhang stehen (z. B. Diabetes)			

Epikrise
Nähere Angaben zum Todesfall, beim nichtnatürlichen Tod zum Hergang und zur Ursache der Schädigung.

Obduktion wird angestrebt nein ja Herzschrittmacher ja nein

Todesart
Gibt es Anhaltspunkte für ein nichtnatürliches Geschehen im Zusammenhang mit dem Todeseintritt? (Selbsttötung, Unglücksfall oder Tod durch äußere Einwirkung, bei der das Verhalten eines Dritten eine Ursache gesetzt haben könnte, Spätod nach Verkehrsunfall, Lungenembolie durch unfallbedingtes Krankenlager etc.)

nein ja (welche?)

Todesart ungeklärt (weshalb?)

Quelle: Prüfungsunterlagen, nicht öffentlich zugänglich

Abbildung 4. Beispielbild zur Diagnosestellung, Station 7



Quelle: Prüfungsunterlagen, nicht öffentlich zugänglich

⁶ Die Frenzel- oder Nystagmusbrille nach Frenzel ist eine rundum geschlossene Leuchtbrille mit vergrößernden Plusgläsern von mindestens 15,00 Dioptrien, die der Untersuchung unwillkürlicher Augenbewegungen dient. Diese Augenbewegungen sind ein typisches Symptom von Störungen des Gleichgewichtssinns. Sie ist benannt nach Ihrem Erfinder, dem Göttinger Otologen Hermann Frenzel (1895–1967) (Kaufmann, 1986).

Auf der EBM-Station (Station 8) wurde schließlich ein Laptop mit Internetzugang bereitgestellt, mittels dem die Teilnehmer:innen ihre Recherche nach aktuellen Leitlinien zum Thema Tabakentwöhnung durchführen konnten (siehe Studierendenanweisung im Anhang).

Auf allen Stationen waren maximal 20 Punkte erreichbar. Für alle Stationen galt die gleiche Zeitvorgabe von 5 Minuten. Die Ergebnisse ihrer Prüfung erfuhren die Prüflinge schriftlich erst einige Wochen nach der Teilnahme, auf Anfrage wurde Ihnen eine ausführliche Beurteilung durch eine ärztliche Fachkraft zur Verfügung gestellt.

4.3. Teilnehmer:innen

Es gelang die Rekrutierung von 101 Hamburger Medizinstudent:innen beiderlei Geschlechts vor ihrem Eintritt ins Praktische Jahr (PJ). Die Stichprobe setzte sich zu gleichen Teilen aus Regelstudiengangs- und iMED-Studiengangsabsolvent:innen zusammen. Beide Teilstichproben wurden querschnittlich untersucht. Geplant war der Einschluss von jeweils 25 Personen an 4 Erhebungszeitpunkten (E1 – E4). Zu E1 und E2 wurden die Regelstudiengangsabsolvent:innen getestet und zu E3 und E4 die iMED-Absolvent:innen.

Die Rekrutierung begann im Herbst 2016, etwa 4 Wochen vor der ersten Erhebung, und endete im Sommer 2018. Mit Beginn sechs bis acht Wochen vor jedem Testzeitpunkt bis zum eigentlichen Tag der jeweiligen Testung erfolgten über Emails an das Studierendenskollektiv via Fachschaftsrat, Flugblätter und die gezielte Vorstellung des Forschungsvorhabens in Vorlesungen, sowie durch die persönliche Ansprache von Studierenden der medizinischen Fakultät am Universitätsklinikum Hamburg-Eppendorf in Lehrveranstaltungen umfangreiche Rekrutierungsbemühungen. Die Teilnahme war freiwillig. Nach anfänglichen sehr wenigen interessierten Rückmeldungen geeigneter Testpersonen wurden als Anreiz finanzielle Incentives (einmalig 50 Euro in Form von Amazon-Gutscheinen) und eine individualisiertes Feedback zu den im Testverfahren gezeigten Leistungen angeboten, das der besseren Vorbereitung auf das anstehende Staatsexamen dienen sollte. Diese Anpassung des Prozederes war erfolgreich und führte zu einer deutlichen Steigerung der Rückmeldungen.

4.4. Konkrete Durchführung des Vergleichs-OSCE und Fragebogenerhebung

Die Datenerhebung bestand aus der Prüfungsleistung an den 8 OSCE-Stationen und dem Ausfüllen eines Fragebogens. Die 4 konsekutiven Erhebungszeitpunkte zwischen Herbst 2016 und Sommer 2018 verliefen identisch: nach der Ankunft der Testteilnehmer:innen und der Vergabe einer personalisierten Startnummer zur späteren Zuordnung der Testungsergebnisse, erfuhren sie eine Einweisung durch prüfungserfahrenes Fachpersonal. Das allgemeine

Vorgehen ist analog dem etablierten OSCE-Format und war den Teilnehmer:innen daher aus ihrem Studium vertraut.

Die Durchführung entsprach einem Zirkeltrainingsablauf, jede:r Teilnehmer:in durchlief jede Stationen genau einmal, allerdings in unterschiedlichen Reihenfolgen. Es wurden parallel 8 Personen gleichzeitig getestet, pro Prüfungstag fanden 3-4 aufeinanderfolgende Schichten statt, um den organisatorischen Aufwand zu minimieren.

Die Testpersonen hatten 5 Minuten Zeit pro Prüfungsstation für die Bewältigung der Prüfungsaufgaben sowie anschließend 90 Sekunden Wechselzeit, in der die Teilnehmer:innen die Prüfungsräume verließen und sich vor die (noch geschlossene) Tür der nächsten Teststation begaben. Dort lasen sie die Aufgabenstellung für die nächste Prüfungsstation, die in Aushängen an den Türen des jeweiligen Prüfungsraumes angebracht waren. In den eigentlichen Prüfungsräumen warteten die Beurteiler:innen mit den Prüfungsunterlagen und den jeweiligen Proband:innen, Simulationspatient:innen und eventuell benötigten Zusatzmaterialien. Die Student:innen starteten gleichzeitig auf ein Signal der Prüfungsaufsicht hin und verließen die Stationen nach Ablauf der 5-minütigen Testungszeit auf ein erneutes Signal hin zeitgleich. Auch dieses Vorgehen ist das im Studiengang übliche Prüfungsformat.

Nachdem jede Person alle Teststationen durchlaufen hatte, wurde von allen unter Aufsicht einer Testleitung anschließend der Fragebogen mit ergänzender Psychometrie und Soziodemografie in einem separaten Raum ausgefüllt. Nach der Abgabe des ausgefüllten Fragebogens erhielten die Testteilnehmer:innen ihren Amazon-Gutschein und konnten die Räumlichkeiten verlassen. Die gesamte Dauer belief sich auf etwa 1,5 Stunden von Beginn bis Ende der Testung.

Durchgeführt wurde der Vergleichs-OSCE zu allen Testzeitpunkten in den Räumlichkeiten des MediTreFF (Medizinisches Trainingszentrum eigener Fähigkeiten und Fertigkeiten), am UKE, der sich durch kleine, direkt nebeneinanderliegende Räume mit geeigneter Ausstattung besonders für dieses Prüfungsformat anbietet.

4.5. Benotung

Auf jeder Prüfungsstation konnte ein sehr guter Teilnehmer bei bester Leistung maximal 20 Punkte erreichen (entsprechend waren 160 Punkte die mögliche Höchstpunktzahl über alle Stationen hinweg. Etwa 15 Punkte machen dabei die Differenz von einer Schulnote zur nächsten aus.). Die Bewertung in Schulnoten wurde für die Einzelstationen und die Gesamtleistung analog zur Approbationsordnung (Güntert, 2003) angelegt (vergleiche Tabelle 2): 100 – 90% der Punkte mit „sehr gut“, also einer „1“ benotet; 89 – 80% mit „gut“ oder „2“; 79-70% mit „befriedigend“ oder einer „3“, 69 – 60% mit „ausreichend“, also einer „4“, alles darunter galt als

„nicht bestanden“. Die Beurteilungen wurden durch überwiegend die gleichen prüfungserfahrenen Rater:innen zu allen Testzeitpunkten durchgeführt, um personenabhängige Schwankungen im Beurteilungsverhalten zu minimieren. Alle Rater:innen wurden vorher geschult und nutzten standardisierte Beurteilungsbögen.

Tabelle 2. *Benotung des Vergleichs-OSCE's, angelehnt an Approbationsordnung*

Schulnote	Wertung	Punktränge Gesamtergebnis	Punkt- range Ein- zelstatio- nen	Prozentränge
1	sehr gut	144-160	18-20	90-100 %
2	gut	128-143	16-17	80-89 %
3	befriedigend	112-127	14-15	70-79 %
4	ausreichend	96-111	12-13	60-69 %
5	ungenügend	80-95	10-11	50-59 %
6	schlecht	0-79	0-9	0-49 %

Eigene Darstellung

4.6. Psychometrische Instrumente

Im Anschluss an die Teilnahme an der OSCE-Prüfung wurde den Studierenden ein 8-seitiger Selbstauskunftsfragebogen vorgelegt, der normierte und validierte Skalen zur Ausprägung psychologischer Persönlichkeitskriterien, globaler Lebenszufriedenheit, sozialer Unterstützung, Angst und Depressivität, angestrebter Facharztweiterbildungsrichtung und soziodemografischen Aspekten enthielt. Das Ausfüllen des Papierfragebogens dauerte etwa 20 Minuten. Eine durchgängig anwesende Testleiterin überwachte den Ausfüllprozess, beantwortete vor Ort Fragen und wies kurz in den Fragebogen ein, und notierte die Startnummer des jeweiligen Studierenden auf den dazugehörigen Fragebögen, um das spätere Zusammenführen mit den Leistungsdaten aus dem OSCE zu ermöglichen. Die Teilnehmer wurden nicht über die enthaltenen Instrumente und Hypothesen informiert, um das Antwortverhalten nicht zu beeinflussen, sondern erhielten nur die neutrale Auskunft, dass es bei den Testungsinhalten um den „Zusammenhang zwischen Persönlichkeit und Arztreife“ ginge, ohne diese näher auszuführen (s. Teilnehmer:inneninformation).

Aufgrund eines internen Absprachefehlers wurden nur das NEO-Fünf-Faktoren-Inventar und der soziodemografische Abschnitt des Fragebogens bei E1 eingesetzt, die anderen psychometrischen Instrumente wurden ab E2 ergänzt und weisen daher kleinere Stichprobengrößen auf.

Tabelle 3. Übersicht der verwendeten psychometrischen Instrumente

Kürzel	Name	Autoren	Skalen und Itemanzahl
NEO-FFI-30	Neo-Fünf-Faktoren-Inventar, 30-Item-Kurzversion	Körner et al. (2007)	Neurotizismus, Extraversion, Offenheit, Gewissenhaftigkeit und Verträglichkeit, 30 Items
SWLS	Satisfaction with Life Scale	Schuma- cher et al. (2003)	Lebenszufriedenheit, 5 Items
F-SoZu K14	Fragebogen zur Erfassung sozialer Unterstützung, Kurzform	Fydrich et al. (2007)	Soziale Unterstützung, 14 Items
HADS-D	Hospital Anxiety and De- pression Scale	Herrmann- Lingen et al. (2011)	Angst und Depression, 14 Items

Eigene Darstellung

4.6.1. NEO-FFI 30 (NEO-Fünf-Faktoren-Inventar)

„NEO“ steht für Neurotizismus, Extraversion und Offenheit. Die ursprüngliche Intention des Verfahrens war der Einsatz als Breitbanddiagnostikum. Die deutsche Übersetzung von Borkenau & Ostendorf (1993) des englischsprachigen Ursprungsinstruments NEO-PI-R wurde von Körner et al. faktorenanalytisch überarbeitet und kondensiert und zu der vorliegenden Kurzform aufbereitet. Das Ausfüllen dauert etwa 10 Minuten.

Der Fragebogen erfasst interindividuelle Ausprägungen der fünf als disjunkt angenommenen Persönlichkeitseigenschaften Neurotizismus, Extraversion, Offenheit, Gewissenhaftigkeit und Verträglichkeit in 5 Skalen mit jeweils 6 Items (Borkenau & Ostendorf, 1993). Er überzeugt durch eine hohe Spezifität und Sensitivität, und eine ausgezeichnete Reliabilität, sowie faktorielle und Konstruktvalidität.

Beispielitems für die Skala Neurotizismus sind *„Ich fühle mich oft angespannt und nervös.“* oder *„Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.“*

Die Skala Extraversion enthält unter anderem *„Ich habe gern viele Leute um mich herum.“* und *„Ich bin gerne im Zentrum des Geschehens“*. Verträglichkeit wird durch Aussagen wie *„Ich bekomme häufiger Streit mit meiner Familie und meinen Kollegen.“* und *„Ich versuche stets rücksichtsvoll und sensibel zu handeln.“* abgefragt. *„Ich halte meine Sachen ordentlich und sauber.“* sowie *„Wenn ich eine Verpflichtung eingehe, so kann man sich auf mich bestimmt verlassen.“* sind Beispielitems der Gewissenhaftigkeitsskala.

Offenheit beinhaltet Aussagen wie „*Poesie beeindruckt mich wenig oder gar nicht.*“ und „*Ich habe oft Spaß daran, mit Theorien oder abstrakten Ideen zu spielen.*“ (siehe Anhang: Fragebogen, 2. Fassung).

Jedes Item kann auf einer 5-stufigen Likert-Skala von „starker Ablehnung“ über „Ablehnung“, „neutral“, „Zustimmung“ bis „starker Zustimmung“ beantwortet werden. Den Ausprägungen sind die Werte 0 – 4 zugeordnet, einzelne Items sind negativ formuliert und müssen umgepolt werden. Die Auswertung erfolgt durch eine Summenbildung der Itemrohwerte pro Skala (nach Umpolung der invertiert formulierten Items), die dann durch die Anzahl der Items geteilt wird, sodass für jedes Individuum für jede Skala ein Mittelwert vorliegt. Der resultierende Mittelwert kann von mindestens 0 bis maximal 4 Punkten reichen, je höher die Zahl, desto höher die gemessene Ausprägung des Attributs.

4.6.2. Satisfaction with Life Scale (SWLS)

Die Lebenszufriedenheit der Teilnehmer:innen erhoben wir ab dem zweiten Erhebungszeitpunkt mit der Satisfaction with Life Scale (SWLS), einem ökonomischen eindimensionalen Selbstbeurteilungsinstrument zur Erfassung der allgemeinen Lebenszufriedenheit. Diese liegt in der deutschen Übersetzung von Schumacher et al. (2003) vor und basiert auf der englischsprachigen Originalversion von Diener et al. (1985).

Die Autoren entwickelten den standardisierten Fragebogen thematisch anknüpfend an die Theorie des subjektiven Wohlbefindens (*Subjective Well-Being*, SWB), die die Zusammensetzung dieses Zustandes aus zwei Hauptkomponenten postuliert: einerseits der affektiv-emotionalen (die sich wiederum in „positiven-“ und „negativen Affekt“ sowie das längerfristig anhaltende „Glück“ aufteilt) und andererseits einer kognitiv-evaluierenden Komponente (Diener 1984, Pavot 2013). Die SWLS bildet die kognitiv-evaluative Komponente des SWB und damit die globale Lebenszufriedenheit in der Selbstbeurteilung ab.

Die Validität der SWLS konnte in unterschiedlichen Erhebungen faktorenanalytisch bestätigt werden. In der studentischen Originalstichprobe luden alle 5 Items auf dem gleichen Faktor zwischen .61 und .84. Die erklärte Gesamtvarianz betrug 66% (Diener et al., 1985). Spätere Erhebungen ermittelten noch höhere Ladungen zwischen .68 und .93 und konnten durch den Generalfaktor 65% beziehungsweise sogar 74% der Gesamtvarianz erklären (vergleiche Pavot et al., 1991). Die Untersuchung der diskriminanten Validität durch Lucas et al. (1996) ermöglicht eine deutliche Abgrenzung sowohl vom emotionalen Wohlbefinden als auch vom Optimismus und dem. Auch die konvergente Validität, gemessen durch die Korrelation zum einschlägigen Life Satisfaction Index A (LSI-A; Neugarten et al., 1961), war mit $r = .84$ befriedigend.

Als Datenbasis für die Entwicklung des SWLS dienten zwei studentische Stichproben (N = 336) und eine im Durchschnitt 75-jährige Gruppe von Personen (N = 53). Nach wiederholter Evaluation wird die Skala unverändert eingesetzt (Pavot & Diener, 1993).

Durch seine Standardisierung wird von einer hohen Objektivität des Instruments ausgegangen. Diener et al (1985) berichten für die englische Originalfassung zudem eine hohe interne Konsistenz. Bei einem Erhebungsabstand von vier Jahren beträgt die Retest-Reliabilität .58, was von den Autoren als empirische Bestätigung der theoretisch postulierten zeitlichen Stabilität des Konstruktes der globalen Lebenszufriedenheit interpretiert wurde.

Das Instrument umfasst 5 Aussagen. Die Ausfüllenden müssen das Ausmaß ihrer Zustimmung auf einer 7-stufigen Likert-Skala durch Ankreuzen angeben (7= *stimme völlig zu*, 6= *stimme zu*, 5= *stimme eher zu*, 4= *weder/noch*, 3= *stimme eher nicht zu*, 2= *stimme nicht zu*, 1= *stimme überhaupt nicht zu*). Das Ausfüllen dauert circa 1-2 Minuten. Die Items lauten:

(1) *In den meisten Bereichen entspricht mein Leben meinen Idealvorstellungen.*

(2) *Meine Lebensbedingungen sind ausgezeichnet.*

(3) *Ich bin mit meinem Leben zufrieden.*

(4) *Bisher habe ich die wesentlichen Dinge erreicht, die ich mir für mein Leben wünsche.*

(5) *Wenn ich mein Leben noch einmal leben könnte, würde ich kaum etwas ändern.*

Die Auswertung erfolgt über eine Addition der Itemrohwerte. Der berechnete Summenscore drückt mit steigender Größe auch steigende Lebenszufriedenheit aus, wobei die Autoren die ermittelten Ergebnisse zur leichteren Einordnung und Interpretation in 7 Ausprägungsstufen von „extrem unzufrieden“ bis „extrem zufrieden“ kategorisieren.

4.6.3. F-SoZu K14, Fragebogen zur Erfassung sozialer Unterstützung

Ab dem zweiten Erhebungszeitpunkt ermittelten wir die soziale Unterstützung unserer Teilnehmer:innen. Dazu nutzen wir den standardisierten Selbstauskunftsfragebogen F-SoZu-K14 von Fydrich et al. (2007), mittels dem sich in 14 Aussagen, die auf einer 5-fach abgestuften Likertskala (von „*trifft nicht zu*“ bis „*trifft genau zu*“) bewertet werden, die subjektiv wahrgenommene Güte des sozialen Netzes der Ausfüllenden beurteilen lässt. Aus Zeitökonomiegründen verwendeten wir auch bei diesem Instrument, wie oben beschrieben, die Kurzform mit 14 Items als Papier-Bleistift-Verfahren. Das Ausfüllen kostete die Teilnehmer etwa 3 Minuten.

Eine hohe soziale Unterstützung wird schon seit längerem in der Forschungsliteratur als wichtige protektive Ressource diskutiert, deren Ausprägung nicht nur psychische, sondern auch körperliche Erkrankungssymptome beeinflussen kann. Personen, die über ein gutes soziales Netz verfügen, können resilienter auf Belastungen reagieren, egal, ob es sich dabei um kleine

Alltagssorgen („*daily hassles*“) oder größere biografische Einschnitte („*live-events*“) handelt (Brown, 1987). Von besonderer Bedeutung ist dabei nicht nur das tatsächliche Maß an Unterstützung durch das soziale Umfeld, sondern insbesondere dessen subjektive Wahrnehmung (Lakey & Cassidy, 1991). Daher wird mit dem Instrument ein Augenmerk auf die Erfassung ebendieser subjektiven Komponente gelegt.

Soziale Unterstützung in der Operationalisierung des F-SoZu untergliedert sich in drei Subskalen: (1) Emotionale Unterstützung, (2) Praktische Unterstützung, und (3) Soziale Integration, wobei die verwendete Kurzform nur einen Gesamtwert abbilden kann. In der Kurzform des Instruments liegen keine invertierten Items vor, zudem wurden gegenüber der Langform die besonders trennscharfen Items ausgewählt, um die ursprünglich rechtsgipflige Verteilungskurve zu nivellieren. Zusätzlich sind die Aussagen schwerer formuliert. Damit soll das Instrument soziale Unterstützung reliabel erfassen, ohne verschiedene Konstruktbereiche explizit zu differenzieren. Nach der neuen Normierung des Fragebogens fand sich allerdings nach wie vor eine linksschiefe Verteilung, weshalb die Differenzierungsfähigkeit des Instruments laut den Autoren besonders im Bereich niedriger sozialer Unterstützung gegeben ist.

Die Objektivität des Instrumentes ist durch die standardisierte Auswertung sicher gestellt. Die Validität konnte in einer Vielzahl von Studien bestätigt werden (zum Beispiel Fydrich et al. 1987, 1991). In der deutschen Normierung der Kurzform (Fydrich et al., 2009) fanden sich exzellente psychometrische Eigenschaften: die Reliabilität (Cronbachs α) betrug .94; die Retestreliaibilität nach einer Woche war mit .96 sehr zufriedenstellend. Die untersuchten 2507 bevölkerungsrepräsentativen Teilnehmer:innen im Alter zwischen 14 - 92 ($M = 48.8$, $SD = 17.9$, 55 % ♀) Jahren erreichten im Mittel 3.97 Punkte ($SD = 0.68$), was

Die folgenden ausgewählten Beispielitems illustrieren das Konstrukt sehr eindrücklich: „*Ich finde ohne weiteres jemanden, der sich um meine Wohnung (zum Beispiel Blumen, Haustiere, Post) kümmert, wenn ich mal nicht da bin*“, „*Es gibt eine Gruppe von Menschen (Freundeskreis, Clique), zu der ich gehöre und mit denen ich mich häufig treffe*. oder „*Wenn ich mal sehr bedrückt bin, weiß ich, zu wem ich damit ohne weiteres gehen kann*.“.

Die Auswertung der Ergebnisse erfolgt durch eine Addition der Itemrohwerte und Division durch die Anzahl der Items, dann wird mittels Normtabelle ein Prozentrang (PR) zugeordnet. 12 der 14 Items müssen beantwortet sein. Eine nach Skalen getrennte Auswertung wird für die Kurzform explizit nicht empfohlen. Auch hier indizieren höhere Werte stärkere soziale Unterstützung. Es besteht mit dem Testverfahren nur eine geringe Differenzierungsfähigkeit im Bereich hoher sozialer Unterstützung, dafür differenziert das Verfahren gut bei geringer sozialer Unterstützung (S. 36, Manual).

4.6.4. Hospital Anxiety and Depression Scale (HADS-D)

Die *Hospital Anxiety and Depression Scale* ist ein kurzer schriftlicher Selbstauskunftsfragebogen, mit dem Angst und Depression erfasst werden können. Auch dieses Verfahren war erst ab E2 Teil des Fragebogens. Ursprünglich von Zigmond und Snaith in 1983 für die Testung somatischer Patienten in klinischen Kontexten entwickelt, findet dieses ökonomische Testverfahren mittlerweile auch in vielen Forschungskontexten jenseits der Untersuchung von Patientengruppen Anerkennung, zum Beispiel auch in der Erfassung der Belastung medizinischen Fachpersonals (Paiva et al., 2018). Dabei zeichnet sich das etablierte und vielfach eingesetzte Instrument insbesondere durch seinen pragmatischen Ansatz, der dem Ressourcenmangel in klinischen Kontexten Rechnung trägt, sowie seine gute Akzeptanz bei (somatischen) Patienten aus. Andere tradierte Messverfahren dieser Skalen mit psychiatrischem oder medizinpsychologischem Hintergrund setzen oft psychopathologische Schwerpunkte oder sind sehr speziell oder umfangreich angelegt. Damit schrecken diese die Befragten ab und verändern ihr Antwortverhalten oder sind als wenig zeitökonomisch, beziehungsweise ungeeignet für die Aufdeckung leichter Beeinträchtigungen einzuschätzen (Herrmann et al., 2011). Der HADS-D schließt die diagnostische Lücke eines kurzen, ohne Vorwissen einsetzbaren und interpretierbaren Instrumentes. Herrmann-Lingen et al. stellte die hier verwandte deutsche Übersetzung in 2011 vor.

Der HADS-D enthält zwei Subskalen: Angst (HADS-D/A) und Depression (HADS-D/D). Erstere enthält 6 Items, die psychische Manifestationen von Angstneurosen generalisiert abbilden sollen, sowie ein zusätzliches zur Erfassung von Panikattacken. Die 7 Items der zweiten Skala beleuchten die mögliche Ausprägung von Leitsymptomen einer milden depressiven Episode mit reduziertem hedonischem Tonus.

Wie auch bei den anderen psychometrischen Skalen, die in dieser Arbeit Verwendung fanden, sind die Durchführungs- und Auswertungsobjektivität des Verfahrens durch die schriftliche Einweisung und die standardisierte Auswertung sichergestellt.

Bearbeitung erfolgte im Bleistift-Papier-Verfahren mit der Ankreuzung der Auftretenshäufigkeit oder Zustimmung auf einer 4-stufigen ordinalen Likertskala. Die Teilnehmer:innen werden gebeten, die Fragen und Aussagen in Hinblick auf die vergangene Woche zu beantworten und zu beurteilen. Das Instrument umfasst 14 Items, deren Anker sich je nach Inhalt unterscheiden. Dabei werden 7 Items für jede Subskala sowohl in alternierender Abfolge präsentiert, als auch jeweilig die Schlüsselrichtung wechselnd. Beispielitems für die Skala Angst sind: „*Ich habe manchmal ein ängstliches Gefühl in der Magengegend*“ und „*Ich fühle mich angespannt oder überreizt*“. Die Skala Depression enthält Aussagen wie „*Ich fühle mich in meinen Aktivitäten gebremst*“ und „*Ich habe das Interesse an meiner äußeren Erscheinung verloren*“.

Die Auswertung geschieht durch Addition der (wo nötig umgepolten) Items jeder Skala gesondert. Eine Interpretation der Werte wird durch eine Kategorisierung in drei Klassen (≤ 7 auf

jeder Subskala = *unauffällig*, 8 - 10 *grenzwertig* und ≥ 11 *auffällig*) ermöglicht, zusätzlich empfehlen die Autoren eine weitere Differenzierung in „*schwere*“ (Wertebereich 11 - 14) und „*sehr schwere*“ (Wertebereich 15 - 21) Symptome. Darüber hinaus stehen Perzentilwerte von verschiedenen Vergleichskollektiven zum Abgleich zur Verfügung. Die Ergebnisse sind laut den Autoren dabei nicht als diagnoseweisend, sondern vielmehr als Orientierungshilfen zu verstehen.

4.7. Soziodemografie und angestrebte Facharztrichtung

Basisdokumentarische Daten zur Kontrolle relevanter Einflussgrößen auf die Vergleichs-OSCE-Performanz jenseits der Vorbereitung durch die absolvierte Lehre wurden basierend auf den Empfehlungen zu demografischen Standards des statistischen Bundesamtes (2010) erhoben. Abgefragt wurden Geschlecht, Alter, Migrationshintergrund, Bildungsabschlüsse und Familienstand mit 11 Items. Darauf folgte die Erhebung der angestrebten Facharztrichtung: „*Wenn Sie sich jetzt entscheiden müssten, welche Facharztweiterbildung würden Sie anstreben?*“. Vor der Auszählung der Einzelnennungen kategorisieren wir diese offenen Antworten formal und inhaltlich am Vorbild der KarMed-Studie, einer einschlägigen, umfangreichen längsschnittlichen Erhebung mit dem Schwerpunkt der Erhebung des Weiterbildungserfolgs deutscher angehender Ärzt:innen (siehe van den Bussche et al., 2012). Der Fragebogen schloss mit einer offenen Frage zu weiteren Anmerkungen.

4.8. Eingabe und Auswertung

Die Beurteilungsbögen des Vergleichs-OSCEs wurden nach Abschluss jeder Erhebung mittels der vergebenen Startnummern mit den von den Teilnehmer:innen ausgefüllten psychometrisch-soziodemografischen Fragebögen zusammengeführt. Im Anschluss wurde durch medizinisches Fachpersonal die OSCE-Note anhand der schriftlichen Beurteilungen und vergebenen Punkte ermittelt und dokumentiert. Das gesamte Testungsmaterial lag in Papierform vor und wurde händisch zur Auswertung in die Statistiksoftware IBM SPSS Statistics (zunächst Version 26, später wurde Version 28 für die Analysen genutzt) eingegeben, mittels der im Anschluss auch die statistische Auswertung durchgeführt wurde. Die Daten aller 4 Erhebungszeitpunkte wurden dafür nach Abschluss der Datenerhebung zusammengeführt, auf Plausibilität kontrolliert und bereinigt, bevor die Ergebnisse der psychometrischen Skalen anhand ihrer Auswertungsvorgaben errechnet wurden.

4.9. Analyseplan und tatsächliches statistisches Vorgehen

Ursprünglich geplant war die Rekrutierung von insgesamt 120 Teilnehmer:innen, um statistisch durch potentiell zu niedrige Gruppengrößen nicht eingeschränkt zu sein. Die statistische Auswertung sollte dabei mit Schwerpunkt auf Korrelationsrechnungen, t-Tests und Varianzanalysen zwischen OSCE-Ergebnissen und Fragebogen geschehen, bei komplexeren Fragestellungen aber auch weitere Berechnungen wie Faktorenanalysen und Regressionsmodelle einschließen.

Teilnehmer:innen, die an einer anderen Universität den Großteil ihres Studiums absolviert hatten (ein Umstand, der bei der Rekrutierung nicht absehbar war), wurden bei den Berechnungen, deren Inhalt nicht mit dem Ausbildungsweg zusammenhing, eingeschlossen, bei den anderen ausgeschlossen.

Je nach Resultat des Levene-Tests auf Varianzgleichheit wurde entweder der T-Test für gleiche Varianzen oder für nicht gleiche Varianzen gewählt.

4.9.1. Kategorisierung der KUMplusKOM-Komponenten

Vor dem Hintergrund der Analyse der KUMplusKOM Komponente des iMED-Curriculums wurde eine Kategorisierung der Einzelitems nach rein kommunikativen oder klinisch-praktischen Fertigkeiten vorgenommen und dann mittels t-Tests auf Mittelwertsunterschiede zwischen den Studiengangabsolvent:innen die beiden operationalisierten Konstrukte „kommunikative Fertigkeiten“ und „klinisch-praktische Fertigkeiten“ überprüft. Dabei wurde unter der Prämisse agiert, dass ausschließlich auf der Station 2 der Psychosomatischen Anamnese bei Fußschmerz rein kommunikative Items abgefragt wurden, während die anderen Stationen die „rein“ klinischen Fertigkeiten abbilden.

4.9.2. Umgang mit geringen und ungleichen Gruppengrößen

Die Untersuchung demografischer Einflussgrößen auf die Persönlichkeitsausprägungen der Gesamt- oder Teilstichprobe war aufgrund der Homogenität der Stichprobe, aber vor allem der teilweise sehr geringen Gruppengrößen statistisch nicht sinnvoll. Unter diesen Umständen ist es nicht empfehlenswert, wie geplant mit dem prädiktiven Wert des Instruments zu arbeiten, da sich die Studiengangsteilnehmer:innen bei den meisten Skalen erst in der zweiten Dezimalstelle überhaupt im Mittelwert unterscheiden. Eine logistische Regression kann aufgrund der Beschaffenheit der Daten (also sehr ähnliche Ausprägungen pro Score) keinen validen p-Wert berechnen (eine bekannte Schwäche der WALD-Statistik). Einzig auf Geschlechtsunterschiede hin war eine nähere Betrachtung praktikabel.

Eine weitere Analyse der möglichen soziodemografischen Einflussgrößen auf psychometrie oder OSCE-Ergebnis ist weiterhin ebenfalls nicht sinnvoll, denn durch die Spezifikationen der Stichprobe kann nicht mit dem prädiktiven Wert des Instruments gearbeitet werden, da sie wieder sehr ungleiche Gruppengrößen und schiefe Verteilungen der Ausprägungen aufweist: Eine logistische Regression kann darüber hinaus aufgrund der Beschaffenheit der Daten (der zu guten kompletten Trennung der Gruppen) für die Psychometrie keinen validen p-Wert berechnen.

4.10. Datenschutz und Genehmigung durch die Ethik-Kommission

Das Vorhaben wurde der Hamburger Ethikkommission zur Prüfung vorgelegt und als unbedenklich ohne separaten Antrag eingestuft und zur Durchführung freigegeben.

5. Ergebnisse

5.1. Allgemeine Beschreibung der Stichprobe

Die nachfolgend abgebildete Tabelle zeigt demografische Daten und ausgewählte Parameter zum Ausbildungsstand der Gesamtstichprobe. Die Gesamtteilnahme an allen Erhebungstagen belief sich auf 100 Personen, die Hälfte der Teilnehmer:innen waren weiblich. Das arithmetische Mittel des Alters betrug 29,1 Jahre ($SD = 3,7$, Range = 24–48, Median = 28,0).

Tabelle 4. Beschreibung der Stichprobe, $N = 100$

Variable	%/n
Geschlecht weiblich	50
Teilnehmer:innen E1	25
Teilnehmer:innen E2	22
Teilnehmer:innen E3	36
Teilnehmer:innen E4	15
Zugehörig zu RSG-Studiengang	52
Zugehörig zu iMED-Studiengang	48 (-1)
Migrationshintergrund angegeben	16
Familienstand ledig	86
Voriger Berufsabschluss vorhanden	33
Promotion (abgeschlossen oder laufend)	5
Facharztweiterbildung angestrebt	98

Eigene Darstellung

Die Gruppengrößen zwischen den Testzeitpunkten variierten leicht, mit durch das parallele Auslaufen des RSGs steigendem Anteil der iMED-Absolvent:innen: an E1 nahmen 25 Personen teil (davon 100 % RSG-Absolvent:innen), E2 belief sich auf 22 Teilnehmer:innen (100 % RSG-Absolvent:innen), E3 dann auf 36 (5,6 % RSG-Absolvent:innen), und schließlich E4 auf 15 Teilnehmer:innen (20 % RSG-Absolvent:innen). 52 % der Studienteilnehmer:innen insgesamt waren iMED-Angehörige. Es ergibt sich für die Gesamtgruppe über alle Testzeitpunkte demnach nahezu eine Gleichverteilung. Ein Teilnehmer musste nachträglich ausgeschlossen werden, da er den Großteil seines Medizinstudiums an einer anderen Universität absolviert und erst zum Studienende nach Hamburg gewechselt hatte.

Im Folgenden werden alle Teilnehmer:innen der verschiedenen Testzeitpunkte, wenn methodisch sinnvoll, zum leichteren Verständnis entweder zu einer Gesamtgruppe aller Personen

beider Studiengangsformen, oder den Angehörigen der jeweiligen Studiengangsformen zusammengefasst. Dies ermöglicht zudem durch die dadurch entstehende höhere Gruppengröße mehr und umfassendere statistische Operationen.

12 % der Befragten gaben an, eine andere Muttersprache als deutsch aufzuweisen, weitere 4 % berichteten von einer bilingualen Muttersprachlichkeit. Analog dazu beschrieben 15 % der Teilnehmer:innen eine doppelte oder andere Staatsbürgerschaft als die deutsche. Die am häufigsten außer (oder zusammen mit) Deutsch gesprochene Muttersprache war Persisch bzw. Afghanisch, Dari oder Urdu mit 7 Nennungen, gefolgt von Englisch (3 Nennungen) und Polnisch (2 Nennungen). Jeweils einmal erwähnt wurden zudem Finnisch, Indonesisch, Isländisch, Italienisch, Russisch, und Spanisch. Die offene Frage nach einer anderen Muttersprache als Deutsch wurde von 19 Personen beantwortet, was im Gegensatz zu den obengenannten 16 % bzw. Personen steht.

86 (86 %) Personen gaben als Familienstand „ledig“ an – 10 % befanden sich zum Umfragezeitpunkt in einer Ehe, 1 % in eingetragener gleichgeschlechtlicher Lebensgemeinschaft und eine weitere Person war geschieden. Mehr als zwei Drittel (67 %) der Befragten gaben an, über keinen weiteren beruflichen Ausbildungsabschluss zu verfügen. Die häufigste weitere Nennung war ein bereits vorhandener Universitätsabschluss mit 11 %.

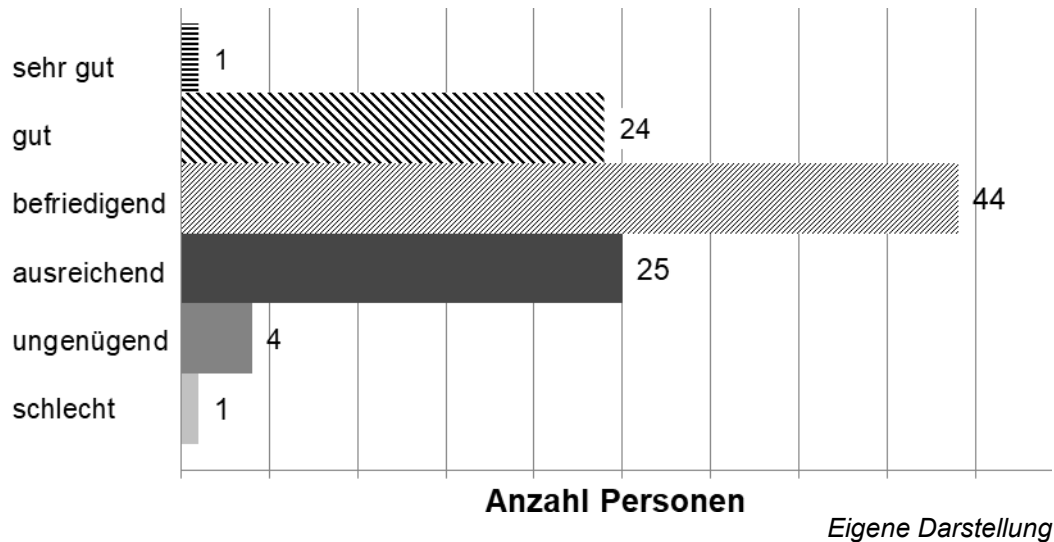
Schließlich wurde eine Facharztweiterbildung von der überwältigenden Mehrheit von 98 % der Teilnehmer:innen angestrebt, näheres wird unten in 5.4 ausgeführt.

5.2. Die Performanz im Vergleichs-OSCE

Von den möglichen 160 Gesamtpunkten bei einer perfekten Leistung in allen Stationen wurden von den Teilnehmer:innen durchschnittlich 117,6 Punkte im Vergleichs-OSCE erreicht ($SD = 13$, Range = 68-148), was einer Durchschnittsnote von „befriedigend“ entspricht (für das Benotungsverfahren vergleiche Tabelle 2, für die Übersicht über die Notenverteilung vergleiche Abbildung 5). Wenn man das der Approbationsordnung gemäße Bewertungsschema anlegt, bestanden 94 von 99 Teilnehmer:innen den OSCE in seiner Gesamtheit, 5 Personen nicht. Die Punkteverteilung der gesamten Stichprobe entsprach gemäß dem Kolmogorov-Smirnov-Test einer Normalverteilung ($p = .200$), diese fand sich aber nicht mehr bei der Kategorisierung nach Schulnoten ($p = <.001$). Anteilig fielen deutlich mehr iMED-Absolvent:innen als die des

RSG durch (6,5 % Durchfallquote versus 1,9 % bei RSG), dies ist aufgrund der geringen Fallzahl (3 versus 1 Person) nicht praktisch relevant. Eine statistische Signifikanzprüfung erübrigt sich wegen der nicht gegebenen Interpretierbarkeit.

Abbildung 5. OSCE-Schulnotenverteilung unter allen Teilnehmer:innen, N = 99



Es fand sich ein Unterschied beim Einfluss des Geschlechtes auf die Gesamtperformance: Frauen erreichten geringfügig höhere Punktzahlen als Männer (♀: $M = 120,2$, $SD = 10,2$, Range 68 - 143 versus ♂: $M = 115,5$, $SD = 15,4$, und Range 68 - 148), signifikant wurde dieser Unterschied allerdings nur bei der Blutentnahmestation (ANOVA $p = .007$) und Schwindelstation (ANOVA $p = .003$). Die jeweiligen Unterschiede machten allerdings weniger als den Unterschied zwischen zwei Schulnoten aus und sind daher nicht als „klinisch“ bedeutsam einzustufen. Darüber hinaus war dies der einzige auffällige Unterschied im Abschneiden beim Vergleichs-OSCE der mit soziodemografische Einflussgrößen, also hier dem Geschlecht, assoziiert war. Damit ist die Hypothese 4: „Inwiefern sind Unterschiede in der OSCE-Performance mit soziodemografischen Charakteristika assoziiert?“ untersucht, denn für weitere Mittelwertunterschiede zum Beispiel in Bezug auf Migrationsstatus, Alter, Bildungs- oder Familienstand waren die Gruppengrößen und Fallzahlen aufgrund der großen Homogenität der Teilnehmer:innen zu gering (siehe 5.1: Allgemeine Beschreibung der Stichprobe) und eine weitere statistische Exploration nicht indiziert.

Am besten schnitt die gesamte Gruppe aller Teilnehmer:innen in der Blutabnahmestation ab, wo sie im Durchschnitt eine 2 (Schulnote „gut“, $M = 17$ Punkte; $SD = 3,1$) erreichte. Am schlechtesten waren die Ergebnisse auf der Schwindelanamnesestation der Neurologie – mit einem Mittelwert von 11,4 Punkten (Schulnote „ungenügend“ $M = 11,4$ Punkte; $SD = 3,6$) reichte es für viele Teilnehmer:innen nicht einmal zum Überwinden der Bestehensgrenze von

12 Punkten, dicht gefolgt von der EbM-Stationen mit einem Durchschnitt von 12,4 Punkten (Schulnote „ausreichend“, $SD = 5,3$). In allen anderen Stationen bestand der oder die Durchschnittsteilnehmerin mit einer 3 (Schulnote „befriedigend“). Auffallend ist dabei der Range: grundsätzlich erreichte in jeder Station mindestens eine Person das maximal mögliche Punktergebnis von 20, also einen perfekten Score. In der Station Totenschein und bei der Blutentnahme wurde das andere Ende der Benotungsskala mit 0 beziehungsweise 1 erreichtem Punkt ebenfalls ausgeschöpft.

Tabelle 5. Übersicht der OSCE-Leistung der Teilnehmer:innen

Station	N	M	Note	Range (0-20 bzw. 0-160)	SD
Blutdruckberatung*	99	15,6	3	9-20	3,1
Somatisierungsstörung	100	15,2	3	8-20	2,5
Blutentnahme	100	17,0	2	1-20	1,8
Schwindelanamnese	100	11,4	5	4-20	3,6
Reanimation	100	15,5	3	5-20	3,1
Anamnese Abdomen	100	14,6	3	7-20	3,1
Totenschein	100	15,3	3	0-20	4,1
Evidenzbasierte Medizin	99	12,3	4	2-20	5,3
Gesamt	99	117,6	3	68-148	13,2

Eigene Darstellung.

*(*Ein Teilnehmer musste vorzeitig die Testung beenden und fehlt daher in zwei Stationen.)*

5.3. Vergleich zwischen den Studiengangsformen

Unsere eingangs vorgestellte Hypothese 1, dass es signifikante Gruppenunterschiede bezüglich der gezeigten Leistung im Vergleichs-OSCE zwischen den Absolvent:innen der verschiedenen Studiengangsformen geben würde, konnte nicht bestätigt werden. Es fanden sich keine signifikanten Unterschiede bei der Gesamtnote (RSG: $M = 3,08$ $SD = 0,84$; iMED: $M = 3,17$ Punkte, $SD = 1,0$; ANOVA $p = .616$) oder Punktzahl (RSG: $M = 118,31$ Punkte, $SD = 12,9$; iMED: $M = 117,46$ Punkte, $SD = 13,4$; ANOVA $p = .747$). Auch die Streuung ist gering und die beiden Gruppen als sehr homogen zu bezeichnen.

Auf Stationsebene fand sich ein einziger signifikanter Unterschied ($p = .010$) bei der Abdomen-Station: die Absolvent:innen des RSG schnitten um 1,5 von 20 Punkten besser ab als ihre Gegenparts aus iMED. Dieser Unterschied der Gruppen ist als gering und nicht relevant einzuschätzen und macht keinen Notensprung aus.

Absolvent:innen der verschiedenen Curricula unterschieden sich nicht hinsichtlich ihrer Noten in kommunikativen (RSG: $M = 2,56$, $SD = 1,26$ sowie iMED: $M = 2,7$, $SD = 1,32$; ANOVA $p = .582$) oder klinisch-praktischen Fertigkeiten (RSG: $M = 2,83$, $SD = 0,73$, sowie iMED: $M = 2,9$,

$SD = 0,74$; ANOVA $p = .677$). Daher konnten wir auch Hypothese 1a und 1b nicht unterstützen: weder im RSG noch in iMED ergibt sich ein auffälliger (also weder statistisch noch klinisch relevanter) Unterschied zu anderen Gruppe in Bezug auf die KUMplusKOM-Kompetenzen.

5.4. Angestrebte Facharzttrichtung

Tabelle 6. Angestrebte Fachweiterbildungsrichtung der Teilnehmer:innen, $N = 98$

Fachrichtung	Nennungen / %
Allgemeinmedizin	8
Anästhesie	18
Frauenheilkunde und Geburtshilfe	6
Innere Medizin / Kardiologie / Gastroenterologie	19
Neurologie	10
Pädiatrie	11
Psychiatrie / Psychosomatik & Psychotherapie (auch Kinder- und Jugend)	5
Schneidende Disziplinen (mit und ohne Schwerpunkt)	12
Sonstige (Augenheilkunde, Dermatologie, Hämatologie, Radiologie, Tropenheilkunde, Kinderdiabetologie)	7
Weiß nicht	2

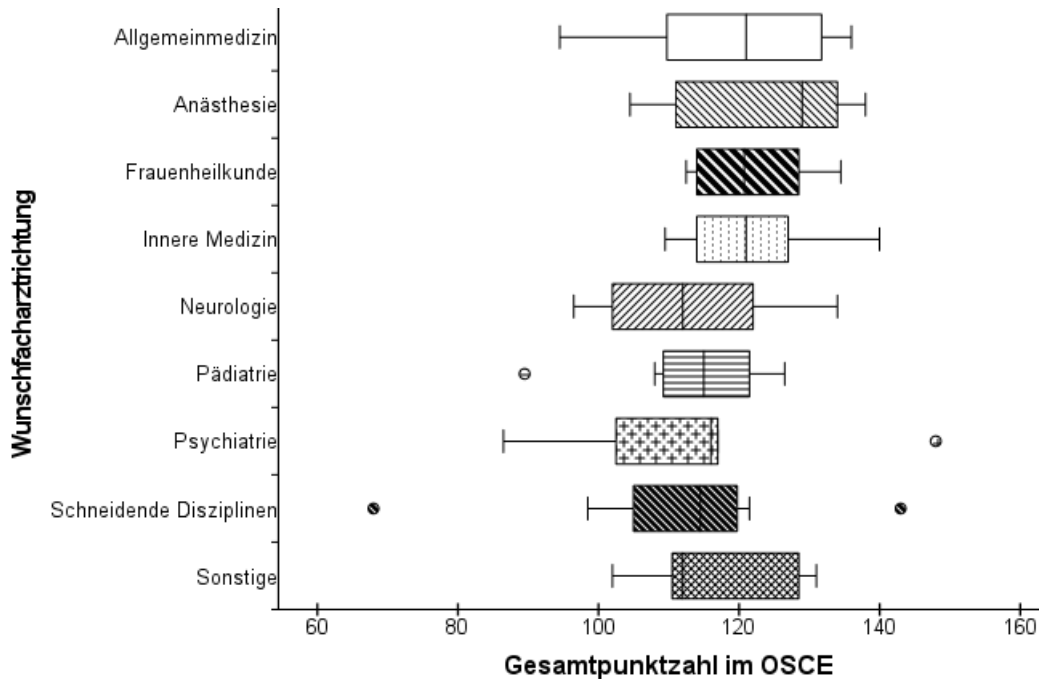
Eigene Darstellung

98 % der Befragten gaben an, nach dem Abschluss ihres Studiums eine Facharztweiterbildung anstreben zu wollen. Vor der Auszählung der Einzelnennungen kategorisierten wir diese sowohl anhand formaler als auch inhaltlicher Kriterien (siehe 4.7). Danach ergaben sich 9 verschiedene Kategorien, sowie die Nennung „weiß nicht“ (vergleiche Tabelle 6).

Die beiden am häufigsten genannten angestrebten Fachrichtungen waren „Innere Medizin / Kardiologie / Gastroenterologie“ mit 19 Nennungen, als auch die Richtung „Anästhesie“ (18 Nennungen). Mit etwas Abstand folgten „Schneidende Disziplinen (mit und ohne Schwerpunkt)⁷“ (12 Nennungen), „Pädiatrie“ (11 Nennungen) sowie „Neurologie“ (10 Nennungen). 71,43 % (70 Personen) der Befragten entfielen auf diese 5 Kategorien.

⁷ Die Begrifflichkeit der „Schneidenden Disziplinen“ ist dabei an van den Bussches Operationalisierung (2006) angelehnt. Vielleicht wäre „chirurgische Fächer“ als Bezeichnung treffender, da zum Beispiel auch Gynäkolog:innen „schneiden“, davon wurde aber aus Gründen der Vergleichbarkeit an dieser Stelle Abstand genommen.

Abbildung 6. Zusammenhang zwischen angestrebter Wunschfachrichtung und OSCE-Note, N= 98



Eigene Darstellung

Auf Wunschfachebene rechneten wir mit Mittelwerten des Gesamtergebnisses, da eine Normalverteilung angenommen werden kann (ANOVA $p = .201$). Es ergaben sich signifikante Unterschiede im Gesamtabschneiden beim Vergleichs-OSCE zwischen den verschiedenen Wunschfachrichtungsgruppen, allerdings nur gemessen an der Schulnote: Anästhesie-, Frauenheilkunde-, Innere Medizin-Interessierte erreichten die höchsten Durchschnittsnoten und -Punktzahlen (Anästhesie: $M = 124,21$, $SD = 11,4$; Frauenheilkunde: $M = 121,83$, $SD = 8,7$; Innere Medizin: $M = 121,3$, $SD = 8,9$). Die Teilnehmer:innen, die sich für die Schneidenden Disziplinen- und die an der Neurologie-Interessierten, erzielten die geringsten Durchschnittsnoten, bzw. -Punktzahlen (Schneidende Disziplinen: $M = 111,1$, $SD = 18,5$; Neurologie: $M = 112,8$, $SD = 12,1$). Allerdings belief sich der ermittelte Unterschied zwischen dem niedrigsten Mittelwert (also Punktzahl im Gesamt-OSCE pro Wunschfach) und dem höchsten auf 12 von maximal 160 Gesamtpunkten und ist damit geringer als der Abstand zwischen einer Schulnote zur anderen (15 Punkten).

5.5. Psychometrie

Im Folgenden beleuchten wir unsere ungerichtete Hypothese 2: Inwieweit ist die Persönlichkeit der Student:innen mit der OSCE-Performance assoziiert? Zeigen zum Beispiel Personen mit hohen Ausprägungen der Persönlichkeitsvariable Gewissenhaftigkeit bessere Ergebnisse im Test zur Arztreife? Oder bessere Ergebnisse in einzelnen Stationen? Hypothese 3 bewegt sich im Anschluss daran von der Performanzperspektive weg, um sich mit den Zukunftsvorstellungen und Interessen der Teilnehmerinnen zu beschäftigen: Präferieren unterschiedliche Persönlichkeiten unterschiedliche Weiterbildungsfächer?

5.5.1. NEO-FFI 30: Erfassung der Persönlichkeitseigenschaften

Für diesen Teil des Fragebogens konnten wir 97 Teilnehmer:innen auswerten (50 RSG-, 46-iMED-Studierende), Für einen Überblick für die Skalenergebnisse siehe Tabelle 7.

Tabelle 7. Ergebnisse des NEO-FFI-30, $N = 97$

Skala	<i>M</i>	<i>SD</i>	Range (0-4)
Neurotizismus	1,14	0,67	0 – 2,8
Extraversion	2,56	0,52	1,17 – 3,5
Offenheit	2,61	0,75	0,33 – 4
Verträglichkeit	1,46	0,53	0,67 – 3
Gewissenhaftigkeit	2,71	0,34	1,5 – 3,5

Eigene Darstellung

Auffallend an den Ergebnissen ist die sehr geringe Streuung. Die Gesamtstichprobe erreichte auf der Skala Neurotizismus einen Mittelwert von 1,14 ($SD = 0,67$) von maximal 4 Punkten, einzelne Teilnehmer:innen erzielten sogar 0 Punkte, während der höchste Wert eine 2,83 war. Die untersuchte Stichprobe setzte sich aus tendenziell wenig als neurotisch einzustufenden Personen zusammen. Die Skala wurde nach oben hin nicht ausgeschöpft und die Ergebnisse sind insgesamt als homogen einzuschätzen.

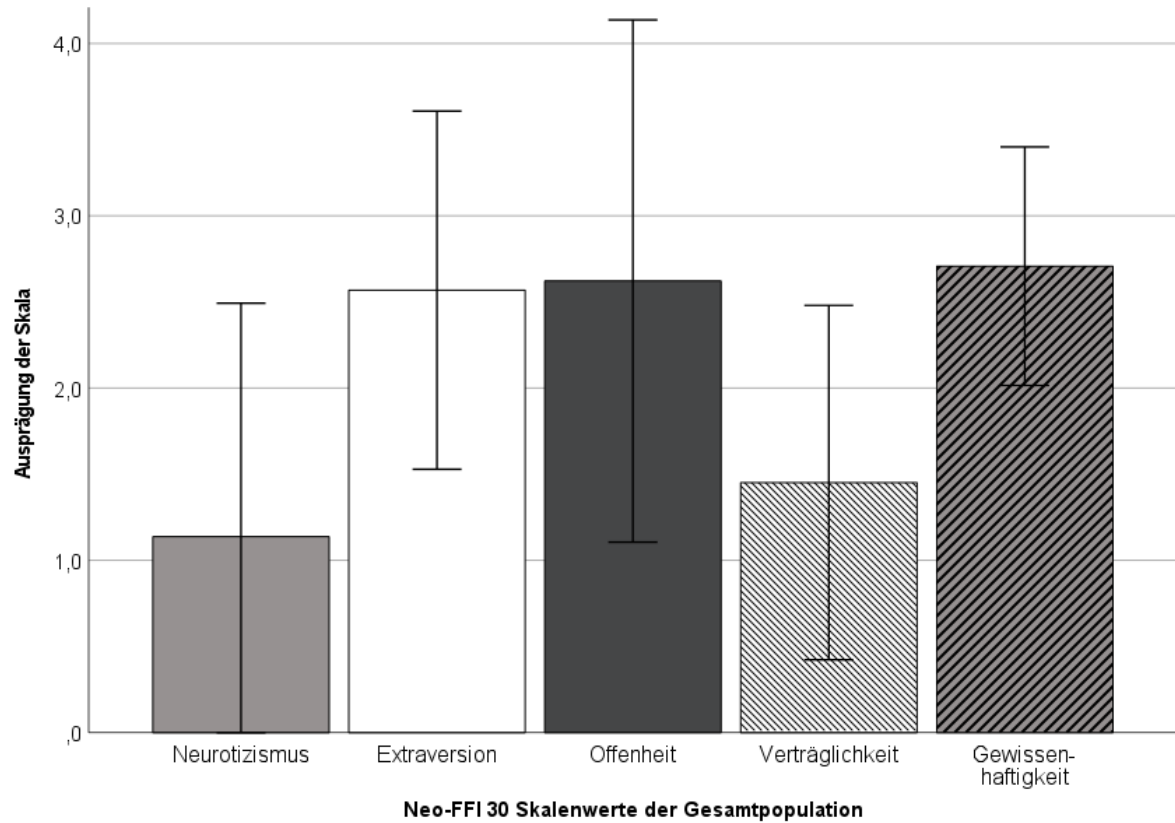
Für die Skala Verträglichkeit zeigte sich ein ähnlicher Trend: ein niedriger Mittelwert von 1,46; ($SD = 0,53$); ein Minimum von 0,67 und Maximum von 3. Die Teilnehmer:innen sind also im Durchschnitt als wenig verträglich einzuschätzen.

Bei der Gewissenhaftigkeitsskala und den beiden folgenden Facetten zeigte sich ein umgekehrtes Ergebnis: die Teilnehmer:innen waren im Mittel recht gewissenhaft ($M = 2,71$; $SD = 0,34$; Range = 1,5 – 3,5).

Ebenfalls können wir davon ausgehen, dass unsere Teilnehmer:innen mit einem Mittelwert von 2,56 überdurchschnittlich extravertiert waren ($SD = 0,52$; Range = 1,17 – 3,5).

Die Persönlichkeitsfacette Offenheit schließlich erreichte die zweithöchste Ausprägung unter den Teilnehmer:innen mit einem Mittelwert von 2,61 Punkten ($SD = 0,75$; Range 0,33 - 4), wies allerdings den höchsten Range auf.

Abbildung 7: NEO-FFI 30-Ergebnisse der Teilnehmer:innen, (N=97, N=99)

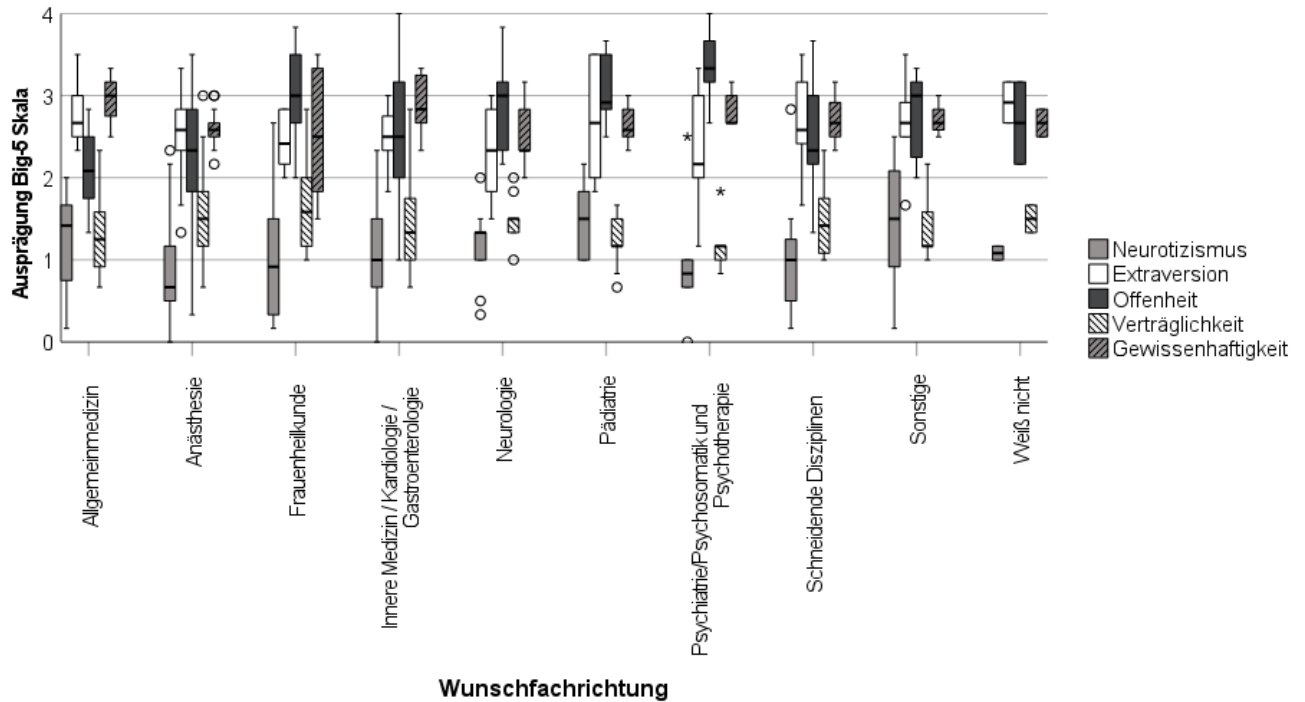


Eigene Darstellung

Zusammenfassend wiesen also alle Teilnehmer:innen eher niedrige Neurotizismus- und Verträglichkeitswerte, aber hohe Gewissenhaftigkeits-, Offenheits- und Extraversionswerte auf. Zwischen den Teilgruppen aus den verschiedenen Studiengangsformen bestanden keine signifikanten Unterschiede, ein Levene-Test ergibt Varianzgleichheit. Bei der Untersuchung der Hypothese 3, in der wir uns mit dem Zusammenhang zwischen Persönlichkeit und Wunschweiterbildungsfach beschäftigten, ergab sich keine sichtbare Differenzierung zwischen Studieren-

den: unabhängig von der Wunschfachrichtung sind unterschiedliche Persönlichkeitsfacetten unterschiedlich stark ausgeprägt. Die Ergebnisse folgen dem Muster des „Gesamtprofils“ aller Teilnehmer:innen (vergleiche Abbildung 8).

Abbildung 8. Big-5-Ausprägung und Wunschfachrichtung der Stichprobe, N =97



Eigene Darstellung

Die nach Studiengangsformen getrennte Analyse mittels Levene-Test ergab wiederum Varianzgleichheit. Es gibt keine signifikanten Unterschiede zwischen den Angehörigen beider Studiengänge bezüglich der Ausprägung der verschiedenen Neo-FFI-Skalen (Signifikanzwerte der Skalen: Neurotizismus: ANOVA $p = .743$; Extraversion: ANOVA $p = .155$; Offenheit, ANOVA $p = .422$; Verträglichkeit ANOVA $p = .895$; und Gewissenhaftigkeit ANOVA $p = .327$).

Dabei ergab sich für die Persönlichkeitsskalen Extraversion (♀: $M = 2,57$, $SD = 0,51$ beziehungsweise ♂: $M = 2,58$, $SD = 0,51$, ANOVA $p = .913$) und Offenheit (♀: $M = 2,6$, $SD = 0,83$ beziehungsweise ♂: $M = 2,63$, $SD = 0,67$, ANOVA $p = .831$) allerdings keine signifikanten Unterschiede zwischen den Geschlechtern. Für die Facette Gewissenhaftigkeit fand sich ein statistisch signifikanter Unterschied zwischen den Ausprägungen der Geschlechter. Männliche Teilnehmer erreichten durchschnittlich um 0,5 Punkte geringere Werte (95%-CI[-0,73, -0,23]), $t(97) = -3,79$, $p < .001$). Die Männer unserer Stichprobe waren demnach geringfügig weniger gewissenhaft als die Frauen. Auch bei der Facette Neurotizismus fand sich ein entsprechender Unterschied von weniger neurotischen Männern als Frauen (♂: $n = 50$, $M = 0,9$, $SD = 0,6$; Range: 0 – 2.5. versus ♀: $n = 49$, $M = 1,4$, $SD = 0,7$; Range 0,2 -; 2,8).

Anders war der Unterschied bei der Skala Verträglichkeit beschaffen: hier war die Ausprägung der Männer durchschnittlich 0,35 Punkte höher (95%-CI/[0,15, 0,55]), $t(97) = 3,44, p < .001$), die Männer also verträglicher.

5.5.2. Satisfaction with Life Scale (SWLS): Globalrating Lebenszufriedenheit

Bei der Auswertung des SWLS konnten wir 98 Teilnehmer:innen in die Auswertung einschließen. Eine Person hatte diesen Abschnitt des Fragebogens nicht vollständig ausgefüllt und wurde ausgeschlossen. Die Rohsummen der Items wurden anschließend aufsummiert und dann gemäß der Auswertungsvorgaben interpretiert (vergleiche Tabelle 8).

Tabelle 8. SWLS-Auswertungskategorien und Ergebnisse der Teilnehmer:innen, $n = 98$

Lebenszufriedenheit nach SWLS	Punkte	<i>n</i>
extrem zufrieden	31 - 35	26
zufrieden	26 - 30	50
eher zufrieden	21 - 25	16
neutral	20	6
eher unzufrieden	15 - 19	nicht vergeben
unzufrieden	10 - 14	nicht vergeben
extrem unzufrieden	5 - 9	nicht vergeben

Eigene Darstellung

Dabei ergab sich ein vergleichsweise hoher Mittelwert von 27,8 Punkten (entspricht „zufrieden“) für die Gesamtstichprobe der Teilnehmer:innen aus beiden Studiengangsformen ($SD = 4.10$; Range = 20 - 35). Die Auszählung auf Einzelebene ergibt eine extrem schiefe Verteilung: 26 Personen sind „extrem zufrieden“; weitere 50 „zufrieden“, 16 „eher zufrieden“, 6 „neutral“. Dies bedeutet, dass 76 % der Teilnehmer höchstzufrieden bis zufrieden sind, und die beiden höchsten Kategorien erreichen. Die untersten drei Kategorien wurden gar nicht vergeben, niemand (!) ist „unzufrieden“.

Auch hier ist eine Analyse der Ergebnisse nach Studiengangsform nicht sinnvoll: der Levene-Test ergibt Varianzgleichheit. Es gibt keine signifikanten Unterschiede zwischen den Absolvent:innen beider Studiengangsformen in Bezug auf ihre Lebenszufriedenheit.

5.5.3. F-SoZu K14: Fragebogen zur Erfassung sozialer Unterstützung

Tabelle 9. Gemessene soziale Unterstützung der Stichprobe, $n = 74$

Niveau der sozialen Unterstützung	Prozentränge	% / (Anzahl)
außergewöhnlich hoch	98	16,22 % (12)
überdurchschnittlich hoch	85 - 97	28,38 % (21)
normal / unauffällig	16 - 84	54,05 % (40)
unterdurchschnittlich niedrig	3 - 15	1,35 % (1)

Eigene Darstellung

Die Auswertung des Instruments erfolgte über die Bildung des Mittelwerts durch Aufsummierung aller 14 Items und die Division durch die Itemanzahl, anschließend ordneten wir mittels der Normtabelle einen Prozentrang zu. 12 der 14 Items mussten beantwortet sein.

Prozentränge zwischen 16 und 84 können als normal/unauffällig interpretiert werden – sie entsprechen einer Standardabweichung vom Mittelwert. Von Prozentrang 3 - 15 gehen die Fragebogenautoren von deutlicher Abweichung, bei Prozenträngen darunter sehr auffällige Abweichung von der Norm im Bereich geringer sozialer Unterstützung aus. Analog stehen Prozentränge 85 - 97 für überdurchschnittlich hohe, darüber ab Prozentrang 97 für außergewöhnlich hohe Unterstützung jenseits der Bevölkerungsnorm (Fydrich et al., S. 36).

Dadurch, dass wir das Instrument erst ab E2 erhoben, war unsere Stichprobengröße mit $n = 74$ etwas geringer als bei den anderen Instrumenten. Die Gruppengrößen unterschieden sich: 27 der Teilnehmer:innen waren dem auslaufenden RSG, 46 dem iMED-Studiengang zuordbar. Eine weitere Person hatte nicht in Hamburg studiert, wurde aus inhaltlichen Gründen aber für die meisten Berechnungen jenseits der Hypothesenprüfung eingeschlossen. Die Stichprobe erreichte einen Mittelwert von 4.57 Rohpunkten ($SD = .42$; Range = 2,86 - 5), was einem Prozentrang (PR) von 84 entspricht. Das Mittel der Gesamtstichprobe entspricht damit einer normalen sozialen Unterstützung an der Grenze zu überdurchschnittlich hoher sozialer Unterstützung. 12 Personen unserer Stichprobe erzielten einen PR über 97 und berichteten damit von außergewöhnlich hoher Unterstützung. 21 Personen wiesen ein überdurchschnittlich hohes Maß an sozialer Unterstützung bei \geq PR 85 auf; weitere 40 Personen erreichten ein normales Maß. Eine einzige Person erzielte einen sehr niedrigen PR von 6 und damit unterdurchschnittlich geringe soziale Unterstützung (vergleiche Tabelle 9).

Die Mittelwerte sind für beide Studiengänge (iMED= 4,49 (PR 72); $SD = .42$ und RSG = 4,69 (PR 81); $SD = 0.42$) vergleichbar. Zwischen den Studiengangsformen bestanden auch hier keine signifikanten Unterschiede (ANOVA $p = .054$), ein Levene-Test ergab wiederum Varianzgleichheit.

Durch die extrem linksschiefe Verteilung konnten wir wiederum nicht mit dem prädiktiven Wert des Instruments arbeiten. Eine logistische Regression kann aufgrund der Beschaffenheit der Daten keinen validen p-Wert berechnen.

5.5.4. Hospital Anxiety and Depression Scale (HADS-D)

Die Auswertung erfolgte für beide Skalen getrennt durch Invertierung der negativ gepolten Items, anschließend der Aufsummierung der Rohitemwerte, und schließlich der Division durch die Itemanzahl. Werte von ≤ 7 sind als unauffällig einzuschätzen, 8 -10 Punkte gelten als grenzwertig, und Werte ≥ 11 deuten auf Auffälligkeiten in der Ausprägung der Facetten und weiteren Untersuchungsbedarf hin (vergleiche Tabelle 10). Maximal konnten sehr beeinträchtigte Teilnehmer:innen auf jeder Skala 21 Punkte erreichen.

Tabelle 10. *Ergebnisse Auswertung HADS-D, n = 74*

Subskala HADS-D	Ausprägung	% (n)
Angst	unauffällig (≤ 7 Punkte)	87,8 % (65)
	grenzwertig (8-10 Punkte)	9,5 % (7)
	auffällig (≥ 11 Punkte)	2,7 % (2)
Depression	unauffällig (≤ 7 Punkte)	95,9 % (71)
	grenzwertig (8-10 Punkte)	4,0 % (3)
	auffällig (≥ 11 Punkte)	nicht vergeben

Eigene Darstellung

Auch hier war die Stichprobengröße mit $n = 74$ etwas geringer, als zum Beispiel beim NEO-FFI-30. Die Gruppengrößen in der Zusammenfassung über alle Testzeitpunkte unterscheiden sich wieder, da das Instrument erst in E2 hinzukam (wie beim F-SoZu-K14 in RSG = 27, iMED = 46 Personen, sowie ein auswärtiger Student).

Es gibt keine signifikanten Unterschiede zwischen den Studiengängen bezüglich beider Subskalen (iMED: Angst: $M = 4,8$; Depression: $M = 2,3$; $SD = 2,2$ sowie $2,0$; $p = .285$ beziehungsweise $.605$ und RSG: Angst: $M = 4,11$; Depression: $M = 2,0$; $SD = 2,2$ und $2,0$). Der Range war linksschief mit 0 – 9 Punkten auf der Depressionsskala und 0 – 13 Punkten auf der Angstfacette. Damit deuten die Ergebnisse für keine Teilgruppe auf eine Belastung hin. Fast 96 % (95,95) der Untersuchten berichteten unauffällige Depressionswerte, die untere Kategorie auffälliger Werte wurde hier nicht vergeben. Die Werte der Angstskala waren etwas weniger schief, dennoch berichteten auch hier fast 90 Prozent (87,8) der Befragten von keiner Belastung. Die zusätzlich von den Autoren empfohlene weitere Differenzierung in

„schwere“ (Wertebereich 11 - 14) und „sehr schwere“ (Wertebereich 15 - 21) Symptome ist für beide Gruppen aus Mangel an diesen Ergebnissen nicht durchführbar, einzig auf der Angstskala erreichte eine Einzelperson 13 Punkte. Unsere Teilnehmer:innen sind im Mittel also nicht im Geringsten als ängstlich oder depressiv einzuschätzen.

Auch hier konnten wir nicht inferenzstatistisch mit dem prädiktiven Wert des Instruments arbeiten, da wir keine „kranken“ Teilnehmer:innen erfassen konnten. Eine logistische Regression kann aufgrund der Beschaffenheit der Daten keinen validen p -Wert berechnen

6. Diskussion

Nachdem die Methodik und alle relevanten Daten in den letzten Kapiteln in aller Ausführlichkeit berichtet wurden, beginnt dieser Abschnitt nun mit einem allgemeinen Überblick und einem Fokus auf die wichtigsten Ergebnisse, die dann genauer besprochen und diskutiert werden, ausgehend von den methodischen Stärken und Schwächen der Arbeit. Darauf folgend werden die psychometrischen und soziodemografischen Ergebnisse besonders beleuchtet. Zum Abschluss werden die Schlussfolgerungen und Implikationen weiter ausgeführt. Die Arbeit endet mit einer kurzen allgemeinen Zusammenfassung.

6.1. Überblick: Die wichtigsten Ergebnisse

1. Wenn man das OSCE-Gesamtergebnis als Kriterium für die Allgemeine Ärztliche Reife der Teilnehmer:innen annimmt, ist durch die erreichte Durchschnittsnote von 3 („befriedigend“) anzunehmen, dass die weit überwiegende Anzahl der Teilnehmer:innen diese erreicht haben.
2. Es finden sich keine signifikanten oder klinisch bedeutsamen Unterschiede zwischen den Absolvent:innen beider Studiengangsformen bezüglich der Performanz im Vergleichs-OSCE, der Psychometrie, Soziodemografie oder hinsichtlich der angestrebten Wunschfacharzttrichtung.
3. Die Teilnehmer:innen verfügen über ein „Persönlichkeitsnormprofil“ (eher niedrige Neurotizismus- und Verträglichkeitswerte, aber hohe Gewissenhaftigkeits-, Offenheits- und Extraversionswerte). Die restliche Psychometrie deutet auf eine sehr psychisch gesunde Stichprobe mit extrem geringen Angst- und Depressionswerten, sehr hoher sozialer Unterstützung, und sehr hoher Lebenszufriedenheit hin.

6.2. Methodische Stärken und Schwächen der Arbeit

6.2.1. Allgemeine methodische Herausforderungen

Ursprünglich geplant war die Rekrutierung von insgesamt 120 Teilnehmer:innen, um statistisch durch potentiell zu niedrige Gruppengrößen nicht eingeschränkt zu sein. Dieses Ziel wurde trotz intensiver Bemühungen leider nicht erreicht, am Ende wurden 99 Testpersonen, also knapp 50 Personen pro Studiengang erfolgreich eingeschlossen. Diese Reduktion der angestrebten Teilnehmerzahl um 12,6 % erschwerte die statistische Auswertung und Aussa-

gekraft der Ergebnisse stellenweise enorm. Durch die kleinen Gruppengrößen und eine zusätzlich insgesamt sehr homogene Stichprobe war es nicht sinnvoll, mit diesen Daten Regressionen oder Faktorenanalysen zu rechnen.

Allerdings ist auch festzuhalten, dass die Rekrutierung einer solchen Menge geeigneter Personen aus einem ohnehin kleinen Pool von Kandidat:innen, zu einem schwierigen Ausbildungszeitpunkt (konkurrierend mit PJ und Abschlussprüfungen) und mit Aussicht auf ein für den oder die Einzelne durchaus auch belastendes prüfungsähnliches Testungsverfahren besonders herausfordernd scheint. Die nachträgliche Inzentivierung der Testteilnahme durch geldwerte Gutscheine und die persönliche Auswertung der Testungsergebnisse zur Prüfungsvorbereitung kann allerdings als erfolgreiche Maßnahme beurteilt werden, da sie den Teilnehmer:innenfluss erheblich verbessert hat. Dies passt zu existierenden Forschungsergebnissen, dass ein Großteil der Rekrutierungsschwierigkeiten in Wissenschaftskontexten auf Motivationsmangel der geeigneten Testpersonen, zu harte Einschlusskriterien oder eine Überschätzung des Probandenpools rückführbar sind (vergleiche zum Beispiel Briel et al. (2016)).

6.2.2. Effekte der Testkonstruktion

Grundsätzlich lässt sich diskutieren, ob das Vorgehen, die durch die systematische Literaturanalyse theoretisch identifizierten KUMplusKOM-Kompetenzen der Allgemeinen Ärztlichen Reife mit den Lehrinhalten des iMED-Curriculums zu verschränken und daraus ein Modell für die Inhalte der Vergleichs-OSCE-Testung zu erstellen, dasselbe aber nicht ebenso für den Regelstudiengang zu machen oder zumindest abzugleichen, möglicherweise nicht weitreichend genug oder einseitig war. Dadurch, dass das Konzept der Arztreife in der Literatur quasi nicht existent ist, war es schließlich im Prinzip eine Modellentwicklung entlang der KUMplusKOM-Komponenten des iMED-Studienganges.

Es besteht die nicht unwahrscheinliche Möglichkeit, dass die im Regelstudiengang vermittelten ärztlichen Kompetenzen implizit dieselben, oder zumindest in großen Teilen vergleichbar waren. Schließlich existierte das RSG-Curriculum in seiner letzten Form auch erst einige Jahre und wurde damals ebenfalls von Expert:innen sorgfältig mit dem Ziel entwickelt und reformiert, hochqualifizierte Ärzt:innen auszubilden. Das könnte erklären, warum sich die Absolvent:innen beider Studiengangsformen hinsichtlich ihrer Performanz beim Vergleichs-OSCE kaum unterschieden. Darüber hinaus ist davon auszugehen, dass zu großen Teilen dieselben Lehrpersonen an der Ausbildung beider Studiengangsteilnehmer:innen beteiligt waren – insbesondere, da wir die ersten mit den letzten Absolvent:innen der jeweiligen Studiengangsform verglichen haben und alleine durch die zeitliche Nähe weniger Varianz bezüglich des Personals zu erwarten ist. Es ist wahrscheinlich (und wünschenswert), dass diese sicher nicht bewusst

„schlechtere“ Lehre im Regelstudiengangsformat angeboten haben, auch wenn sich die Reihung und Präsentation der Lehrinhalte strukturell unterschieden. Da es einige Hinweise darauf gibt, dass akademischer Lernerfolg und Wissenserwerb nicht nur durch die reine Präsentation des zu lernenden Wissens, sondern auch durch das Rollenvorbild und die Persönlichkeit des oder der Lehrenden beeinflusst wird, erscheint ein größerer Einfluss der Person der Dozent:innen nicht abwegig (Harden & Lilley, 2018). Auch motiviert insbesondere die Abschlussprüfung zum Lernen bestimmter Inhalte (Jünger, 2018; Raupach, Brown, Anders, Hasenfuss, & Harendza, 2013), und die Prüfungen haben sich wenig verändert. Daraus ließe sich ableiten, dass die vorhergehenden Generationen der erfolgreichen Absolvent:innen des Regelstudiengangs ebenso über gute Qualifikationen hinsichtlich ihrer ärztlichen Reife verfügen müssten. Auf der anderen Seite könnte es natürlich auch bedeuten, dass die Reformation des Medizincurriculums in Hamburg bisher keine großen messbaren Unterschiede bei den Absolvent:innen erzeugt hat. Andererseits könnten sich noch stärkere Unterschiede zwischen den Lehrinhalten und Dozierenden entwickeln, sobald der Modellstudiengang iMED fest etabliert ist und die Übergangsphase zwischen den Studiengängen mit allen praktisch-strukturellen Herausforderungen länger abgeschlossen ist, als zum Zeitpunkt der Datengewinnung.

6.2.3. Selektionseffekte

Anknüpfend an die oben ausgeführten Bedenken bezüglich der möglichen Beeinflussung der Ergebnisse durch die Testkonstruktion, muss an dieser Stelle auch dem Selektionsbias in der Datengewinnung dieser Arbeit Sorge getragen werden. Es liegt in der Natur der Stichprobe, nämlich dem Vergleich der letzten Absolvent:innen des einen, „alten“ Studiengangs mit denen des neuen begründet, dass vermeintlich besonders langsame Studierende mit denen, die das neue Studium auf Anhieb in Regelstudienzeit absolviert hatten, verglichen werden mussten. Genauso ist es denkbar, dass die Absolvent:innen aus iMED im Verlauf ihres Studiums großzügiger bewertet wurden, da sie in dem neu konzipierten Studiengang direkt nach dessen Einführung manche Herausforderungen in der Organisation aushalten mussten.

Es ist vorstellbar, dass in dieser Arbeit besonders leistungsschwache oder in anderen Lebensbereichen benachteiligte Studierende mit besonders Leistungsstarken verglichen werden. Möglicherweise liegt sogar ein regelrechter Exzellenzbias vor, und es haben sich ohnehin nur solche Studierende für eine Teilnahme entschieden, die besonders leistungsstark und motiviert waren. Auch ist anzunehmen, dass sich introvertierte Studierende, und/oder solche mit eher neurotisch-ängstlichen Persönlichkeiten mit geringerer Wahrscheinlichkeit zu einer Testteilnahme anmeldeten. Es ist möglich, dass das finanzielle Incentive eher Student:innen angesprochen hat, die das Geld besonders gut gebrauchen konnten. Andererseits ist vorstellbar, dass Student:innen, die gerade besonders hohe Stressniveaus aufweisen, viel neben dem

Studium arbeiten, oder an Prüfungsängsten leiden, sich gegen eine Teilnahme entschieden haben. Sind die Teilnehmer:innen also überhaupt repräsentativ für das Gros der Medizinstudierenden? Tatsächlich passen die hier aufgedeckten Persönlichkeitseigenschaften gut zu Arbeiten von (Bussenius & Harendza, 2019): in einer Untersuchung an Hamburger Medizinstudiengangsbewerber:innen konnten diese nachweisen, dass zum Studium zugelassene Bewerber:innen signifikant höhere Extraversionswerte und geringer Depressionswerte aufwiesen als die abgelehnte Vergleichsgruppe. Gleichzeitig verfügten die zugelassenen Kandidat:innen über höhere Raten sogenannten adaptiven Perfektionismus, ein Konzept, das inhaltliche Schnittmengen mit der Big 5 Skala „Gewissenhaftigkeit“ aufweist. Diese Ähnlichkeiten lassen vermuten, dass die in dieser Arbeit untersuchten Personen durchaus repräsentativ für zumindest die Hamburger Medizinstudierenden sind.

Von zusätzlicher Bedeutung für die generelle Verallgemeinerbarkeit ist die besondere Beschaffenheit unserer ohnehin grundlegend stark vorselektierten studentischen Ausgangsstichprobe. Medizinstudierende sind auch ohne die oben genannten Selbstselektionsprozesse eine nicht repräsentative Stichprobe für Studierende an sich (oder generell für den Durchschnittsmenschen dieser Altersgruppe in Deutschland, aber mit diesem ist ein Vergleich im Rahmen dieser Arbeit inhaltlich ohnehin nicht sinnvoll). Eine große Anzahl der Bewerber:innen für Medizinstudiengangsplätze durchlaufen in der Regel einen umfangreichen Auswahlprozess, der einerseits kognitive Fertigkeiten wie Abiturnote oder die Performanz im HAM-Nat überbetont (die Auswahl der Studierenden geschieht an der Fakultät nicht nur unter Berücksichtigung des Numerus clausus: aktuell werden 30 % ausschließlich über ihre hervorragende Abiturnote zugelassen), implizit aber auch durch den hohen Druck Persönlichkeitsunterschiede wie Gewissenhaftigkeit, Neurotizismus, aber auch Stresstoleranz, und soziale, finanzielle und Bildungsressourcen bedingt wird (Hissbach et al., 2012). Zusätzlich sind diese interindividuellen Eigenschaften und Gegebenheiten auch Einflussgrößen auf die Performanz in Prüfungssituationen, Weiterbildungserfolg und kommunikative Fertigkeiten – alles essenzielle Aspekte zur Erlangung der Arztreife auch nach Antritt des Studiums.

6.3. Allgemeine ärztliche Reife und Vergleichs-OSCE

Die Teilnehmer:innen erreichten bei einer normalverteilten Notenstreuung beim Vergleichs-OSCE durchschnittlich ein „befriedigend“ als Gesamtnote. Dies entspricht tatsächlich der Approbationsrealität – man könnte dieses Ergebnis als einen Hinweis auf die Güte des hier verwendeten selbst entwickelten Verfahrens werten, wenn man davon ausgeht, dass die Approbationsprüfungen tatsächlich auch die allgemeine Arztreife erfassen.

Entgegen der Arbeitshypothese waren die wenigen Unterschiede zwischen den Gruppen oder Teilnehmer:innen in Bezug auf die Performanz bei der Vergleichs-OSCE-Testung sehr gering

– und verfügten selbst bei statistischer Signifikanz über keine praktische Relevanz, da sie nicht annähernd Notensprünge ausmachten. Tatsächlich sind die Ergebnisse zwischen den Studiengängen und Teilnehmern vielmehr als überraschend homogen zu bezeichnen. Bezüglich der Hauptfragestellung, inwieweit sich Absolvent:innen der unterschiedlichen Studiengangsformen in der allgemeinen ärztlichen Reife unterschieden, konnten durch diese Untersuchung keine Hinweise auf Differenzen erbracht werden. Dadurch konnte auch die Vermutung, dass die Absolvent:innen des iMED-Curriculums besser abschneiden würden, nicht bestätigt werden. Einige mögliche Gründe dafür (Curriculumsinhalte und Lehrpersonal, individuelle Unterschiede, Testkonstruktion oder Gruppengrößen) wurden oben schon genauer beleuchtet.

Es finden sich darüber hinaus in der internationalen Literatur vielfältige Ansätze, um medizinischen Studienerfolg beziehungsweise später dann erfolgreiche ärztliche Tätigkeit zu definieren (vergleiche zum Beispiel Steiner-Hofbauer et al., 2018;). Es ist denkbar, dass der für diese Arbeit verwendete Arztreifebegriff trotz sorgfältiger Untermauerung Unzulänglichkeiten aufwies, und die Inhaltsvalidität nicht in vollem Umfang sichergestellt war. Hierzu passt, dass in der internationalen Literatur zum damaligen Zeitpunkt keine operationalisierten Korrelate aufzufinden waren.

Lediglich durch Geschlecht und Wunschfachrichtung fanden sich nennenswerten Einflüsse auf die Performanz, die in den folgenden Unterkapiteln näher ausgeführt werden.

Ein weiterer Aspekt, auf den hier bisher noch nicht eingegangen wurde, ist, dass mögliche Unterschiede zwischen den Testungsteilnehmer:innen zu einem anderen Zeitpunkt der jeweiligen Berufsbiografie spürbar werden könnten, was durch den querschnittlichen Charakter dieser Arbeit nicht erhoben werden konnte. Es ist zum Beispiel durchaus vorstellbar, dass in der späteren Berufsausübung durch ihre gezieltere universitäre Vorbereitung die iMED-Absolvent:innen erfolgreicher in Teams arbeiten können, oder konstruktiver mit Angehörigen von Patient:innen kommunizieren – Fertigkeiten, die beim Vergleichs-OSCE nicht explizit abgeprüft wurden.

6.3.1. Ergebnisse der einzelnen Vergleichs-OSCE-Stationen

Es wurden kaum Zusammenhänge zwischen der Performanz an einzelnen Teststationen und Studierendencharakteristika gefunden. Allerdings fanden sich Geschlechtsunterschiede bei einigen wenigen: so schnitten weibliche Teilnehmerinnen signifikant besser in der Schwindel- und Blutentnahmestation ab, als ihre männlichen Mitstreiter, und zwar ungeachtet der Studiengangszugehörigkeit. Es finden sich in der Literatur einige mögliche Erklärungen für diesen Unterschied: so könnte zum Beispiel eine bessere Feinmotorik für spezifische Aufgaben und im Durchschnitt kleinere Hände unter den Frauen für ihre erfolgreichere Blutentnahme sorgen

(Tomasits & Haber, 2016). Höhere kommunikative Fertigkeiten könnten auf der Schwindelstation die besseren Ergebnisse erzeugt haben (Graf et al., 2017). Auch ist mittlerweile sehr belastbar nachgewiesen worden, dass Frauen im Durchschnitt grundsätzlich besser bei Prüfungsvorbereitungen und akademischem Lernen (Wang & Degol, 2017) abschneiden. Dieser Effekt müsste allerdings dann auch bei den anderen Stationen spürbar und könnte durch die Vorselektion der Stichprobe durch Abiturnote und akademische Leistung nivelliert worden sein. Zuletzt besteht die Möglichkeit, dass ein Bewertungsbias vorlag, entweder auf Ebene der Testkonstruktion, oder durch die prüfenden Personen.

Die Blutentnahmestation war diejenige, die auch unabhängig vom Geschlecht durchschnittlich die besten Ergebnisse, also die höchsten Punktzahlen und Noten, hervorbrachte. Dieses Ergebnis überrascht nur bedingt, da viele der Teilnehmer:innen über frische Klinikerfahrung verfügen und Blutabnahme eine Tätigkeit ist, die vom etablierten Klinikpersonal bekanntermaßen gerne an Studierende weitergegeben wird. Zudem arbeiten viele Studierende bereits vor oder parallel zu Ihrer Studienzeit in medizinischen Berufsfeldern und können dort diese praktische Fertigkeit gut erlernen und verfestigen. Leider wurden die Vorkenntnisse der Teilnehmer:innen nicht erfasst und können daher nicht berücksichtigt werden.

Das allgemeine Notenspektrum pro Station variiert stark, nur bei der Totenscheinstation und bei der Blutentnahme wurde der Range (fast) komplett ausgereizt (siehe Tabelle 5. Übersicht der OSCE-Leistung der Teilnehmer:innen), gefolgt von der Schwindelstation. Es wurde auf allen Stationen die höchste Punktzahl mindestens einmal vergeben, die überwiegend meisten Ergebnisse bewegten sich beginnend in den besseren zwei Dritteln oder der besseren Hälfte der Punktskala: es fand sich also ein leichter Deckeneffekt der Notenvergabe. Die idealen Streuungsergebnisse bei Prüfungen aus einer Testkonstruktionsperspektive lassen sich umfangreich diskutieren (siehe zum Beispiel Eid & Schmidt, 2014). Streut die Variable (also hier die Benotungskriterien auf der jeweiligen Station) schlecht, wenn der Punktrange nicht ausgeschöpft wird – ein Hinweis darauf, dass die Rater:innen die Bewertungskriterien oder eigentliche Aufgabenstellung der Station schlecht konzipiert haben? Oder liegt es vielmehr an der Qualität der Lehre, oder der Vorerfahrung der Studierenden, die einfach in manchen Fächern unabhängig von der Testung schlechter oder im Fall dieser Arbeit besser sind, dass die Streuung gering ist? Und was ist wünschenswert? Das zur typischen Approbationsleistung sehr ähnliche Gesamtergebnis könnte ein Hinweis darauf sein, dass die leistungsstarken, frisch aus dem Studium kommenden Teilnehmer:innen der Testung einfach gut sind, in dem was sie tun – selbst bei einem unvorbereiteten Wissens- und Fertigkeitstest.

Das vergleichsweise niedrige Durchschnittsergebnis auf der neurologischen Schwindelstation könnte auf einen anderen Grund zurückzuführen sein, der von den oben genannten Einflüssen entkoppelt scheint. Die dort für den 3. Testzeitpunkt eingesetzte Prüferin war entgegen ihrer Instruktion durch die Prüfungsorganisation unangemessen anspruchsvoll in ihren Erwartungen

an die Leistung der Teilnehmer:innen und verfügte nur über geringe Erfahrung mit ihrer Rolle und bewertete in der Folge sehr streng. Bei den Erhebungen an der gleichen Station, zu den anderen Testzeitpunkten und mit anderen Prüfer:innen, erzielten die Teilnehmer:innen im Durchschnitt deutlich bessere Ergebnisse, vergleichbar zu den Ergebnissen in den anderen Stationen. Die Bewertungskriterien konnten diese individuelle Abweichung leider nicht soweit kompensieren, dass die Ergebnisse davon unbeeinflusst geblieben wären.

6.3.1. Wunschfacharztweiterbildung

Die Befragung nach der angestrebten Wunschfacharztweiterbildung ergab die Nennung erstaunlich weniger Fächer insgesamt. Ein überraschendes Ergebnis: denn je nach Definition stehen derzeit fast 50 verschiedene Weiterbildungsrichtungen zur Auswahl. Das könnte in der Unkenntnis der Studierenden mit den „Orchideenfächern“ ihrer Profession begründet liegen, denn auch im Studium wird das besonders oft gelehrt, was besonders oft vorkommt. Es ist davon auszugehen, dass die spätere Zuordnung der Absolvent:innen mehr Varianz aufweisen könnte. Auch steht in Frage, ob die Nennung eines Wunschfaches auch automatisch impliziert, dass die befragte Person tatsächlich dort landet: schließlich spielen viele auch pragmatische Gründe von Verfügbarkeit bis örtlicher Anbindung eine Rolle bei der endgültigen Wahl (van den Bussche et al., 2017).

Passend zu den von van den Bussche et al. in diversen Arbeiten gefundenen Ergebnissen strebten auch unsere Teilnehmer:innen am häufigsten Innere Medizin und Anästhesiologie als Wunschfacharztweiterbildungsrichtungen an. Ein Hauptgrund scheint dafür die erhöhte Familienvereinbarkeit dieser Fachrichtungen zu sein. Damit würde die Untersuchungsstichprobe den vermeintlichen Wertewandel der jüngeren Ärzt:innengenerationen gut repräsentieren ("Zum Nachwuchsmangel in Unfallchirurgie und Orthopädie", 2010). Allerdings könnten sich Teilnehmer:innen mit solchen Interessen von den Vertreter:innen der Allgemeinmedizin, die sie rekrutiert haben, auch schlicht am ehesten angesprochen gefühlt haben - ein weiterer möglicher Selektionseffekt.

Auf der Leistungsebene ist festzuhalten, dass sich die Teilnehmer:innen der verschiedenen Wunschfachzuordnungen signifikant in der Performanz unterschieden (allerdings nur bei der Analyse auf Schulnotenniveau): bei 12 von 160 Punkten, was keinen Notensprung ausmacht. Deshalb ist nicht von einer besonderen Relevanz dieses Ergebnisses auszugehen. Größere Stichprobengrößen wären nötig, um diese Unterschiede belastbarer untersuchen zu können.

6.4. Psychometrische und soziodemografische Ergebnisse

6.4.1. Persönlichkeitsmessung

Wie eingangs bereits angeschnitten, war die Varianz in der Ausprägung der Persönlichkeitsattribute der Studienteilnehmer:innen überraschenderweise sehr gering: alle Teilnehmer:innen wiesen vergleichsweise niedrige Neurotizismus- und Verträglichkeitswerte, aber hohe Gewissenhaftigkeits-, Offenheits- und Extraversionswerte auf - insbesondere, wenn man diese Werte mit den Normwerten der deutschen Normalpopulation vergleicht (Körner et al., 2008). Man könnte von einem „Normprofil“ der untersuchten Personen sprechen, da sie, alle eine sehr ähnliche Zusammensetzung an Persönlichkeitsschwerpunkten und Verläufen aufweisen (siehe Abbildung 7: NEO-FFI 30-Ergebnisse der Teilnehmer:innen, ($N=97$, $N=99$)). In Bezug auf die recht hoch ausgeprägte Skala Extraversion ist davon auszugehen, dass die Stichprobe einem Selektionseffekt unterliegt: schüchterne (introvertierte) Personen sind vermutlich naturgemäß weniger motiviert an der Exposition durch ein OSCE-Verfahren, das in diesem Fall im wahrsten Sinne ein Stresstest ist, und werden mit geringerer Wahrscheinlichkeit freiwillig teilnehmen. Zudem scheinen höhere Extraversionswerte hilfreich in einem Berufsfeld, in dem Patientenkontakt die Regel darstellt. Es ist vorstellbar, dass die Selektion bereits vor Studienbeginn stattfindet, weil sich Personen mit niedrigeren Extraversionswerten eher für andere Berufsrichtungen mit weniger gezwungenen Sozialkontakten entscheiden.

Ähnlich begründbar erscheint die Häufung niedrigerer Verträglichkeitswerte – Personen mit geringeren Ausprägungen dieses Attributes legen weniger Wert auf die Sympathie und Meinung anderer und sind eher bereit, sich für eigene Interessen einzusetzen oder abzugrenzen. Auch hier erscheint es naheliegend, dass Teilnehmer:innen mit dieser Eigenschaft eher bereit sind, an einem Testverfahren wie dem Vergleichs-OSCE mitzuwirken, weil sie der Meinung anderer (Prüfer:innen und Kommiliton:innen) potenziell weniger Bedeutung beimessen. Die niedrigeren Verträglichkeitswerte unserer Stichprobe sind also kein „schlechtes“ Attribut an Ärzt:innen, da eine hohe Durchsetzungsfähigkeit eine:n abgegrenztere:n Mediziner:in mit höherer psychischer Stabilität ergibt. Ein weiterer Hinweis auf die Legitimität dieser Ergebnisse ist die stimmige Eingliederung in existierende Forschungsbefunde: (Harendza et al., 2019) konnten zum Beispiel belegen, dass Medizinstudent:innen unter anderem schlechteren Leistungen in Bezug auf Teamarbeit, aber bessere bei Adhärenz zu Prozeduren zeigten, als die mit ihnen verglichenen streng selektierten Bewerber:innen für Fluglotsenausbildungsplätze. Auch bei diesem Instrument fand sich ein statistisch signifikanter Geschlechtsunterschied: männliche Teilnehmer erzielten geringere Ergebnisse bei Neurotizismus und Gewissenhaftigkeit, waren dafür aber verträglicher. Damit konnten wir die üblicherweise berichteten Geschlechtsunterschiede in der Literatur replizieren (Weisberg et al., 2011), ein Hinweis auf die Güte des Instrumentes und der Testdurchführung.

Das Persönlichkeitsprofil der Teilnehmer:innen scheint, wie im Stand der Forschung bereits kurz angeschnitten, besonders vorteilhaft für den Erfolg der medizinischen Ausbildung. Liegt auch hier ein potenzieller Selektionseffekt vor? Laut Sobowale et al. (2018) scheinen hohe Gewissenhaftigkeitswerte und niedrige Neurotizismusausprägungen von Vorteil beim Absolvieren der akademischen Ausbildung zu sein. Entweder könnten Studierende mit abweichenden Attributen weniger oder gar nicht erfolgreich im Studium gewesen sein und es nicht an den Punkt schaffen können, oder nur die (durch die „richtige“ Persönlichkeit) leistungsstarken nehmen an der Testung teil, weil sie nicht mit Misserfolg rechnen müssen. Dass Sorgfalt, Fleiß und geringere Besorgtheit oder Ängstlichkeit zu besseren Studienleistungen führen, scheint naheliegend.

Eine wichtige Einschränkung bei der Interpretation der Persönlichkeitsdaten muss allerdings benannt werden: in der Fachliteratur wird von Korrelationen der Persönlichkeitsausprägungen mit Bildungsstand und Alter berichtet (Wrzus et al., 2022). Da unsere Stichprobe in Bezug auf beides sehr homogen ist, überrascht vor diesem Hintergrund das Ergebnis wenig, da eine Replikation erwartbar ist. In Bezug auf die Interpretation der Daten hinsichtlich einer möglichen Berufseignung sollte darüber hinaus mit Vorsicht vorgegangen werden. Gleichwohl der NEO-FFI als der Goldstandard in der Persönlichkeitsmessung zu betrachten ist, wird von den Testautoren zu Schlüssen für die Berufseignung abgeraten. Was die Studieneignung oder den Studienerfolg anbetrifft, könnte man allerdings angesichts der hier erläuterten Daten Hinweise auf Zusammenhänge annehmen.

Eine Frage wirft die Homogenität der Persönlichkeitsprofile aller Teilnehmer:innen darüber hinaus auf: kann es sein, dass die Persönlichkeit der Studierenden die Performance beim OSCE so beeinflusst, dass die Unterschiede zwischen den Studiengängen gar nicht zum Tragen kommen?

6.4.2. Lebenszufriedenheit, soziale Unterstützung und Angst und Depression

Da die Ergebnisse der psychometrischen Instrumente alle dem gleichen Muster folgen, werden sie im Folgenden gemeinsam diskutiert. Extreme Mittelwerte und die kaum existente Varianz bei den Ergebnissen der drei Testverfahren auf Ausprägungen von Lebenszufriedenheit (SWLS), Angst und Depression (HADS-D), und sozialer Unterstützung (F-SoZu K14) machten prädiktive Werte sinnlos – es gibt keine schlicht „kranken“ Teilnehmer:innen. Eine Aufschlüsselung der Ergebnisse nach Absolvent:innen verschiedener Studiengangsformen oder demografischen Parametern ist durch die auffällige Homogenität der Daten und unterschiedlichen Gruppengrößen ebenfalls statistisch nicht praktikabel. Dennoch ist die Diskussion der psychometrischen Daten aufgrund ihrer recht extremen positiven Natur sinnvoll.

Wie bereits erwähnt, erzielten drei Viertel der Teilnehmer:innen in Bezug auf die Untersuchung der Lebenszufriedenheit die beiden höchsten Kategorien, während die untersten drei Kategorien gar nicht vergeben wurden, also kein einziger der Teilnehmer:innen „unzufrieden“ war. Im Vergleich mit der deutschen Normierung lässt sich festhalten, dass die Stichprobe über eine wirklich außergewöhnlich hohe Lebenszufriedenheit berichtet (Glaesmer et al., 2011; Fydrich et al., 2009). Dafür gibt es einige mögliche Erklärungen: denkbar ist es zum Beispiel, dass die Reihung der Fragebogenuntersuchung im Anschluss an die praktische Vergleichs-OSCE Testung zu Verzerrungen geführt haben könnte. Die Teilnehmer:innen könnten durch die absolvierte (erfolgreiche) Testung erleichtert und positiv gestimmt gewesen sein und dadurch optimistischere Einschätzungen zu ihrer Lebenszufriedenheit abgegeben haben, also einem Primingeffekt unterlegen sein. Auch ist davon auszugehen, dass solche Personen, die besonders unzufrieden und dadurch belastet sind, sich vermutlich die freiwillige Teilnahme an einem stressreichen Testverfahren eher weniger zumuten und daran nicht teilnehmen.

Darüber hinaus ist nach genauerer Prüfung der für den SWLS verwendeten Items ein möglicher Einfluss des Alters nicht unwahrscheinlich: Alter könnte bei Items wie *„Bisher habe ich die wesentlichen Dinge erreicht, die ich mir für mein Leben wünsche“* oder *„Wenn ich mein Leben noch einmal leben könnte, würde ich kaum etwas ändern.“* eine größere Rolle für das Antwortverhalten spielen, und unsere Stichprobe war mit einem Durchschnittsalter von knapp 29 Jahren recht jung.

Eine Rekapitulation der Ergebnisse für die Untersuchung auf Angst und Depression mittels des HADS-D zeigt ein ähnliches Muster: Fast alle (96 Prozent) Teilnehmerinnen erzielten unauffällige Depressionswerte, ohne die Skala auszuschöpfen. Bei der Angstskala berichteten analog fast 90 Prozent der Befragten von keiner Belastung. Die Stichprobe liegt im Mittel also ebenfalls deutlich unter den Grenzwerten für Auffälligkeiten und sogar die von Herrmann et al. berichteten Normwerte für Angst und Depression in der gesunden Normalpopulation und bei Medizinstudierenden unterschreitet. Dies passt zu den geringen Neurotizismuswerten, die die Teilnehmer:innen bei der Persönlichkeitsmessung zeigten. Und auch hier könnten die Testungsreihenfolge und Selbstselektionseffekte eine Rolle in der Ausprägung der Ergebnisse gespielt haben.

Die Hospital Anxiety and Depression Scale, ist wie der Name schon sagt, ursprünglich zur einfachen Anwendung in klinischen Settings konzipiert worden – daher ist die Anwendung im Rahmen dieser Studie nicht die ursprünglich intendierte. Allerdings ist die Anwendung in anderen Kontexten vielfach erprobt und für hinreichend zu Screeningzwecken befunden worden (Snaith, 2003). Da unsere Teilnehmer:innen aber ohnehin einen Deckeneffekt in Bezug auf die Ausprägung ihrer Werte aufweisen, scheint die vermeintliche Einschränkung der Aussagekraft des Testverfahrens noch weniger bedenklich. Besonders eindrücklich sind die Ergebnisse, da sie sich fundamental von diversen internationalen Befunden unterscheiden, nach

denen die mentale Gesundheit von untersuchten Medizinstudent:innen grundsätzlich geringer ist als die vergleichbarer nicht-medizinisch tätiger Personen (Dyrbye et al., 2014, Van Remortel et al., 2018), und höhere Depressionsraten bis hin zu Suizidalität leider übliche Eigenschaften dieser Gruppe zu sein scheinen (Rotenstein et al., 2016). Erfreulicherweise konnten diese Ergebnisse im Rahmen dieser Arbeit nicht repliziert werden.

Schlussendlich ergibt sich in der Betrachtung der Ergebnisse für die Testung der sozialen Unterstützung der Teilnehmer:innen mittels des F-SoZu K14 wiederum die gleiche Tendenz, wie bei den anderen psychometrischen Skalen, mit entsprechend ähnlichen Erklärungsansätzen: Das Mittel der Gesamtstichprobe entspricht einer normalen sozialen Unterstützung an der Grenze zu überdurchschnittlich hoher sozialer Unterstützung. Die deutsche Normierung kommt im Vergleich auf einen deutlich niedrigeren Wert in der Durchschnittsbevölkerung (Fydriich et al., 2009). Hier berichten die Testautoren allerdings schon selbst von einer durch das Instrument an sich grundsätzlich höher eingeschätzten sozialen Unterstützung bei Personen mit geringerer psychischer Belastung, einer Eigenschaft, die nach den restlichen Daten wohl auf die Stichprobe zutrifft.

Ein weiterer interessanter Aspekt in Bezug auf die hohe soziale Unterstützung dieser Stichprobe: Medizinstudent:innen brechen ihr Studium sehr viel seltener ab, als die anderer Fachdisziplinen (Tolciu & Sode, 2011). Ist das Ergebnis hoher Zugangshürden und damit *sunk-cost effect*⁸ oder vielleicht auch Ausdruck besserer sozialer Unterstützung?

6.4.3. Soziodemografie

Gerade in Hinblick auf die besondere Homogenität der Ergebnisse lohnt sich eine besondere Betrachtung der soziodemografischen Eigenschaften der Stichprobe, die dem Homogenitätsmuster aller Daten folgt (oder, wenn man so will, voransteht). Die Teilnehmer:innen ähneln sich stark: sie sind jung, ledig, und verfügen weit überwiegend über keinen Migrationshintergrund.

Bezüglich des Vergleichs der beiden Absolvent:innengruppen ist dieses homogene Ergebnis als sehr erfreulich zu werten, da die Vergleichbarkeit dadurch umso höher ist. Gleichzeitig bestätigt es Arbeiten, die zu Bildungsungleichheit in Deutschland bezugnehmen und die geringe Varianz im höheren akademischen Sektor bemängeln (Pollak, 2010). Die demografische Analyse zeigt einen deutlich niedrigeren Anteil an Migrationshintergründen als im deutschen

⁸ (= S.) [engl. Versunkene Kosten], Steigerung des Einsatzes, irrationale Beharrlichkeit. Laut. Arkes und Blumer (1985) die Tendenz, ein Vorhaben (z. B. ein Projekt, eine Investition, eine Beziehung) fortzusetzen, wenn bereits eine Investition in Form von Geld, Anstrengung (Energie) oder Zeit getätigt wurde, also versunkene Kosten entstanden sind. (Kirchler & Stark 2019)

Durchschnitt (Kutzner, 2009). Auch sind Selektionseffekte angesichts dieser Homogenität wahrscheinlich.

6.5. Schlussfolgerung

Trotz einiger erwartbarer methodischer Herausforderungen liegt mit dieser Arbeit ein gelungener Vergleich der beiden Hamburger Medizincurricula vor. Insbesondere die Konzeption und Durchführung des praktischen Fertigkeitstest, des Vergleichs-OSCEs, liefert wichtige Anhaltspunkte und ein verlässliches Vorbild für erfolgreiche Praxisüberprüfungen der akademischen Leistungen von Medizinstudierenden.

Gleichzeitig konnte gezeigt werden, dass der Bedeutung des Arztreifebegriffes mehr Aufmerksamkeit geschenkt werden muss. Es ist denkbar, dass der für diese Arbeit verwendete Arztreifebegriff trotz sorgfältiger Untermauerung Unzulänglichkeiten aufwies, und die Inhaltsvalidität nicht in vollem Umfang sichergestellt war.

Die psychometrischen Ergebnisse reihen sich gut in existierende Forschungsbefunde ein und unterstreichen die Güte und Relevanz der Arbeit. Was die Studieneignung oder den Studien-erfolg von angehenden Mediziner:innen anbetrifft, könnte man Hinweise auf Zusammenhänge zur Persönlichkeit annehmen. Eine weitere empirische Exploration dieser Zusammenhänge, auch in anderen medizinnahen Disziplinen, scheint vielversprechend, um den zukünftigen Herausforderungen des deutschen Gesundheitssystem und allgemein des medizinischen Berufsfelds bessere Personalauswahlkriterien erarbeiten zu können.

Eine Frage wirft die Homogenität der Persönlichkeitsprofile aller Teilnehmer:innen darüber hinaus auf: kann es sein, dass die Persönlichkeit der Studierenden die Performance beim OSCE so beeinflusst, dass die Unterschiede zwischen den Studiengängen gar nicht zum Tragen kommen?

Eine reizvolle Überlegung im Sinne einer Verlaufskontrolle wäre die erneute Durchführung eines OSCEs mit den aktuellsten Medizinstudiumsabsolvent:innen, sowohl in Hamburg, als auch an anderen Standorten, wobei die Pandemiesituation der letzten Jahre sicherlich auch als Einflussgröße auf die Qualität der Lehre nicht zu unterschätzen ist. Dabei sollte der Vermeidung von Prüfereffekten besondere Beachtung geschenkt werden.

7. Zusammenfassung

Hintergrund: Am Universitätsklinikum Hamburg-Eppendorf löste 2012 der integrierte Modellstudiengang Medizin (iMED) den Regelstudiengang (RSG) ab. Basierend auf der neuen Approbationsordnung, orientiert er sich als einer der ersten deutschlandweit in der Vermittlung der medizinischen Lehrinhalte an modernsten Konzepten für die Ausbildung von Ärzt:innen mit dem Ziel, bereits zu Studienbeginn die theoretischen Grundlagenfächer mit der klinischen Praxis zu verschränken. Insbesondere wurde ein Schwerpunkt auf die integrierte Vermittlung von kommunikativen und praktischen Fertigkeiten gelegt.

Fragestellung: Ziel der Untersuchung war die Analyse der durch den absolvierten Studiengangstyp und die Persönlichkeit bedingten Unterschiede in der Performanz von Studierenden bei einem klinisch-praktischen OSCE (*objective structured clinical examination*), mittels dem die Allgemeine Ärztliche Reife erfasst wurde.

Methode: In den Jahren 2016 - 2018 wurden freiwillige Absolvent:innen aus beiden Studiengangsgruppen (RSG vs. iMED) zu 4 querschnittlichen Testzeitpunkten für die Teilnahme an einem selbstkonzipierten Vergleichs-OSCE mit anschließender Fragebogenuntersuchung rekrutiert. Es wurden die Allgemeine Ärztliche Reife, das Persönlichkeitsprofil (NEO-FFI 30), die Lebenszufriedenheit (*Satisfaction with Life Scale* (SWLS)), soziale Unterstützung (F-SoZu K14), Angst- und Depressionswerte (*Hospital Anxiety and Depression Scale* (HADS-D)), Soziodemografie und angestrebte Facharztweiterbildungsrichtung erfasst.

Ergebnisse: Im Durchschnitt haben die Teilnehmer:innen (N = 100, 50 % weiblich) eine befriedigende Leistung (Schulnote 3) im Test zur Arztreife gezeigt, unabhängig ihrer Studiengangszugehörigkeit. Dabei erzielte die Gesamtstichprobe außergewöhnlich positive psychometrische Werte und zeigte ein für alle geltendes „Persönlichkeitsnormprofil“ mit ähnlichen Ausprägungen.

Schlussfolgerung: Die Umstellung des Hamburger Medizincurriculums hatte nicht zu Folge, dass sich die Leistungen der untersuchten Absolvent:innen beider Studiengänge maßgeblich in der Ärztlichen Reife unterschieden. Eine vertiefende Untersuchung erscheint sinnvoll, gerade in Hinblick auf das gefundene Persönlichkeitsmuster unter den Teilnehmer:innen.

8. Erklärung des Eigenanteils an dieser Arbeit

Folgende Vorarbeiten wurden von mir übernommen: die gesamte Fragebogenkonzeption und dessen Überarbeitung geschah durch mich in Abstimmung mit Prof. Dr. Scherer. Ich war verantwortlich für die Anfrage und Einholung der Freigabe der Ethikkommission für das gesamte Forschungsvorhaben. Genauso lag die Erstellung von Teilnehmer:inneninformationen und Einverständniserklärungen für alle Testungsteile, sowie große Teile der allgemeinen Materialbereitstellung bei mir.

Des Weiteren geschah die inhaltliche Konzeption der Simulationspatient:innenstation und die Schulung und Ausbildung der eingesetzten Simulationspatient:innen, sowie die administrative Betreuung dieser durch mich, in Zusammenarbeit mit dem Simulationspatient:innenprogramm und den psychosomatischen Fachvertreter:innen des UKE. Zusätzlich war ich mitverantwortlich für die Organisation und administrative Betreuung von Untersuchungsmodellen für die Anamnesestation.

Bei allen vier Testzeitpunkten des Vergleichs-OSCEs von 2016 bis 2018 war ich für die Organisation und Vor-Ort-Betreuung der Testung und eingesetzten Prüfenden, Simulationspatient:innen, Untersuchungsprobanden und Testteilnehmer:innen zusammen mit Katharina Dahl, Maren Ehrhardt und Kerstin Schrom verantwortlich.

Dabei lag die Verantwortung für die Durchführung und Betreuung der an die praktische Testung anschließenden Fragebogenerhebung allein bei mir und geschah ohne weitere personelle Unterstützung im Sinne einer Testleitung.

Schlussendlich war ich alleinverantwortlich für die Datensammlung, -Ablage, -Dokumentation und -Eingabe von sowohl der Vergleichs-OSCE-Testung als auch der Fragebogenerhebung für alle 4 Erhebungszeitpunkte. Die Datenbereinigung, -Zusammenführung, und die komplette statistische Analyse (letzteres betreut von Dr. Sigrid Boczor), sowie abschließend die Aufbereitung der Daten beider Studienteile für alle 4 Testzeitpunkte, geschah ausschließlich durch mich.

Zu guter Letzt war ich für die Literaturrecherche und Konzeption, sowie eigentliche Erstellung dieser Arbeit und die wiederholte Überarbeitung nach Rückmeldungen externer Begutachter:innen allein verantwortlich.

9. Literaturverzeichnis

- Asendorpf, J. B., & Neyer, F. J. (2012). *Psychologie der Persönlichkeit*. Springer-Verlag.
- Ainsworth, M. (1978). The Bowlby-Ainsworth attachment theory. *Behavioral and Brain Sciences*, 1(3), 436-438.
- Antonovsky, A. (1974). Conceptual and methodological problems in the study of resistance resources and stressful life events. In B. S. Dohrenwend & B. P. Dohrenwend (Hrsg.), *Stressful life events: Their nature and effects*. John Wiley & Sons.
- Barbosa-Camacho, F. J., Miranda-Ackerman, R. C., Vazquez-Reyna, I., Jimenez-Ley, V. B., Barrera-Lopez, F. J., Contreras-Cordero, V. S., Gonzalez-Ojeda, A. (2020). Association between HEXACO personality traits and medical specialty preferences in Mexican medical students: a cross-sectional survey. *BMC Psychology*, 8(1), 23.
- Bergold, S., & Steinmayr, R. (2018). Personality and Intelligence Interact in the Prediction of Academic Achievement. *Journal of Intelligence*, 6(2), 27.
- Borkenau, P., & Ostendorf, F. (1993). NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae: Handanweisung.
- Briel, M., Olu, K.K., Von Elm, E., Kasenda, B., Alturki, R., Agarwal, A., Bhatnagar, N., Schandelmaier, S., (2016). A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *Journal of Clinical Epidemiology* 80, 8–15.
- Brown, G. W. & Harris, T. O. (1978). *The social origins of depression: A study of psychiatric disorders in women*. London: Tavistock.
- Bundesärztekammer (2018). Musterweiterbildungsordnung 2018. www.bundesaerztekammer.de/themen/aerzte/aus-weiter-fortbildung/weiterbildung/muster-weiterbildungsordnung/ (letzter Zugriff 07.09.2022)
- Bussenius, L., & Harendza, S. (2019). The relationship between perfectionism and symptoms of depression in medical school applicants. *BMC Medical Education*, 19(1), 370.
- Callaghan, P., & Morrissey, J. (1993). Social support and health: a review. *Journal of Advanced Nursing*, 18(2), 203-210.
- Cleland, J. A. (2009). The use of simulated patients in medical education: AMEE Guide No 42. *Medical Teacher* 31(6): 477-486.
- Cooke, P. J., Melchert, T. P., & Connor, K. (2016). Measuring Well-Being. *The Counseling Psychologist*, 44(5), 730-757.
- Costa, P. T., & McCrae, R. R. (1989). The NEO PI/FFI manual supplement. Odessa: Psychological Assessment Resources.
- Diener, E. (1985). The satisfaction with life scale. *Journal of personality assessment* 49(1): 71-75.

- Diener, E., Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being*, 3, 1–43.
- Dyrbye LN, West CP, Satele D, Boone S, Tan L, Sloan J, Shanafelt TD. (2014). Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Academic Medicine*, Mar;89(3):443-51.
- Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Hogrefe Verlag.
- Eisenberg, D., Golberstein, E., & Hunt, J. B. (2009). Mental health and academic success in college. *The BE Journal of Economic Analysis & Policy*, 9(1).
- Epstein, R. M., & Krasner M. S. (2013). Physician resilience: what it means, why it matters, and how to promote it. *Academic Medicine*, Mar;88(3):301-3.
- Ferguson E., James D., Madeley L. (2002). Factors associated with success in medical school: systematic review of the literature. *The BMJ*. Apr 20;324(7343):952-7.
- Fishbein, D., Nambiar, S., Mckenzie, K., Mayorga, M., Miller, K., Tran, K., Schubel, L., Agor, J., Kim, T., Capan, M., (2019). Objective measures of workload in healthcare: a narrative review. *International Journal of Health Care Quality Assurance* 33, 1–17.
- Fydrich, T., Geyer, M., Hessel, A., Sommer, G. und Brähler, E. (1999). Fragebogen zur sozialen Unterstützung (F-SozU): Normierung an einer repräsentativen Stichprobe. *Diagnostica*, 45, 212-216.
- Fydrich, T., Sommer, G., Tydecks, S., & Brähler, E. (2009). Fragebogen zur sozialen Unterstützung (F-SozU): Normierung der Kurzform (K-14). *Zeitschrift für Medizinische Psychologie*, 18(1), 43-48.
- Geißler, R. (2006). Bildungschancen und soziale Herkunft. *Archiv für Wissenschaft und Praxis der sozialen Arbeit* 4(2006): 34-49.
- Glaesmer, H., Grande, G., Braehler, E., & Roth, M. (2011). The German version of the Satisfaction with Life Scale (SWLS): Psychometric properties, validity, and population-based norms. *European Journal of Psychological Assessment*, 27(2), 127–132.
- Guilera, T., Batalla, I., Forne, C., & Soler-Gonzalez, J. (2019). Empathy and big five personality model in medical students and its relationship to gender and specialty preference: a cross-sectional study. *BMC Medical Education*, 19(1), 57.
- Graf, J., Smolka, R., Simoes, E., Zipfel, S., Junne, F., Holderried, F., Herrmann-Werner, A.. (2017). Communication skills of medical students during the OSCE: Gender-specific differences in a longitudinal trend study. *BMC Medical Education*, 17(1).
- Grochowska, A., Gawron, A., Bodys-Cupak, I., (2022). Stress-Inducing Factors vs. the Risk of Occupational Burnout in the Work of Nurses and Paramedics. *International Journal of Environmental Research and Public Health* 19, 5539.
- Güntert, A. (2003). *Approbationsordnung für Ärzte (ÄAppO), Bundesärzteordnung (BÄO): mit Erläuterungen und praktischen Hinweisen*, Deutscher Ärzteverlag.

- Guilera, T., Batalla, I., Forne, C., & Soler-Gonzalez, J. (2019). Empathy and big five personality model in medical students and its relationship to gender and specialty preference: a cross-sectional study. *BMC Medical Education*, 19(1), 57.
- Harden, R. M. (1988). What is an OSCE?. *Medical Teacher*, 10(1): 19-22.
- Harden, R. M., Lilley, P. (2018). *The eight roles of the medical teacher: the purpose and function of a teacher in the healthcare professions*. London: Elsevier Health Sciences.
- Harandi, T. F., Taghinasab, M. M., Nayeri, T. D. (2017). The correlation of social support with mental health: A meta-analysis. *Electronic Physician*. Sep 25;9(9):5212-5222.
- Harendza, S., Soll, H., Prediger, S., Kadmon, M., Berberat, P. O., & Oubaid, V. (2019). Assessing core competences of medical students with a test for flight school applicants. *BMC Medical Education*, 19(1), 9.
- Hamburger Lernzielkatalog, (2009).
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjJ55uj-oL6AhVgSPEDHQuYCSqQFnoECA8QAQ&url=https%3A%2F%2Fwww.uke.de%2Fdateien%2Fzielgruppen%2Fstudium-lehre%2Fdokumente%2Flernzielkatalog_091104_mat.pdf&usg=AOvVaw0yi6EehRcilwb9lqGKYIYK, (abgerufen am 7.9.2022)
- Hayat, A. A., Kohoulat, N., Amini, M., & Faghihi, S. A. A. (2020). The predictive role of personality traits on academic performance of medical students: The mediating role of self-efficacy. *Medical Journal of the Islamic Republic of Iran*, 34, 77.
- Hermann-Lingen, C. (Hrsg.). (2011). *HADS-D: Hospital Anxiety and Depression Scale - German Version*. Bern: Huber.
- Hoffmeyer-Zlotnik, J. H. (Hrsg.). (2010). Demographische Standards Ausgabe 2010: eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute eV, der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute eV (ASI) und des Statistischen Bundesamtes, Statistisches Bundesamt.
- Jafrani, S., Zehra, N., Zehra, M., Abuzar Ali, S. M., Abubakar Mohsin, S. A., & Azhar, R. (2017). Assessment of personality type and medical specialty choice among medical students from Karachi; using Myers-Briggs Type Indicator (MBTI) tool. *Journal of Pakistan Medical Association*, 67(4), 520-526.
- Jünger, J. (2018). Kompetenzorientiert prüfen im Staatsexamen Medizin. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 61(2), 171-177.
- Kaufmann, H. (Hrsg.). (1986). *Strabismus*. Stuttgart.
- E., & Stark, J. (2019, July 01). sunk-cost effect. In M. A. Wirtz (Hrsg.): *Dorsch Lexikon der Psychologie*. Bern: Hogrefe. Retrieved from:
<https://dorsch.hogrefe.com/stichwort/sunk-cost-effect>
- Kuhnigk, O., Weidtmann, K., Dietsche, S., Guse, A. H., Mihalache, I., Schultz, J., Harendza, S. (2006). Vom "vorklinischen Studienabschnitt" zu "Medizin I": Umstellung auf die neue ÄAppO an der Medizinischen Fakultät der Universität Hamburg. *GMS Zeitschrift für Medizinische Ausbildung*. 2006;23(2):Doc36.

- Kusier, A. O., & Folker, A. P. (2021). The Satisfaction with Life Scale: Philosophical Foundation and Practical Limitations. *Health Care Analysis*, 29(1), 21-38.
- Kutzner, C. (2009). Die demografische Entwicklung in Deutschland. In *Demografischer Wandel und Kultur*. (pp. 15-33). Verlag für Sozialwissenschaften.
- Lakey, B. & Cassady, P. B. (1990). Cognitive processes in perceived social support. *Journal of Personality and Social Psychology*, 59, 337- 343.
- Luhmann, M., Lucas, R. E., Eid, M., & Diener, E. (2013). The Prospective Effect of Life Satisfaction on Life Events. *Social Psychological and Personality Science*, 4(1), 39-45.
- Mayer, J. D., & Skimmyhorn, W. (2017). Personality attributes that predict cadet performance at West Point. *Journal of Research in Personality*, 66, 14-26.
- Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, 55, 1–22.
- Nawaiseh, M. B., Haddadin, R. R., Al Droubi, B., Nawaiseh, H. B., Alarood, S., Aborajoo, E., Abu-Yaghi, N. E. (2020). The Association Between Personality Traits and Specialty Preference Among Medical Students in Jordan. *Psychology Research and Behavior Management*, 13, 599-607.
- Neugarten, B. L. (1961). The measurement of life satisfaction. *Journal of gerontology*.
- O'Connor, M.C.; Paunonen, S.V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43, 971–990.
- Oishi, S., Schimmack, U., & Colcombe, S. J. (2003). The contextual and systematic nature of life satisfaction judgments. *Journal of Experimental Social Psychology*, 39, 232–247.
- Oishi, S., Diener, E., & Lucas, R. E. (2007). The Optimum Level of Well-Being: Can People Be Too Happy? *Perspectives on Psychological Science*, 2(4), 346-360.
- Paiva, C. E., et al. (2018). Doctor, are you healthy? A cross-sectional investigation of oncologist burnout, depression, and anxiety and an investigation of their associated factors. *BMC Cancer* 18(1): 1-11.
- Pavot, W., & Diener, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social indicators research*, 28(1): 1-20.
- Pavot, W., & Diener E. (2013). Happiness experienced: The science of subjective well-being. *The Oxford handbook of happiness*, 134-151.
- Petermann, F. (2015). Hospital Anxiety and Depression Scale, deutsche Version (HADS-D). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*.
- Pollak, R. (2010). Kaum Bewegung, viel Ungleichheit: Eine Studie zu sozialem Auf-und Abstieg in Deutschland, *Schriften zu Wirtschaft und Soziales*.
- Raupach, T., Brown, J., Anders, S., Hasenfuss, G., & Harendza, S. (2013). Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. *BMC Medicine*, 11(1), 61.

- Rotenstein LS, Ramos MA, Torre M, (2016). Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *JAMA*. 2016;316(21):2214-2236
- Scherer, M., Ehrhardt, M. & Lühmann, D. (2014). *Entwicklung eines Prüfungs- und Messinstruments zur Outcome-bezogenen Evaluation des KUMplusKOM Lehrkonzepts im i-med Modellstudiengang –Folgeantrag*. FFL2 Antrag, Institut und Poliklinik für Allgemeinmedizin, Universitätsklinikum Hamburg-Eppendorf (nicht öffentlich zugänglich).
- Schimmack, U., Diener, E., & Oishi, S. (2002). Life-Satisfaction Is a Momentary Judgment and a Stable Personality Characteristic: The Use of Chronically Accessible and Stable Sources. *Journal of Personality*, 70(3), 345-384.
- Schumacher, J. (2003). SWLS–Satisfaction with life scale. *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden*.
- Sobowale, K., Ham, S. A., Curlin, F. A., & Yoon, J. D. (2018). Personality Traits Are Associated with Academic Achievement in Medical School: A Nationally Representative Study. *Academic Psychiatry*, 42(3), 338-345.
- Sommer, G. und Fydrich, T. (1991). Entwicklung und Überprüfung eines Fragebogens zur sozialen Unterstützung. *Diagnostica*, 37, 160-178.
- Stadtfeld, C., Vörös, A., Elmer, T., Boda, Z., & Raabe, I. J. (2019). Integration in emerging social networks explains academic failure and success. *Proceedings of the National Academy of Sciences*, 116(3), 792-797.
- Steiner-Hofbauer V, Schrank B, Holzinger A. (2018). What is a good doctor? *Wiener Medizinische Wochenschrift*. Nov;168(15-16):398-405.
- Thoits, P. A. (1985). Social support and psychological well-being: Theoretical possibilities. In *Social support: Theory, research and applications* (pp. 51-72). Springer, Dordrecht.
- Tolciu, A., & Sode, M. (2011). *Mehr Studienanfänger: Mehr Studienabbrecher?* (No. 61). HWWI policy paper.
- Tomasits, J., & Haber, P. (2016). Frauen betreiben Sport. In: *Leistungsphysiologie*. Springer, Berlin, Heidelberg.
- Uchino, B. N., Cacioppo, J. T., & Kiecolt-Glaser, J. K. (1996). The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health. *Psychological Bulletin*, 119(3), 488.
- van den Bussche, H., et al. (2006). Medizinische Fakultäten: Der Ausbildungserfolg im Vergleich (II). *Deutsches Ärzteblatt International*, 103(34-35): 2225.
- van den Bussche, H., Ziegler, S., Krause-Solberg, L., & Scherer, M.. (2017). Stabilität und Wechsel der von angehenden Fachärztinnen und Fachärzten angestrebte fachärztliche Anerkennung nach vier Weiterbildungsjahren. *Das Gesundheitswesen*, 79(10), 865-870.
- Van Hoek, G., Portzky, M., & Franck, E.. (2019). The influence of socio-demographic factors, resilience and stress reducing activities on academic outcomes of undergraduate

- nursing students: A cross-sectional research study. *Nurse Education Today*, 72, 90-96.
- Vedel, A., & Poropat, A. E. (2017). Personality and academic performance. *Encyclopedia of personality and individual differences*, 1-9.
- Wang, H., Wu, S., & Liu, Y. (2003). Association Between Social Support and Health Outcomes: A Meta-analysis. *The Kaohsiung Journal of Medical Sciences*, 19(7), 345-350.
- Wang, M., & Degol, J. L. (2017). Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions. *Educational Psychology Review*, 29(1), 119-140.
- Walkiewicz, M., Tartas, M., Majkiewicz, M., Budzinski, W., (2012). Academic achievement, depression and anxiety during medical education predict the styles of success in a medical career: A 10-year longitudinal study. *Medical Teacher*, 34, e611–e619..
- Wilks, S. E. (2008). Resilience amid Academic Stress: The Moderating Impact of Social Support among Social Work Students. *Advances in Social Work*, 9(2), 106-125.
- Wissenschaftsrat (2014): *Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland auf Grundlage einer Bestandsaufnahme der humanmedizinischen Modellstudiengänge*; Dresden.
URL: <https://www.wissenschaftsrat.de/download/archiv/4017-14.html>
- Zigmond, A. S. & R. P. Snaith (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica* 67(6): 361-370.
- Zum Nachwuchsmangel in Unfallchirurgie und Orthopädie. (2010). *Zeitschrift für Orthopädie und Unfallchirurgie*, 148(01), 17-18.

Anhang

Vergleichs-OSCE: Prüfungsanweisungen für die Studierenden

(aus datenschutzrechtlichen Gründen nicht enthalten)

X

Fragebogen, 2. Fassung

Lieber Teilnehmer, liebe Teilnehmerin,

herzlichen Dank, dass Sie sich bereit erklärt haben, an dieser Umfrage teilzunehmen!
Bitte lesen Sie sich die folgenden Fragen gründlich durch und beantworten sie so wahrheitsgemäß wie möglich. Wenn Sie sich zwischen mehreren Alternativen nicht entscheiden können, wählen Sie bitte die aus, die Ihnen am meisten entspricht.
Das Ausfüllen des Fragebogens wird circa 20 Minuten dauern, bei Fragen können Sie sich jederzeit an den Testleiter wenden.

Im Folgenden erhalten Sie 30 Aussagen, die sich zur Beschreibung Ihrer eigenen Person eignen könnten. Bitte lesen Sie jede Aussage genau durch und kreuzen Sie als Antwort die Kategorie an, die Ihre Sichtweise am besten ausdrückt. Lassen Sie keine Aussage aus. Auch wenn Ihnen einmal die Entscheidung schwer fallen sollte, kreuzen Sie trotzdem immer eine Antwort an, die noch am ehesten auf Sie zutrifft. Es gibt bei diesem Fragebogen keine „richtigen“ oder „falschen“ Antworten.

- **Starke Ablehnung**, wenn Sie der Aussage auf keinen Fall zustimmen oder sie für völlig unzutreffend halten.
- **Ablehnung**, wenn Sie der Aussage eher nicht zustimmen oder sie für zutreffend halten.
- **Neutral**, wenn Sie die Aussage weder richtig noch falsch, also weder zutreffend noch unzutreffend ist.
- **Zustimmung**, wenn Sie der Aussage eher zustimmen oder sie für zutreffend halten.
- **Starke Zustimmung**, wenn Sie der Aussage nachdrücklich zustimmen oder sie für völlig zutreffend halten.

	starke Ablehnung	Ablehnung	neutral	Zustimmung	starke Zustimmung
1. Ich habe gern viele Leute um mich herum.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Ich halte meine Sachen ordentlich und sauber.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Ich fühle mich anderen oft unterlegen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	starke Ablehnung	Ablehnung	neutral	Zustimmung	starke Zustimmung

	Ablehnung				stimmung
4. Ich bin leicht zum Lachen zu bringen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Ich finde philosophische Diskussionen langweilig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Ich bekomme häufiger Streit mit meiner Familie und meinen Kollegen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Ich kann mir meine Zeit recht gut einteilen, sodass ich meine Angelegenheiten rechtzeitig beende.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Mich begeistern die Motive, die ich in der Kunst und in der Natur finde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Manche Leute halten mich für selbstsüchtig und selbstgefällig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Ich versuche, alle mir übertragenen Aufgaben sehr gewissenhaft zu erledigen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Ich fühle mich oft angespannt und nervös.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Ich bin gerne im Zentrum des Geschehens.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Poesie beeindruckt mich wenig oder gar nicht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Im Hinblick auf die Absichten anderer bin ich eher zynisch und skeptisch.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Manchmal fühle ich mich völlig wertlos.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Ich habe oft das Gefühl, vor Energie überzuschäumen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Ich bin ein fröhlicher, gut gelaunter Mensch.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Manche Leute halten mich für kalt und berechnend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Wenn ich eine Verpflichtung eingehe, so kann man sich auf mich bestimmt verlassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Zu häufig bin ich entmutigt und will aufgeben, wenn etwas schief geht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	starke Ablehnung	Ablehnung	neutral	Zustimmung	starke Zustimmung

22. Wenn ich Literatur lese oder ein Kunstwerk betrachte, empfinde ich manchmal ein Frösteln oder eine Welle der Begeisterung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Ich habe wenig Interesse, über die Natur des Universums oder die Lage der Menschheit zu spekulieren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Ich versuche stets rücksichtsvoll und sensibel zu handeln.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Ich bin eine tüchtige Person, die ihre Arbeit immer erledigt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Ich fühle mich oft hilflos und wünsche mir eine Person, die meine Probleme löst.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. Ich bin ein sehr aktiver Mensch.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Ich werde wohl niemals fähig sein, Ordnung in mein Leben zu bringen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Ich habe oft Spaß daran, mit Theorien oder abstrakten Ideen zu spielen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. Um zu bekommen, was ich will, bin ich notfalls bereit, Menschen zu manipulieren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

NRG-FF-30

Bitte kreuzen Sie zu jeder Aussage eine der Alternativen an!	stimmt überhaupt nicht	stimmt nicht	stimmt eher nicht	weder/ noch	stimmt eher	stimmt	stimmt genau
1. In den meisten Bereichen entspricht mein Leben meinen Idealvorstellungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Meine Lebensbedingungen sind ausgezeichnet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Ich bin mit meinem Leben zufrieden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Bisher habe ich die wesentlichen Dinge erreicht, die ich mir für mein Leben wünsche.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Wenn ich mein Leben noch einmal leben könnte, würde ich kaum etwas ändern.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

SWLS

In diesem Teil geht es um Ihre Beziehung zu wichtigen Menschen, also zum Partner, zu Familienmitgliedern, Freunden und Bekannten, Kollegen und Nachbarn. Es wird erfasst, wie Sie diese Be-

ziehung erleben und einschätzen. Bitte geben Sie an, wie sehr die Aussagen jeweils für Sie zutreffen.

	trifft nicht zu	trifft eher nicht zu	trifft teilweise zu	trifft zu	trifft genau zu
1. Ich finde ohne weiteres jemanden, der sich um meine Wohnung kümmert, wenn ich mal nicht da bin.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Es gibt Menschen, die mich ohne Einschränkungen so nehmen, wie ich bin.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Ich erfahre von anderen viel Verständnis und Geborgenheit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Ich habe einen sehr vertrauten Menschen, mit dessen Hilfe ich immer rechnen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Bei Bedarf kann ich mir ohne Probleme bei Freunden oder Nachbarn etwas ausleihen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Ich habe Freunde / Angehörige, die sich auf jeden Fall Zeit nehmen und gut zuhören, wenn ich mich aussprechen möchte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Ich kenne mehrere Menschen, mit denen ich gerne etwas unternehme.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Ich habe Freunde / Angehörige, die mich einfach mal umarmen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Wenn ich krank bin, kann ich ohne Zögern Freunde / Angehörige bitten, wichtige Dinge für mich zu erledigen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Wenn ich mal sehr bedrückt bin, weiß ich, zu wem ich damit ohne weiteres gehen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Es gibt Menschen, die Freude und Leid mit mir teilen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Bei manchen Freunden/ Angehörigen kann ich auch mal ganz ausgelassen sein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Ich habe einen vertrauten Menschen, in dessen Nähe ich mich ohne Einschränkung wohl fühle.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Es gibt eine Gruppe von Menschen (Freundeskreis, Clique), zu der ich gehöre und mit der ich mich häufig treffe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

F-SOZIALK

Bitte beantworten Sie jede der folgenden Fragen so, wie es für Sie persönlich **in der letzten Woche** am ehesten zutrif. Machen Sie bitte nur **ein Kreuz** pro Frage und lassen Sie bitte keine Frage

aus! Überlegen Sie bitte nicht lange, sondern wählen Sie die Antwort aus, die Ihnen auf Anhieb am zutreffendsten erscheint!

Ich fühle mich angespannt oder überreizt <input type="checkbox"/> meistens <input type="checkbox"/> oft <input type="checkbox"/> von Zeit zu Zeit/gelegentlich <input type="checkbox"/> überhaupt nicht	Ich fühle mich in meinen Aktivitäten gebremst <input type="checkbox"/> fast immer <input type="checkbox"/> sehr oft <input type="checkbox"/> manchmal <input type="checkbox"/> überhaupt nicht
Ich kann mich heute noch so freuen wie früher <input type="checkbox"/> ganz genau so <input type="checkbox"/> nicht ganz so sehr <input type="checkbox"/> nur noch ein wenig <input type="checkbox"/> kaum oder gar nicht	Ich habe manchmal ein ängstliches Gefühl in der Magenregion <input type="checkbox"/> überhaupt nicht <input type="checkbox"/> gelegentlich <input type="checkbox"/> ziemlich oft <input type="checkbox"/> sehr oft
Mich überkommt eine ängstliche Vorahnung, dass etwas Schreckliches passieren könnte <input type="checkbox"/> ja, sehr stark <input type="checkbox"/> ja, aber nicht allzu stark <input type="checkbox"/> etwas, aber es macht mir keine Sorgen <input type="checkbox"/> überhaupt nicht	Ich habe das Interesse an meiner äußeren Erscheinung verloren <input type="checkbox"/> ja, stimmt genau <input type="checkbox"/> ich kümmere mich nicht so sehr darum, wie ich sollte <input type="checkbox"/> möglicherweise kümmere ich mich zu wenig darum <input type="checkbox"/> ich kümmere mich so viel darum wie immer
Ich kann lachen und die lustige Seite der Dinge sehen <input type="checkbox"/> ja, so viel wie immer <input type="checkbox"/> nicht mehr ganz so viel <input type="checkbox"/> inzwischen viel weniger <input type="checkbox"/> überhaupt nicht	Ich fühle mich rastlos, muss immer in Bewegung sein <input type="checkbox"/> ja, tatsächlich sehr <input type="checkbox"/> ziemlich <input type="checkbox"/> nicht sehr <input type="checkbox"/> überhaupt nicht
Mir gehen beunruhigende Gedanken durch den Kopf <input type="checkbox"/> einen Großteil der Zeit <input type="checkbox"/> verhältnismäßig oft <input type="checkbox"/> von Zeit zu Zeit, aber nicht allzu oft <input type="checkbox"/> nur gelegentlich/nie	Ich blicke mit Freude in die Zukunft <input type="checkbox"/> ja, sehr <input type="checkbox"/> eher weniger als früher <input type="checkbox"/> viel weniger als früher <input type="checkbox"/> kaum bis gar nicht
Ich fühle mich glücklich <input type="checkbox"/> überhaupt nicht <input type="checkbox"/> selten <input type="checkbox"/> manchmal <input type="checkbox"/> meistens	Mich überkommt plötzlich ein panikartiger Zustand <input type="checkbox"/> ja, tatsächlich sehr oft <input type="checkbox"/> ziemlich oft <input type="checkbox"/> nicht sehr oft <input type="checkbox"/> überhaupt nicht
Ich kann behaglich dasitzen und mich entspannen <input type="checkbox"/> ja, natürlich <input type="checkbox"/> gewöhnlich schon <input type="checkbox"/> nicht oft <input type="checkbox"/> überhaupt nicht	Ich kann mich an einem guten Buch, einer Radio- oder Fernsehsendung freuen <input type="checkbox"/> oft <input type="checkbox"/> manchmal <input type="checkbox"/> eher selten <input type="checkbox"/> sehr selten

Die folgenden Fragen helfen, die Ergebnisse dieser Umfrage zu untersuchen und auszuwerten. Dabei ist es wichtig, dass die Forscher die Antworten auf die bisher beantworteten Fragen nach

Merkmale auswerten können, die gesellschaftliche Gruppen beschreiben. Hierfür benötigen wir detaillierte Angaben zu Ihrer Person, damit wir Sie einer entsprechenden Gruppe zuordnen können. Die Forscher werten die Daten nicht für Ihre Person aus, sondern für solche Gruppen, zu denen man Sie zum Beispiel entsprechend Ihrer Altersgruppe oder Ihrem Geschlecht zuordnen kann.

3. Was ist Ihr Geschlecht? <input type="checkbox"/> männlich <input type="checkbox"/> weiblich <input type="checkbox"/> anderes, und zwar: []
4. Ihr Alter? Geburtsmonat: [] [] Geburtsjahr: [] [] [] []
5. Was ist Ihre Muttersprache? (Mehrfachnennungen möglich) <input type="checkbox"/> deutsch (weiter bei Frage 7) <input type="checkbox"/> andere, und zwar []
6. Wie schätzen Sie Ihre deutschen Sprachkenntnisse ein? <input type="checkbox"/> sehr gut <input type="checkbox"/> gut <input type="checkbox"/> mittelmäßig <input type="checkbox"/> gering <input type="checkbox"/> schlecht
7. Seit wann leben Sie in Deutschland? <input type="checkbox"/> seit meiner Geburt <input type="checkbox"/> seit (Jahr) [] [] [] []
8. Welche Staatsangehörigkeit besitzen Sie? (Mehrfachnennungen möglich) <input type="checkbox"/> die deutsche Staatsangehörigkeit <input type="checkbox"/> andere, und zwar: []
<input type="checkbox"/> andere, und zwar: [] <input type="checkbox"/> keine Staatsangehörigkeit (staatenlos)
9. In welchem Land sind ihre Eltern geboren?

Mutter:

Deutschland

In einem anderen Land, und zwar: _____

Vater:

Deutschland

In einem anderen Land, und zwar: _____

10. Welchen Familienstand haben Sie? Was aus dieser Liste trifft auf Sie zu?

Verheiratet und lebe mit meinem/r Ehepartner/-in zusammen

In eingetragener Lebenspartnerschaft zusammenlebend (gleichgeschlechtlich)

Verheiratet und lebe von meinem/meiner Ehepartner/-in getrennt

Ledig

Geschieden

Verwitwet

Eingetragene Lebenspartnerschaft, getrennt lebend (gleichgeschlechtlich)

Eingetragene Lebenspartnerschaft aufgehoben (gleichgeschlechtlich)

Eingetragene/r Lebenspartner/-in verstorben (gleichgeschlechtlich)

11. Welchen höchsten beruflichen Ausbildungsabschluss haben Sie?

Noch in beruflicher Ausbildung (Berufsvorbereitungsjahr, Auszubildende/r, Praktikant/-in, Student/-in)

Keinen beruflichen Abschluss und bin nicht in beruflicher Ausbildung

Beruflich-betriebliche Berufsausbildung (Lehre) abgeschlossen

Beruflich-schulische Ausbildung (Berufsfachschule, Handelsschule, Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung) abgeschlossen

Ausbildung an einer Fach-, Meister-, Technikerschule, Berufs- oder Fachakademie abgeschlossen

Bachelor an (Fach-)Hochschule abgeschlossen

Fachhochschulabschluss (z. B. Diplom, Master)

Universitätsabschluss (z. B. Diplom, Magister, Staatsexamen, Master)

Promotion

Einen anderen beruflichen Abschluss, und zwar: _____

12. Wenn Sie sich jetzt entscheiden müssten: Welche Facharztweiterbildung würden Sie anstreben?

Ich möchte keine Facharztweiterbildung machen.

Ich würde am liebsten Facharzt für: _____ werden.

Haben Sie noch etwas anzumerken?

Damit haben Sie den Fragebogen vollständig ausgefüllt. Bitte legen Sie ihn in den dafür vorgesehenen Rücklaufkarton.

Herzlichen Dank für Ihre Teilnahme, mit Ihren Angaben tragen Sie zur Verbesserung der medizinischen Ausbildung bei!

Teilnehmer:inneninformation



Liebe Teilnehmer, liebe Teilnehmerin,

wir möchten den Zusammenhang zwischen persönlichen Charakteristika von Medizinstudierenden kurz vor Ende ihres Studiums und ihrer Leistung in einem OSCE-Verfahren zur Messung der Arztreife untersuchen.

Dazu werden Sie nach Ihrer Teilnahme an einem OSCE einem kurzen schriftlichen Fragebogen ausfüllen, der Ihnen durch einen Mitarbeiter des Instituts für Allgemeinmedizin ausgehändigt wird. In dem Fragebogen werden Informationen über Ihre Person und Lebenssituation erhoben, dokumentiert und ausgewertet. Alle Daten der Befragung werden sofort pseudonymisiert¹, das heißt, unter einer persönlichen Studiennummer gespeichert. Ihr Name erscheint nicht in Verbindung mit diesen Daten!

Die Daten werden vom Institut für Allgemeinmedizin der Universitätsklinik Hamburg-Eppendorf ausgewertet und ausschließlich zu Forschungszwecken verwendet. Wissenschaftliche Veröffentlichungen erfolgen ausschließlich in anonymisierter² Form und lassen keinerlei Rückschluss auf meine Person zu. Ihre Daten werden selbstverständlich nicht an Dritte weitergegeben! Nach Ablauf der Aufbewahrungsfrist von 10 Jahren werden alle erhobenen Daten vernichtet.

Sie haben das Recht, über Ihre Daten informiert zu werden. Hierzu können Sie sich an Fr. Keim (s.u.) wenden, die alle weiteren Schritte in die Wege leitet.

Ihre Teilnahme an der Untersuchung ist freiwillig. Ihre Einwilligung können Sie jederzeit und ohne Angabe von Gründen widerrufen, ohne dass es nachteilige Folgen für Sie hat.

Bei Fragen können Sie uns gerne direkt kontaktieren, oder den Testleiter vor Ort ansprechen.

Mit Ihrer Teilnahme ermöglichen Sie die Durchführung einer Promotion am Institut für Allgemeinmedizin und tragen zur Verbesserung der medizinischen Ausbildung bei. Wir würden uns daher sehr freuen, wenn Sie zur Teilnahme bereit wären!

Unsere Namen und Adressen:

Projektleitung: Prof. Dr. med. Martin Scherer (Tel. 7410-52400, m.scherer@uke.de)

Projektdurchführung: Rebecca Keim (Tel. 7410-57639, r.keim@uke.de)

Institut für Allgemeinmedizin
Universitätsklinikum Hamburg-Eppendorf
Martinistraße 52
20246 Hamburg
Fax: 040 7410-53681

Herzlichen Dank und mit freundlichen Grüßen,

(Prof. Dr. Martin Scherer)

(Dipl.-Psych. Rebecca Keim)

¹ **Pseudonymisieren** ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Identifizierung des Betroffenen auszuschließen oder wesentlich zu erschweren (§ 3 Abs. 6a Bundesdatenschutzgesetz).

² **Anonymisieren** ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können (§ 3 Abs. 6 Bundesdatenschutzgesetz).

Danksagung

Ich möchte mich bedanken: bei meinem Doktorvater Prof. Dr. Martin Scherer, für seine pragmatische Nüchternheit angesichts meiner Besorgnis, das wieder-auf-die-Schiene-Setzen und das wohl dosierte Lob und die aufbauenden Emails zum richtigen Zeitpunkt. Bei den beiden anderen Mitgliedern meines hochkarätigen Thesiskomitees: bei Prof. Dr. Sigrid Harendza, deren Engagement und kompetente Freundlichkeit nicht nur inspirierend, sondern auch tröstlich waren. Bei Prof. Dr. Bernd Löwe für seine Ruhe, die klaren Einsichten und kritischen Hinweise. Mein Dank gilt denen, die den Vergleichs-OSCE aus der Taufe gehoben, geplant und durchgeführt haben, allen voran Katharina Dahl, Dr. Maren Ehrhardt, Dr. Dagmar Lühmann, selbstverständlich Prof. Dr. Scherer und Kerstin Schrom (in alphabetischer Reihenfolge!). Und allen, die bei der Prüfung als Rater:innen, Simulationspatient:innen, Untersuchungsprobanden und schließlich Teilnehmer:innen Zeit und Mühe investiert haben. Danke, dass ich von euch, beziehungsweise Ihnen lernen durfte.

Dr. Sigrid Boczor, ohne deren mir immer noch unbegreiflichen Einsatz ich an der Statistik und später am Verschriftlichen dieser vermutlich verzweifelt wäre – danke dir. Deine ruhige und freundliche Art hat mir sehr wohlgetan und deine Kompetenz sehr geholfen. Mögen deine Pflanzen immer gedeihen und dein Tee die perfekte Temperatur halten. Dr. Thomas Zimmermann: danke für deine kritische Deutlichkeit. Durch das Doktorandentreffen konnte ich nie ganz aufhören, und das hat mich sicher daran gehindert, manchmal einfach hinzuwerfen. Danke Stefanie Butz, noch einmal tausend Dank an dich, Dagmar und Dr. Nadine Pohontsch, danke– trotz großem Stress habt ihr euch meine Texte angesehen und ich durch eure kompetenten Rückmeldungen enorm profitiert.

Viele Freunde haben mich in den 5 langen Jahren meiner Promotionszeit unterstützt, nie das Zutrauen verloren und mir in vielen kleinen und großen Dingen geholfen. Allen voran möchte ich dir, Sarah Porzelt, für deine unerschütterliche Freundschaft, Begleitung und Unterstützung danken. Niemand tröstet so gut wie du und mit wenigen kann ich mich so amüsieren. Ich plane fest, es dir bei deiner Promotion genauso zu vergelten!

Und ein letztes, aber sehr wichtiges Dankeschön gilt meinem Ehemann Oliver. Du bist mein Fels in der Brandung, nur zusammen sind wir ein funktionaler Erwachsener.

Lebenslauf

(aus datenschutzrechtlichen Gründen nicht enthalten)

Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: