

Bundesgesundheitsbl 2018 · 61:178–186
<https://doi.org/10.1007/s00103-017-2670-2>
 Online publiziert: 2. Januar 2018
 © Der/die Autor(en) 2017. Dieser Artikel ist eine Open-Access-Publikation.



Anja Schwibbe¹ · Janina Lackamp¹ · Mirjana Knorr¹ · Johanna Hissbach¹ · Martina Kadmon² · Wolfgang Hampe¹

¹ Arbeitsgruppe Auswahlverfahren, Institut für Biochemie und Molekulare Zellbiologie, Zentrum für Experimentelle Medizin, Universitätsklinikum Hamburg Eppendorf, Hamburg, Deutschland

² Dekanat, Medizinische Fakultät, Universität Augsburg, Augsburg, Deutschland

Medizinstudierendenauswahl in Deutschland

Messung kognitiver Fähigkeiten und psychosozialer Kompetenzen

Einleitung

Aktuell befasst sich das Bundesverfassungsgericht mit der Frage, ob die Auswahl von Medizinstudierenden verfassungskonform ist. Seit der Novellierung des Hochschulrahmengesetzes (HRG) im Jahr 2004 werden über die Stiftung für Hochschulzulassung (hochschulstart.de, ehemals Zentralstelle für die Vergabe von Studienplätzen ZVS) nach Abzug von Vorabquoten (z. B. für Nicht-EU-AusländerInnen oder Härtefälle) 20 % der Studienplätze an die Abiturbesten und ebenfalls 20 % aufgrund der Wartezeit seit dem Abitur vergeben. Für die verbleibenden etwa 60 % der Studienplätze können die Fakultäten innerhalb gesetzlicher Vorgaben Auswahlkriterien festlegen.

Das Interesse am Medizinstudium ist enorm: Zum Wintersemester 2017/2018 bewarben sich 43.184 BewerberInnen um die 9176 Studienplätze. In den meisten Bundesländern erhalten nicht alle BewerberInnen mit einer Abiturdurchschnittsnote von 1,0 über die Abiturbestenquote einen Studienplatz. Die für eine Zulassung benötigte Wartezeit beträgt 7–8 Jahre [1]. Obwohl die Fakultäten eine Vielfalt von Auswahlverfahren einsetzen, können AbiturientInnen selbst mit Abiturnoten unter 2,0 praktisch nur über eine Zulassung in der Wartezeitquote Medizin studieren.

Die Ziele bei der Auswahl der BewerberInnen sind vielfältig:

- Die AbsolventInnen sollen kompetente ÄrztInnen werden, die die zukünftige medizinische Versorgung der Bevölkerung sichern.
- Um die vorhandenen Ressourcen gut zu nutzen und die persönliche Belastung der Studierenden durch Misserfolge zu verringern, soll die Anzahl der Studienabbrüche minimiert werden.
- Die Studienleistungen sollen hoch sein und zeitgerecht erbracht werden. Sie werden meist zur Qualitätsermittlung von Auswahlverfahren herangezogen, da einheitliche Kriterien zur Beurteilung des ärztlichen Berufserfolgs nicht definiert sind oder erst viele Jahre nach Studienbeginn messbar wären.
- Die Passung der Studierenden zu Fakultät und Studiengang soll erreicht werden, um eine hohe Studienzufriedenheit und Bindung der Studierenden an ihre Fakultät herzustellen.
- Die Auswahlverfahren müssen mit vertretbarem Aufwand durchgeführt werden können und gleichzeitig hohe Standards der Testgütekriterien erfüllen, um eine faire Auswahl zu gewährleisten.

Im vor kurzem veröffentlichten Masterplan Medizinstudium 2020 einigten sich Bundes- und Landespolitik auf eine Weiterentwicklung der Mediziner Ausbildung [2]. Bei der Zulassung zum Studium sollen soziale und kommunikative Aspekte sowie praktische Erfahrung in medi-

zinnahen Bereichen in Zukunft stärker gewichtet werden. Zusätzlich wird eine „Landarztquote“ vorgeschlagen, in der Annahme, dadurch die hausärztliche Versorgung in ländlichen Regionen zukünftig zu gewährleisten. Die über diese Quote Zugelassenen müssen sich verpflichten, nach dem Studium auf dem Land zu arbeiten. Die Universitäten sollen mit speziellen Auswahlverfahren geeignete BewerberInnen identifizieren. Bisher sind aber keine Verfahren bekannt, die eine spätere hausärztliche Tätigkeit in ländlichen Gebieten vorhersehen. Internationale Erfahrungen deuten nur darauf hin, dass aus ländlichen Regionen stammende BewerberInnen nach dem Studium häufiger auf dem Land arbeiten [3].

Fragestellung

In diesem Beitrag geben wir eine Übersicht über die bisher in Deutschland eingesetzten Auswahlverfahren und beurteilen ihre Qualität unter Berücksichtigung des aktuellen Standes der internationalen Forschung. Im zweiten Teil berichten wir über ein fakultätsübergreifendes Projekt zur Entwicklung eines Situational Judgement Tests (SJT), der mit vertretbarem Aufwand zur Messung psychosozialer Kompetenzen bei großen Bewerberzahlen eingesetzt werden könnte. Der Artikel endet mit einem Vorschlag, wie der zukünftige Zulassungsweg eines potenziellen Medizinstudierenden unter

Tab. 1 Verfahren zur Auswahl von Medizinstudierenden an den 35 staatlichen deutschen Hochschulen zum Wintersemester 2017/2018

Vorauswahlverfahren	Anzahl Fakultäten	Auswahlverfahren	Anzahl Fakultäten
Abiturnote (Cut-off 2,5)	1	Ausschließlich Abiturnote	5
Ortspräferenz		Abiturnote in Kombination mit ^a	
1	3	Einzelfachnoten	3
≤2	2	TMS	22
≤3	4	HAM-Nat	3
Mit Abiturnote	11	Medizinischer Verständnistest/ Bewerbungsschreiben	1
Mit Abiturnote + Boni	3	Klassische Interviews	6
		Multiple Mini-Interviews	5
Keine Vorauswahl	11	Berufsausbildung/-tätigkeit	22
		Soziales Engagement/Dienste	11
		Naturwiss. Engagement	7
		Sportliche Leistungen	3
		Jugend musiziert/Fremdsprachen	1

TMS Test für medizinische Studiengänge, HAM-Nat Hamburger Auswahlverfahren für Medizinstudierende – Naturwissenschaftsteil

^a Viele Fakultäten kombinieren mehrere Kriterien mit der Abiturnote

Berücksichtigung der Erkenntnisse aus-
sehen könnte.

Übersicht der Auswahlverfahren in der Medizin

Im Laufe der letzten Jahre nahmen immer mehr staatliche medizinische Fakultäten ihr Recht in Anspruch, in der Auswahl ihrer StudienbewerberInnen neben der Abiturnote weitere Kriterien zu berücksichtigen. 2008 wählten noch 15 der damals 34 Fakultäten ihre BewerberInnen allein aufgrund der Abiturnote aus [4], heute sind es nur noch fünf von 35 staatlichen medizinischen Fakultäten (■ Tab. 1). Hingegen werden Test- und Interviewverfahren deutlich häufiger eingesetzt. Während 2008 die BewerberInnen nur an sechs Universitäten mit dem „Test für medizinische Studiengänge“ (TMS) und mit dem HAM-Nat (Hamburger Auswahlverfahren für Medizinische Studiengänge – Naturwissenschaftsteil) ihre Chancen auf einen Studienplatz erhöhen konnten, nutzen aktuell 22 Fakultäten die Ergebnisse im TMS und drei Fakultäten die im HAM-Nat als Auswahlkriterien. Klassische Interviews werden aktuell zwar nur noch an sechs statt neun Fakultäten durchgeführt, aber an fünf

weiteren Standorten werden Multiple Mini-Interviews verwendet. Gerade bei Verwendung solch aufwendiger Auswahlverfahren schränken Fakultäten die Zahl der TeilnehmerInnen durch eine Vorauswahl bei den BewerberInnen ein. Häufig werden nur diejenigen berücksichtigt, die den Studienort mit erster bis dritter Ortspräferenz gewählt haben, um sicherzustellen, dass die ausgewählten BewerberInnen die Studienplätze mit hoher Wahrscheinlichkeit annehmen. Auch private Hochschulen, die ein Medizinstudium anbieten, führen eine Auswahl durch, sind jedoch nicht an die Vorgaben des HRG gebunden.

Kognitive Auswahlkriterien

Wenn Kriterien in der Studierendenauswahl berücksichtigt werden, die vorwiegend Wahrnehmung, Schlussfolgern, Verständnis, Denken oder Problemlösen erfordern [5], wird von „kognitiven Auswahlkriterien“ gesprochen. Bei den in der Medizin eingesetzten kognitiven Auswahlverfahren wird zwischen Methoden unterschieden, die eher angeborene Fähigkeiten erfassen, und Methoden, die gelernte Fertigkeiten und Wissen messen [6].

Fähigkeiten sind physische und psychische Eigenschaften, die einen Menschen generell dazu befähigen, körperliche oder geistige Leistungen zu erbringen [7]. Studienfachübergreifende und studienfachspezifische Studierfähigkeitstests sollen kognitive Fähigkeiten zur Vorhersage des Studienerfolgs erfassen [8]. Da Fähigkeiten als über die Zeit hinweg stabil gelten, wird angenommen, dass Studierfähigkeitstests nur wenig trainierbar sind.

Schulnoten und Kenntnistests messen überwiegend Wissen und auf formaler Bildung basierende angeeignete Fertigkeiten. Im Gegensatz zu Fähigkeiten beziehen sich Fertigkeiten auf konkretes Können bestimmter Verhaltensweisen wie Schreiben oder Geschicklichkeit [7]. Neben zugrunde liegenden Fähigkeiten braucht man für ein gutes Abschneiden auch Motivation und Persönlichkeitseigenschaften wie Gewissenhaftigkeit [9]. Fähigkeitsnachteile lassen sich hier in bestimmten Grenzen durch Motivation und Übung ausgleichen [10].

Hochschulen können abwägen, ob sie BewerberInnen bevorzugen möchten, die schon mit den gewünschten Fähigkeiten ausgestattet sind, oder ob sie der Meinung sind, dass sich durch Lehre und Übung gewisse Fähigkeitsdefizite auffangen, Fertigkeiten erwerben und gewünschte Leistungen erbringen lassen.

Schulnoten

Die Abiturdurchschnittsnote nimmt auch nach der Novellierung des HRG weiter eine Sonderstellung ein, da ihr in der Auswahlentscheidung ein „maßgeblicher Einfluss“ gegeben werden muss [11]. Eine Metaanalyse aus dem deutschsprachigen Raum belegt generell die prognostische Validität der Abiturdurchschnittsnote für den Studienerfolg [12]. Die durchschnittliche Prognosekraft der Abiturnote für den Medizinstudienabschluss fällt mit $r = 0,55$ (für Kriteriumsreliabilität und Varianzeinschränkung durch Selektion korrigierter Validitätskoeffizient) und im vorklinischen Bereich mit $r = 0,58$ noch stärker aus als in anderen Studiengängen [12]. Auch international sagen Schulabschlussnoten Studienleistungen und -abbruch in der Medizin voraus [13]. Abschlussnoten

Bundesgesundheitsbl 2018 · 61:178–186 <https://doi.org/10.1007/s00103-017-2670-2>
 © Der/die Autor(en) 2017. Dieser Artikel ist eine Open-Access-Publikation.

A. Schwibbe · J. Lackamp · M. Knorr · J. Hissbach · M. Kadmon · W. Hampe

Medizinstudierendenauswahl in Deutschland. Messung kognitiver Fähigkeiten und psychosozialer Kompetenzen

Zusammenfassung

Das Bundesverfassungsgericht überprüft aktuell, ob das Vergabeverfahren der Medizinstudienplätze mit dem Grundrecht auf freie Berufswahl vereinbar ist, da BewerberInnen ohne sehr gute Abiturnoten meist sieben Jahre warten müssen. Bei einer Umstellung des Zulassungssystems möchte die Politik, dem Masterplan Medizinstudium 2020 folgend, psychosoziale Auswahlkriterien stärker gewichten. Welche Erfahrungen gibt es mit den aktuellen Auswahlverfahren? Wie können Situational Judgement Tests in Zukunft zur Validität der Auswahlverfahren für deutsche Medizinfakultäten beitragen? Untersuchungen zeigen, dass die Abiturnote der beste Prädiktor von Studienleistungen ist. Die Auswahl der BewerberInnen nach

Abiturnote gerät jedoch verstärkt in die Diskussion, da die Vergleichbarkeit der Abiturnoten infrage steht und die Bewerberzahl mit Bestnote stark anwächst. Fähigkeits- und Kenntnistests können inkrementell zur Abiturnote die Vorhersage der Studienleistungen verbessern, wobei insbesondere die naturwissenschaftlichen Anteile bedeutsam sind. Die Messung psychosozialer Kompetenzen in klassischen Interviews ist wenig reliabel und valide. Die reliableren Multiple Mini-Interviews können vor allem praktische Studienergebnisse besser vorhersagen. Im Ausland eingesetzte Situational Judgement Tests (SJTs) werden als reliabel und valide eingeschätzt, die Korrelation eines in Hamburg pilotierten

deutschen SJT mit dem Multiple Mini-Interview stimmt vorsichtig positiv. Ein vom Medizinischen Fakultätentag und der Bundesvertretung der Medizinstudierenden vorgeschlagenes Modell berücksichtigt diese Ergebnisse. In dem Modell soll die Studierendenauswahl zu je 40 % auf der Abiturnote und einem kognitiven Test sowie zu je 10 % auf einem SJT und berufspraktischer Erfahrung beruhen. Zusätzlich können die Fakultäten spezifische Verfahren einsetzen.

Schlüsselwörter

Medizin · Studierendenauswahl · Situational Judgement Test · Masterplan Medizinstudium 2020 · Kognitive und psychosoziale Auswahlkriterien

Selection of medical students. Measurement of cognitive abilities and psychosocial competencies

Abstract

The German Constitutional Court is currently reviewing whether the actual study admission process in medicine is compatible with the constitutional right of freedom of profession, since applicants without an excellent GPA usually have to wait for seven years. If the admission system is changed, politicians would like to increase the influence of psychosocial criteria on selection as specified by the Masterplan Medizinstudium 2020. What experiences have been made with the actual selection procedures? How could Situational Judgement Tests contribute to the validity of future selection procedures to German medical schools? High school GPA is the best predictor of study performance, but is more and more under

discussion due to the lack of comparability between states and schools and the growing number of applicants with top grades. Aptitude and knowledge tests, especially in the natural sciences, show incremental validity in predicting study performance. The measurement of psychosocial competencies with traditional interviews shows rather low reliability and validity. The more reliable multiple mini-interviews are superior in predicting practical study performance. Situational judgement tests (SJTs) used abroad are regarded as reliable and valid; the correlation of a German SJT piloted in Hamburg with the multiple mini-interview is cautiously encouraging.

A model proposed by the Medizinischer Fakultätentag and the Bundesvertretung der Medizinstudierenden considers these results. Student selection is proposed to be based on a combination of high school GPA (40%) and a cognitive test (40%) as well as an SJT (10%) and job experience (10%). Furthermore, the faculties still have the option to carry out specific selection procedures.

Keywords

Medicine · Student selection · Situational judgement test · Masterplan Medizinstudium 2020 · Cognitive and psychosocial selection criteria

sind für alle BewerberInnen leicht verfügbar. Ihre Reliabilität und prädiktive Validität werden auf die Aggregation vieler schulischer Einzelbewertungen zurückgeführt [14]. Trotzdem mehrt sich in den letzten Jahren die Kritik an der fehlenden Diskriminationskraft [15], da die Anzahl der BewerberInnen mit Bestnoten stark zugenommen hat. Zudem unterscheiden sich die Berechnungsmaßstäbe der Gesamtabiturnote zwischen den Bundesländern, die Über-

einstimmung der BeurteilerInnen aus unterschiedlichen Schulen ist mäßig und die Retestreliabilität ist relativ gering [16]. Tent [17] resümiert, dass Noten nicht so schlecht seien wie ihr Ruf, eine Studiauswahlentscheidung aber nicht allein auf ihnen basieren sollte.

Neben der Abiturnote werden an drei deutschen Universitäten Einzelfachnoten in der Auswahl boniert (Tab. 1), da sie als Prädiktoren der fachspezifischen Studierfähigkeit gesehen werden [18].

Die meisten Studien zeigen jedoch, dass die Vorhersage des Studienerfolges durch Einzelnoten allein oder in Kombination mit der Abiturdurchschnittsnote nicht oder nur in bestimmten Studienfächern gering verbessert werden kann [12].

Nicht alle BewerberInnen können die geforderten Fachnoten vorweisen, da nicht alle Schulen alle Fächer anbieten oder BewerberInnen Schulen im EU-Ausland mit anderen Schulsystemen besucht haben, sodass eine generelle

Verfügbarkeit wie bei der Abiturnote nicht gegeben ist [14]. Dennoch werden international Fachnoten oft für die Studierendenauswahl berücksichtigt. Insbesondere in Großbritannien sind die Fächerkombination und -anzahl der Sekundärstufe (A-Level) bedeutsam [19].

Studierfähigkeits- und Kenntnistests

In den letzten Jahren wurden zunehmend Testverfahren eingesetzt, von denen man sich einen zur Abiturnote inkrementellen Erkenntnisgewinn verspricht [14]. Zudem gelten sie als objektiver und fairer hinsichtlich sozialer Selektionseinflüsse [20]. Die beiden in Deutschland am häufigsten eingesetzten Verfahren sind der TMS [21, 22] und der HAM-Nat [23–25].

Der TMS zählt zu den Fähigkeitstests und misst das Verständnis für naturwissenschaftliche und medizinische Problemstellungen, Merkfähigkeit, die Genauigkeit der visuellen Wahrnehmung, das räumliche Vorstellungsvermögen und die Fähigkeit zu konzentriertem und sorgfältigem Arbeiten [21, 22] mithilfe von neun Untertests, ohne dass medizinisches Vorwissen dafür nötig ist. Eine aktuelle Untersuchung aus Heidelberg ergab ein Cronbachs Alpha von 0,75 auf Itemebene über alle Subtests hinweg. Es zeigten sich eine inhaltlich nachvollziehbare Zweifaktorenstruktur und erste Hinweise auf die prädiktive Validität des Verfahrens im heutigen Kontext [22, 26]. Der TMS ermöglicht BewerberInnen mit mittleren Abiturnoten den Zugang zu einem Studienplatz, die dann vergleichbare Studienleistungen wie BewerberInnen der Abiturbestenquote mit mittleren TMS-Ergebnissen erreichen [27]. Der TMS wird jährlich im Frühjahr an mehreren Standorten durchgeführt. Aufgrund ihres Ergebnisses können die BewerberInnen abschätzen, wie stark sich ihre Zugangschancen an den einzelnen Universitäten erhöhen und die Ortspräferenzen bei ihrer Studienplatzbewerbung anpassen [21].

Der HAM-Nat ist hingegen ein Kenntnistest, der für das Medizinstudium relevante naturwissenschaftliche Kenntnisse auf Oberstufenniveau erfasst [23]. Motivierte BewerberInnen können sich in-

tensiv vorbereiten und damit ihre Zulassungschancen und gleichzeitig ihr für das Studium benötigte Vorwissen verbessern. Sowohl Ergebnisse aus Hamburg als auch aus Magdeburg berichten zufriedenstellende interne Konsistenzen zwischen 0,85 und 0,89 [24, 25]. Der HAM-Nat sagt den für den erfolgreichen Abschluss des Medizinstudiums relevanten Studienerfolg nach zwei Jahren voraus und zeigt inkrementelle prädiktive Validität zur Abiturnote [24]. Eine signifikante Wechselwirkung zwischen Abiturnote und Testergebnis in einem Regressionsmodell zur Vorhersage des Studienerfolgs nach sieben Semestern (Odds Ratio = 1,80, $p = 0,037$) zeigt, dass der HAM-Nat besonders bei BewerberInnen mit sehr guten Abiturnoten inkrementelle Validität aufweist [23]. Anders als der TMS wird der HAM-Nat nach Abschluss der Bewerbung an den einzelnen Hochschulen durchgeführt, dort ist die Teilnahme für die Studienplatzvergabe verpflichtend.

Die vier Subtests zu biologischem, chemischem und physikalischem Wissen des in den USA und Kanada eingesetzten Medical College Admission Tests (MCAT) sind mit dem Kenntnistest HAM-Nat vergleichbar, werden aber noch durch Subtests zu verbalem Schlussfolgern und das Schreiben von zwei Essays ergänzt. Eine Metaanalyse zeigt, dass neben der zufriedenstellenden Validität des Gesamttests ($r = 0,39$, 95 % KI 0,21–0,54 für den vorklinischen Abschnitt) insbesondere der Subtest Biologie die höchsten korrelativen Zusammenhänge ($r = 0,32$, 95 % KI 0,21–0,42) aufweist [28]. Mit dem TMS vergleichbar ist der United Kingdom Clinical Aptitude Test (UKCAT) mit den vier Subtests „abstract reasoning“, „decision analysis“, „quantitative reasoning“ und „verbal reasoning“ und einem SJT. Im aktuellen Jahresbericht von 2015/2016 werden zufriedenstellende Reliabilitäten für den kognitiven Teil von 0,89 berichtet, wobei die Berechnungsmethode nicht näher spezifiziert wird [29]. Aber die groß angelegte UKCAT-12-Studie zeigte, dass das Gesamtergebnis mit Leistungen im Medizinstudium eher schwach korreliert ($r = 0,15$) und nur einen sehr kleinen inkrementellen Beitrag im Vergleich zu den Schulnoten leistet [6]. Andere Stu-

dien finden keine Zusammenhänge [30, 31]. Der ebenfalls aus Großbritannien stammende Biomedical Admissions Test (BMAT) ist in erster Linie ein Fähigkeitstest, aber zu einem Drittel auch Wissenstest [32]. Als fast alleiniger Prädiktor der Examensleistungen im ersten und zweiten Studienjahr stellte sich der Wissensteil des BMAT zu Biologie, Chemie, Physik und Mathematik heraus [33, 34]. Auch für den Graduate Australian Medical School Admissions Test (GAMSAT) geht die prädiktive Kraft nur vom Subtest zu naturwissenschaftlichen Kenntnissen aus [35].

In einer Metaanalyse konnten Hell et al. [8] die prognostische Güte des TMS als den bekanntesten Studierfähigkeitstest für den deutschsprachigen Raum nachweisen. Auf internationaler Ebene schätzen Patterson et al. [13] die Befundlage zur prädiktiven Validität von Studieneignungstests allerdings als widersprüchlich ein, wobei sie sich auf Leistungs- und Fähigkeitstests beziehen. Harris et al. [32] gehen noch einen Schritt weiter. Sie fassen die bisherigen Entwicklungen im Bereich der Medizinbewerbersauswahl zusammen und bewerten den Einsatz von Verfahren, die nicht auf Kenntnissen und Wissen basieren, vor dem Hintergrund fehlender Validitätsnachweise als äußerst kritisch.

Psychosoziale Auswahlkriterien

Die Bedeutung nichtakademischer Kriterien wie Kommunikationsfertigkeiten oder professionelles Verhalten sowie bestimmter Persönlichkeitseigenschaften in der Studierendenauswahl zukünftiger MedizinerInnen hat in den letzten Jahren deutlich zugenommen [36]. Immer mehr Universitäten suchen nach Messmethoden, die eine Erfassung dieser Kriterien im Auswahlprozess ermöglichen [13].

Klassische Interviews

Interviews zählen international zu den am häufigsten eingesetzten Verfahren in der Bewerberauswahl für ein Medizinstudium [13] und auch in Deutschland gehören sie zu den im HRG aufgeführten möglichen Auswahlkriterien [11]. Interviews bieten Fakultäten die Möglichkeit, Interessen, Motivation und Kommunika-

tionsvermögen der BewerberInnen einzuschätzen, und bieten BewerberInnen die Chance, mit der Hochschule in direkten Kontakt zu treten und Informationen einzuholen. Sie werden aber häufig wegen des immensen Personal- und Zeitaufwandes nur mit einer vorselektierten Bewerbergruppe durchgeführt [16]. Entlang eines Kontinuums lassen sich freie, teilstrukturierte und strukturierte Interviews unterscheiden, wobei Reliabilität und Validität mit steigender Strukturierung zunehmen [16, 37]. Trotz der großen Verbreitung gelten insbesondere nichtstrukturierte Auswahlgespräche als nicht objektiv, reliabel und prädiktiv valide genug, um in der Medizinbewerbersauswahl eingesetzt zu werden [13].

Multiple Mini-Interviews

Seit gut zehn Jahren wenden sich medizinische Fakultäten weltweit zunehmend den strukturierten Interviews zu, wobei das Multiple Mini-Interview (MMI) eine vorherrschende Rolle einnimmt [13]. Es hat den höchsten Strukturierungsgrad und wird mittlerweile auch in Deutschland von fünf Fakultäten eingesetzt (■ Tab. 1). Das MMI wurde am Vorbild der OSCEs (Objective Structured Clinical Examination) 2002 an der McMaster-Universität in Kanada entwickelt und besteht aus einem Rundkurs mehrerer Interviewstationen mit jeweils neuen Aufgaben und neuen Interviewern und Juroren an jeder Station [38]. Beispielsweise findet an einer Station ein Rollenspiel mit einem Schauspieler statt, an einer anderen Station müssen in einem Interview persönliche Erfahrungen diskutiert und an einer weiteren Station standardisierte Fragen beantwortet werden [39]. Da Verhalten kontextspezifisch ist, lässt sich schlecht vom Verhalten in einer Situation auf das in einer anderen schließen. Das MMI bietet eine Zusammenstellung mehrerer Situationen, um eine reliablere Aussage zu ermöglichen [40]. Reviews zeigen, dass sich mit MMIs je nach Anzahl, Dauer und Inhalt der Stationen und Anzahl der Juroren zufriedenstellende Werte für Reliabilität und prädiktive Validität erreichen lassen, wobei besser Inhalte praktischer Aufgaben aus OSCEs vorhergesagt werden [13, 39]. Knorr und Hissbach [39] berichten Ge-

neralisierbarkeitskoeffizienten zwischen 0,51 und 0,88, wobei Verbesserungen eher an mehr Stationen und weniger an mehr Rater in den Stationen gebunden sind, was aber mit sehr hohen Mehrkosten verbunden ist [41]. Allerdings bleibt die Frage der Konstruktvalidität bisher weitestgehend unbeantwortet. MMIs können ähnlich wie Assessment-Center neben Interviews und Rollenspielen auch Gruppendiskussionen und praktische Aufgaben enthalten. In Hamburg, Oldenburg und Witten/Herdecke werden Gruppendiskussionen als Teil des MMIs durchgeführt und in Münster bearbeiten die BewerberInnen auch praktische Aufgaben.

Berufsausbildung, soziales oder kulturelles Engagement

Es gibt keinen Hinweis darauf, dass eine Berufstätigkeit im medizinischen Bereich oder ein freiwilliger Sozialdienst positiv mit dem Studienerfolg korreliert. StudienbewerberInnen mit Berufserfahrung können sogar signifikant schlechter in den Prüfungen abschneiden [42]. Dies kann zumindest zum Teil auf schlechtere Abiturnoten zurückgeführt werden, da die Berufstätigkeit häufig zur Überbrückung der Wartezeit bis zur Studienzulassung ausgeübt wird. Zudem haben die Berufstätigen meist lange mit anspruchsvollem schulischen oder akademischen Lernen ausgesetzt sowie eine veränderte familiäre Situation oder durch das Einkommen gestiegene Lebensansprüche, sodass sie weniger Zeit für das Studium aufbringen können. Möglicherweise bereichern sie studentische Diskussionen um andere Sichtweisen und erhöhen die Heterogenität der Studierendenschaft, was aber nur schwer zu quantifizieren ist.

Es gibt keine Studien, die zeigen, dass durch Referenzen nachgewiesenes soziales oder kulturelles Engagement den Studien- oder Berufserfolg vorhersagt. Zur prädiktiven Validität von Bewerbungsschreiben gibt es widersprüchliche Ergebnisse, sie werden meist als nicht valide eingestuft [13].

Situational Judgement Tests

In SJTs werden den Testanden papier-, video- oder PC-basiert arbeits- bzw. studi-

enrelevante Situationen dargeboten. Anschließend sollen sie Handlungsoptionen beurteilen [43], ohne dass spezifisches Fachwissen vonnöten ist. Bei den Fragestellungen werden generell der Should-do- und der Would-do-Ansatz unterschieden. Beim ersten gilt es anzugeben, was der Handelnde machen sollte, wobei eher das Wissen um die richtige Reaktion erfasst wird. Beim Would-do-Ansatz wird der Teilnehmer gefragt, was er in der Situation tun würde, wobei die Gefahr einer sozial erwünschten Antwort zunimmt [43].

Auch die Antwortformate lassen von der Auswahl einer richtigen und/oder einer falschen Handlung über das Erstellen von Rangreihen bis hin zu Bewertungen jeder einzelnen Handlungsoption auf mehrstufigen Skalen viel Gestaltungsspielraum. Da sich der Inhalt der Situationen an jeden Bereich anpassen lässt [44], haben sich SJTs in der Personalauswahl weitverbreitet [45] und konnten dort auch ihren Nutzen zur Vorhersage beruflicher Leistungen zeigen [43].

In den letzten Jahren haben sich SJTs im Ausland auch im Bereich der Studienbewerbersauswahl in der Medizin etabliert [13]. In Großbritannien werden Integrität, Perspektivenübernahme und Teamfähigkeit mittels eines SJTs als Teil des UKCAT [46] jährlich bei mehr als 20.000 BewerberInnen gemessen [47]. In Belgien wurde ein SJT zur Messung interpersoneller Fähigkeiten eingesetzt, der mit der Leistung in Studienfächern mit interpersonellen Inhalten zu 0,21 und zur Einschätzung der beruflichen Leistung zu 0,15 korreliert [48]. Die McMaster-Universität in Kanada setzt einen Video-SJT ein, um auf einer Vorauswahlstufe Teamfähigkeit, Kommunikationsfähigkeit, professionelles Handeln und Vertraulichkeit aller BewerberInnen einzuschätzen [49]. Sie konnten moderate Zusammenhänge ($r = 0,30-0,50$) zu inhaltlich verwandten Teilen der Abschlussprüfungen zeigen [50]. Die Universität von Dundee misst Integrität mittels eines SJTs [51], berichtete bisher aber nur Ergebnisse zu Reliabilität und Zusammenhänge zum MMI. Das gestiegene wissenschaftliche Interesse hat bisher noch keine generalisierbaren Aussagen für den medizinischen

Tab. 2 Ausgewählte Dimensionen für die SJT-Entwicklung

1. Kommunikationsfähigkeit
2. Integrität
3. Perspektivenübernahme (Empathie)
4. Teamfähigkeit
5. Effiziente Entscheidungsfindung
6. Kritisches Denken
7. Selbstregulation/Selbstreflexion
8. Problemlösefähigkeit
9. Bewusstsein für ethische Fragen im medizinischen Kontext

Auswahlkontext vergleichbar zur Metaanalyse von McDaniel et al. [43] für den beruflichen Bereich hervorgebracht. Lievens und Motowidlo [45] kritisieren zudem, dass wenig Anstrengung in die Entwicklung theoretischer Erklärungsansätze investiert wird und auch die zu messenden Konstrukte meist nicht klar und eindeutig definiert werden.

Entwicklung von Situational Judgement Tests an deutschen Medizinfakultäten

Auf der Suche nach einer kosten- und personalgünstigeren Alternative zu Interviews müssen im Auswahlkontext immer die Themen Verfälschbarkeit und Trainierbarkeit berücksichtigt werden. Daher sind Selbstberichtsmaße wie Persönlichkeitsfragebögen nicht geeignet. SJTs mit einem Should-do-Fragenformat gelten als Leistungstests, die sich effizient bei großen Personengruppen durchführen lassen, wobei die Trainierbarkeit kontrovers diskutiert wird [13, 52].

Seit 2011 werden an der medizinischen Fakultät in Hamburg verschiedene SJT-Formate untersucht, die bisher nicht auswahlrelevant waren. Den Ausgangspunkt bildete ein papierbasierter SJT, bei dem zwölf typische Situationen aus dem Krankenhaus- und Medizinkontext beschrieben wurden. Die BewerberInnen sollten jeweils aus fünf vorgegebenen Antworten die beste auswählen. Aufgrund zu niedriger Schwierigkeiten, zu geringer Variation in den Antworten und zu geringer Reliabilitätswerte wurde im Folgejahr ein videogestützter SJT mit freiem Antwortformat erprobt. Neben dem enormen zeitlichen Auswer-

tungsaufwand war in Hamburg die Übereinstimmung der beiden unabhängigen JurorInnen so niedrig, dass auch dieser Ansatz nicht weiterverfolgt wurde. Da jedoch die BewerberInnen zurückmeldeten, dass die videobasierte Darstellung ihnen den Zugang zu den Situationen erleichterte, entwickelten wir einen Video-SJT mit einem objektiveren, standardisierteren und weniger zeitaufwendigen Auswertungssystem. Die BewerberInnen sollten auf einer fünfstufigen Likert-Skala die Angemessenheit mehrerer Verhaltensoptionen zur Erreichung vorgegebener Ziele einschätzen. Dieser SJT wurde in zwei aufeinanderfolgenden Jahren mit insgesamt fast 400 StudienbewerberInnen erprobt. Für die Auswertung wurde ein Expertenrating mehrerer PsychologInnen des Uniklinikums ($N = 38$) herangezogen. Die Bewerberleistungen zeigten eine gute Streuung der Schwierigkeiten und gute Varianzen, aber keine zufriedenstellende Reliabilität und konkordante Validität zum ebenfalls im Auswahlverfahren durchgeführten MMI.

In Heidelberg und in Oldenburg wird ein videobasierter SJT zur Messung sozialer Kompetenzen als freiwilliges Online-Self-Assessment mit dem Ziel eingesetzt, Studieninteressierten Informationen zum Medizinstudium zu vermitteln und deren interpersonale Fertigkeiten zu messen [53]. Die TeilnehmerInnen müssen die soziale Angemessenheit von je vier bis sechs Handlungen zu 20 Videos einschätzen, die zwölf auf einer Anforderungsanalyse basierende Kompetenzen abbilden [53]. Verglichen wurden die Bewerberantworten aus Heidelberg mit dem Rating einer Expertengruppe aus 37 ÄrztInnen und 29 klinischen PsychologInnen. Die Ergebnisse zweier Kohorten zeigen zufriedenstellende interne Konsistenzen ($\alpha_{2013} = 0,83$; $\alpha_{2014} = 0,81$) und erste Hinweise zur konvergenten und divergenten Validität. In Oldenburg konnte bisher kein bedeutsamer Zusammenhang zwischen demselben SJT und MMI-Ergebnissen gezeigt werden [54]. Ein in Münster freiwillig durchgeführter SJT war hypothesenkonform unabhängig von Abiturnote und medizin-naturwissenschaftlichem Verständnistest, er korrelierte aber auch nicht mit den Ergebnissen der multiplen Miniaktionen

[54]. Auch die medizinische Hochschule Brandenburg musste feststellen, dass der seit zwei Jahren eingesetzte Paper-Pencil-SJT nicht mit den Einzelinterviews oder der Abiturnote und auch nur schwach mit einer schriftlichen Reflexionsaufgabe zusammenhängt [54].

Standortübergreifende Kooperation zur SJT-Entwicklung

Die oben beschriebenen Erfahrungen zeigen, dass die Entwicklung eines SJTs zur Messung psychosozialer Kompetenzen von StudienbewerberInnen aufgrund der vielen Möglichkeiten der Situationsdarstellung, der Antwortformate und der Auswertungsmethoden sehr komplex und aufwendig ist. Seit Anfang 2016 kooperieren daher im Rahmen der Arbeitsgruppen Studierendenauswahl des Medizinischen Fakultätentages (MFT) und der Gesellschaft für Medizinische Ausbildung (GMA) die sechs medizinischen Fakultäten Göttingen, Hamburg, Heidelberg, Münster, Oldenburg und Witten/Herdecke zur Entwicklung eines fakultätsübergreifenden SJTs. Nach Absprache mit der Work Psychology Group, die den im UKCAT in Großbritannien eingesetzten SJT entwickelt hat, führten wir zunächst eine Literaturrecherche zur Frage: „Welche Kompetenzen braucht ein Arzt in 25 Jahren?“, in der Grundlagenliteratur und in bekannten Rahmenkonzepten wie CanMEDS und dem Nationalen Kompetenzbasierten Lernzielkatalog Medizin (NKLM) durch [55]. Anschließend wurden für die 20 ermittelten Dimensionen die Wichtigkeit und die Messbarkeit durch einen SJT in einer zweistufigen elektronischen Delphi-Befragung durch VertreterInnen von acht Fakultäten bewertet und neun Dimensionen für die SJT-Entwicklung ausgewählt (Tab. 2).

Zu den ausgewählten Dimensionen wurden Interviews mit verschiedenen Stakeholdern der medizinischen Fakultäten geführt und mithilfe der Critical Incident Technique [56] erfolgskritische Ereignisse gesammelt. Nach einem von der Work Psychology Group geleiteten Workshop wurden in den kooperierenden Fakultäten 33 textbasierte Situationen entwickelt, die wiederum von

Situation

Fiona beobachtet seit längerer Zeit, dass ihr Kommilitone Felix den Anschluss im Studium zu verlieren scheint. Er wirkt häufig müde und verpasst Seminare und Vorlesungen. Sie beide gehören zu einer Fünfergruppe, die eine Hausarbeit gemeinsam einreichen muss. Bis wenige Tage vor dem Abgabetermin hat Felix noch nichts beigesteuert und ist auch seit längerer Zeit nicht mehr in den Lehrveranstaltungen aufgetaucht. Die Gruppe erwägt, die Hausarbeit ohne Felix fertigzustellen und dem Dozenten zu sagen, dass Felix keinen Beitrag geleistet hat. Noch am selben Tag sieht Fiona Felix zufällig auf dem Universitätsgelände.

Wie **angemessen** ist jede der folgenden Handlungsoptionen in dieser Situation für Fiona?
 1 = sehr angemessen, 2 = angemessen, 3 = unangemessen, 4 = sehr unangemessen

- a) Felix über die Überlegungen in der Gruppe informieren.
- b) Felix erinnern, dass der Abgabetermin für die Arbeit nächste Woche ist.
- c) So tun, als hätte sie Felix nicht gesehen.
- d) Felix freundlich zunicken und weitergehen.

Abb. 1 ▲ Beispielimitem aus dem Probelauf des Situational Judgement Tests (SJT) 2016 in Hamburg

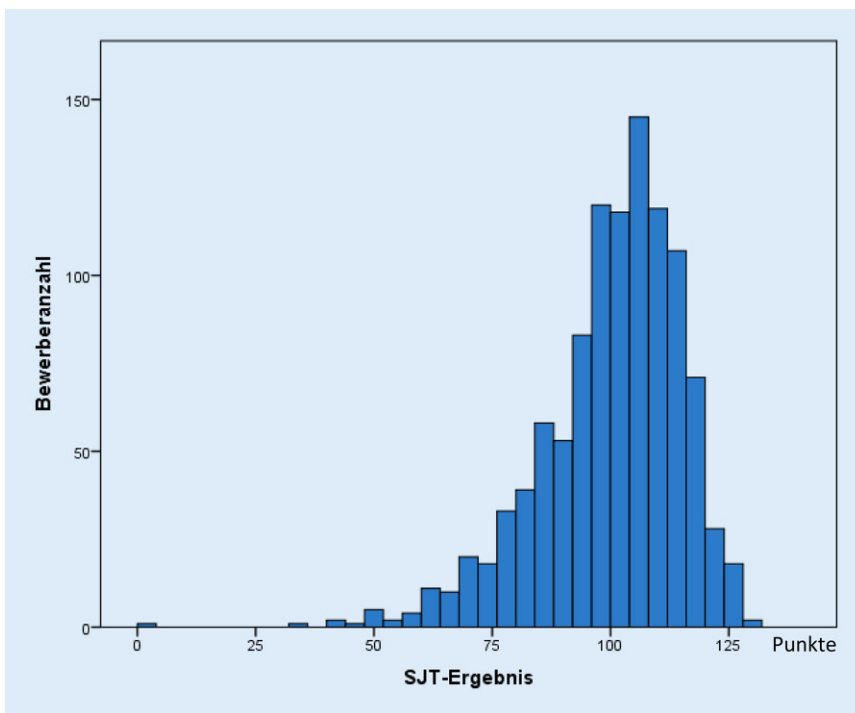


Abb. 2 ▲ Bewertung von 1069 StudienbewerberInnen im Probelauf des Situational Judgement Tests (SJT). Punkteskala: quadrierte Differenz zwischen Bewerber- und Expertenantworten, Mittelwert: 100 (durchschnittliches Ergebnis), Standardabweichung: 15

ExpertInnen hinsichtlich des Verständnisses, der Relevanz, des Realismus, der Fairness und des Schwierigkeitsgrades eingeschätzt und anschließend überarbeitet wurden. Die Angemessenheit der so entwickelten 232 Handlungsoptionen zu 33 Situationen wurde in einem Concordance Panel bestehend aus 29 Mitgliedern der Arbeitsgruppen Studierendenauswahl sowie Lehrenden

und Studierenden der beteiligten Fakultäten bewertet. Anschließend wurden im August 2016 zehn Situationen mit insgesamt 66 Items freiwillig von 1076 Hamburger StudienbewerberInnen bearbeitet (Beispielimitem in [Abb. 1](#)). In der Auswertung wurde für 1069 BewerberInnen (7 Personen hatten die Datenfreigabe nicht unterzeichnet) die quadrierte Abweichung des Urteils vom

Expertenurteil berechnet und in Punkte umgerechnet, wobei eine höhere Punktzahl eine geringere Abweichung und damit eine bessere Leistung bedeutet. Die Häufigkeitsverteilung zeigt, dass der Großteil der BewerberInnen in der Lage war, die Handlungen in Übereinstimmung mit dem Concordance Panel zu bewerten und hohe Punktzahlen zu erreichen, wobei es aber auch hinlänglich viele Personen mit niedrigen Punkten gab ([Abb. 2](#)). Die ersten Untersuchungen zeigen keinen Einfluss des Geschlechts und des Schultyps, aber einen Vorteil für BewerberInnen mit Deutsch als Muttersprache. Der SJT korreliert erwartungsgemäß nicht mit dem zur Studierendenauswahl eingesetzten Verfahren HAM-Nat und der Abiturdurchschnittsnote, deren Leistungen eher auf fachspezifischem Wissen basieren ([Tab. 3](#)). Eine signifikante Korrelation mit dem HAM-Int ($r = 0,22$, $p = 0,003$), das ebenfalls psychosoziale Kompetenzen messen soll, deutet dagegen auf eine konkordante Validität des entwickelten SJTs hin. Untersuchungen zur prädiktiven Validität des SJTs haben begonnen, Ergebnisse liegen aber noch nicht vor, da die Studierenden bisher nur das 1. Studienjahr abgeschlossen haben. Weitere Forschungsanstrengungen sind erforderlich, um einen reliablen und validen SJT für die Studierendenauswahl in Deutschland zu entwickeln.

Fazit

Aktuell steht das Verfahren zur Auswahl von Medizinstudierenden auf dem Prüfstand. Das Bundesverfassungsgericht untersucht die Konformität mit dem Grundrecht auf freie Berufswahl. Die Politik möchte dem Masterplan Medizinstudium 2020 entsprechend die Gewichtung von sozialen Kriterien bei der Auswahlentscheidung stärken und die Gewichtung der immer stärker diskutierten Abiturdurchschnittsnote senken [2]. Abgelehnte StudienbewerberInnen beschwerten sich über die steigende Wartezeit und die Fakultäten klagen, dass über die Wartezeitquote zugelassene Studierende häufiger das Studium abbrechen oder verlängern [1, 27]. StudienbewerberInnen können bei der Vielzahl an Quoten und

Tab. 3 Korrelation der SJT-Ergebnisse aus Hamburg mit weiteren Auswahlkriterien

	Abiturnote (N = 1069)	HAM-Nat (N = 1069)	HAM-Int (N = 178)
<i>SJT</i>	0,03	-0,02	0,22
	<i>p</i> = 0,337	<i>p</i> = 0,576	<i>p</i> = 0,003

HAM-Nat Hamburger Auswahlverfahren für medizinische Studiengänge – Naturwissenschaftsteil,
HAM-Int Hamburger Auswahlverfahren für medizinische Studiengänge – Interviewverfahren

Auswahlverfahren der einzelnen Hochschulen kaum einen Überblick gewinnen, für welche Universität sie sich bewerben sollen, um die persönlichen Chancen zu optimieren.

Vor diesem Hintergrund haben der MFT und die Bundesvertretung der Medizinstudierenden in Deutschland (bvmd e. V.) einen Vorschlag zur Neugestaltung der Studienplatzvergabe erarbeitet, nachdem die Wartezeit- und die Abiturbestenquote abgeschafft werden sollen [57]. Nach diesem Entwurf können die BewerberInnen in unterschiedlichen Kategorien Punkte sammeln: maximal 40 Punkte für die Abiturnote, bis zu 40 Punkte für einen Test kognitiver Fähigkeiten (vorgeschlagen wird eine Kombination des TMS und des HAM-Nat), 10 Punkte für eine einjährige berufspraktische Erfahrung und 10 Punkte für einen bundesweit einheitlichen SJT zur Erfassung sozialer Kompetenzen. Da für Abiturnote und Eignungstests der Forschungsstand zur prädiktiven Validität deutlich besser ist, sollen diese mit der stärksten Gewichtung eingehen. Berufserfahrung und SJT sollen die im Masterplan Medizinstudium 2020 geforderten psychosozialen Kompetenzen abbilden, wobei der Nutzen eines SJTs auch vor dem Hintergrund der hier berichteten Ergebnisse nur vorsichtig positiv einzuschätzen ist. Die Hälfte der Studienplätze soll direkt entsprechend der Punktesumme vergeben werden. Für die andere Hälfte der Studienplätze sollen die Fakultäten die verbleibenden BewerberInnen wiederum aufgrund der Punktesumme zu einem ihrem Profil entsprechenden universitären Auswahlverfahren einladen können, das dann auch weitere Tests zur Ermittlung der psychosozialen und kommunikativen Kompetenzen, wie z. B. Multiple Mini-Interviews, umfassen kann.

In diesem Modell ist das Gewicht der Abiturnote zugunsten anderer Kriterien

abgesenkt, damit auch BewerberInnen mit schlechteren Abiturnoten eine Chance auf Zulassung haben und so dem Grundrecht auf freie Berufswahl entsprochen wird. Nach einem Urteil des Bundesverfassungsgerichts könnten die Bundes- und Landesgesetze weiterentwickelt werden, um die aktuellen Herausforderungen der Medizinstudierendenauswahl zu meistern. Auch innerhalb der Nicht-EU-Ausländerquote und, falls sie eingeführt wird, der Landarztquote sollte die Studienplatzvergabe nach demselben Punktesystem erfolgen, um auch hier eine objektive und faire Auswahl zu treffen. Eine Erhöhung des Anteils anderer als kognitiver Auswahlkriterien kann zudem den Zugang für unterrepräsentierte Minderheiten erleichtern [58]. Unter den Konzepten Widening Participation und Diversity bemühen sich vor allem ausländische Fakultäten um MedizinbewerberInnen, die aufgrund ihres sozioökonomischen und schulischen Hintergrundes mit kognitiven Verfahren nur geringe Chancen auf einen Studiengang hätten.

Korrespondenzadresse

A. Schwibbe

Arbeitsgruppe Auswahlverfahren, Institut für Biochemie und Molekulare Zellbiologie, Zentrum für Experimentelle Medizin, Universitätsklinikum Hamburg Eppendorf Martinistr. 52, 20246 Hamburg, Deutschland a.schwibbe@uke.de

Danksagung. Die AutorInnen bedanken sich bei den Mitarbeitenden der medizinischen Fakultäten Göttingen, Hamburg, Heidelberg, Münster, Oldenburg und Witten/Herdecke und den Mitgliedern der Arbeitsgruppen Studierendenauswahl des MFTs und der GMA, die durch ihre Teilnahme an der Delphi-Umfrage, den Interviews zur Sammlung der kritischen Ereignisse, dem Itementwicklungsworkshop

und dem Concordance Panel an der Entwicklung des SJTs mitgewirkt haben.

Einhaltung ethischer Richtlinien

Interessenkonflikt. A. Schwibbe, J. Lackamp, M. Knorr, J. Hissbach und W. Hampe sind an der Konzeption und Durchführung des Hamburger Auswahlverfahrens beteiligt. M. Kadmon ist an der Durchführung des Tests für medizinische Studiengänge beteiligt.

Die Ethik-Kommission der Ärztekammer Hamburg, PV4983 hat uns bestätigt, dass unsere Auswahlforschung im Allgemeinen „kein Forschungsvorhaben am Menschen“ darstellt.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Heidmann J, Schwibbe A, Kadmon M, Hampe W (2016) Warten aufs Medizinstudium: Sieben lange Jahre. *Dtsch Arztebl Int* 113:1636–1638
- Bundesgesundheitsministerium (2017) Masterplan Medizinstudium 2020. https://www.bmbf.de/files/2017-03-31_Masterplan%20Beschlussstext.pdf. Zugegriffen: 22. Aug. 2017
- Grobler L, Marais BJ, Mabunda S (2015) Interventions for increasing the proportion of health professionals practising in rural and other underserved areas. *Cochrane Database Syst Rev*. <https://doi.org/10.1002/14651858.CD005314.pub3>
- Hampe W, Hissbach J, Kadmon M, Kadmon G, Klusmann D, Scheutzel P (2009) Wer wird ein guter Arzt? Verfahren zur Auswahl von Studierenden der Human- und Zahnmedizin. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 52(8):821–830
- Becker-Carus C, Wendt M (2017) Psychologie als Wissenschaft. In: Becker-Carus C, Wendt M (Hrsg) *Allgemeine Psychologie Eine Einführung*. Springer, Berlin, S 2–30
- McManus IC, Dewberry C, Nicholson S, Dowell JS (2013) The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. *BMC Med* 11:244
- Hutter J (2004) Kompetenzfeststellungen – Expertise inklusive eines Handlungsleitfadens. Bundesinstitut für Berufsbildung, Bonn
- Hell B, Trapmann S, Schuler H (2007) Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empir Pädagog* 21:251–270
- McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HW (2013) Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student

- selection: meta-regression of six UK longitudinal studies. *BMC Med* 11:243
10. McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR (2003) A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ* 327:139–142
 11. Gesetze im Internet (2017) Hochschulrahmengesetz in der Fassung der Bekanntmachung vom 19. Januar 1999 (BGBl. I S. 18), das zuletzt durch Artikel 6 Absatz 2 des Gesetzes vom 23. Mai 2017 (BGBl. I S. 1228) geändert worden ist. <https://www.gesetze-im-internet.de/hrg/HRG.pdf>. Zugegriffen: 22. Aug. 2017
 12. Trapmann S, Hell B, Weigand S, Schuler H (2007) Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Z Pädagog Psychol* 21:11–27
 13. Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J (2016) How effective are selection methods in medical education? A systematic review. *Med Educ* 50:36–60
 14. Rindermann H, Oubaid V (1999) Auswahl von Studienanfängern durch Universitäten – Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. *Z Differ Diagnost Psychol* 20:172–191
 15. McManus C, Woolf K, Dacre JE (2008) Even one star at A level could be “too little, too late” for medical student selection. *BMC Med Educ* 8:16
 16. Wissenschaftsrat (2004) Empfehlungen zur Reform des Hochschulzugangs 2004. <http://www.wissenschaftsrat.de/texte/5920-04.pdf>. Zugegriffen: 22. Aug. 2017
 17. Tent L (2006) Zensuren. In: Rost DH (Hrsg) *Handwörterbuch Pädagogische Psychologie*. Beltz, Weinheim, S 873–880
 18. Wedler B, Troche S, Rammsayer T (2008) Studierendenauswahl – Eignungsdiagnostischer Nutzen von Noten aus Schule und Studium. *Psychol Rundsch* 59:123–125
 19. Sartania N, McClure JD, Sweeting H, Browitt A (2014) Predictive power of UKCAT and other pre-admission measures for performance in a medical school in Glasgow: a cohort study. *BMC Med Educ* 14:116
 20. Gold A, Souvignier E (2005) Prognose der Studierfähigkeit. *Z Entwicklungspsychol Pädagog Psychol* 37:214–222
 21. Cappel S, Kadmon G, Kadmon M (2015) Test für Medizinische Studiengänge: Der TMS im Aufwind. *Dtsch Arztebl Int* 112:1618
 22. Kadmon G, Kirchner A, Duelli R, Resch F, Kadmon M (2012) Warum der Test für Medizinische Studiengänge (TMS)? *Z Evid Fortbild Qual Gesundheitswes* 106:125–130
 23. Hissbach J, Feddersen L, Sehner S, Hampe W (2012) Eignung von HAM-Nat und TMS-Modul „Medizinisch-naturwissenschaftliches Grundverständnis“ für die Studienbewerberauswahl in der Medizin. *GMS Z Med Ausbild* 29:Doc72
 24. Hissbach JC, Klusmann D, Hampe W (2011) Dimensionality and predictive validity of the HAM-Nat, a test of natural sciences for medical school admission. *BMC Med Educ* 11:83
 25. Werwick K, Winkler-Stuck K, Hampe W, Albrecht P, Robra BP (2015) Introduction of the HAM-Nat examination – applicants and students admitted to the Medical Faculty in 2012–2014. *GMS Z Med Ausbild* 32:Doc53
 26. Kadmon G, Kadmon M (2011) Medizinertest: Chance für Bewerber ohne Einser-Abitur. *Dtsch Arztebl Int* 108:2388–2392
 27. Kadmon G, Kadmon M (2016) Studienleistung von Studierenden mit den besten versus mittelmäßigen Abiturnoten: Gleicht der Test für Medizinische Studiengänge (TMS) ihre Prognosen aus? *GMS Z Med Ausbild* 33:Doc7
 28. Donnon T, Paolucci EO, Violato C (2007) The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med* 82:100–106
 29. UKCAT (2016) Annual report and financial statements for the year ended 31 March 2016. <https://www.ukcat.ac.uk/media/1143/annual-report-2015-2016.pdf>. Zugegriffen: 22. Aug. 2017
 30. Lynch B, Mackenzie R, Dowell J, Cleland J, Prescott G (2009) Does the UKCAT predict Year 1 performance in medical school? *Med Educ* 43:1203–1209
 31. Yates J, James D (2013) The UK Clinical Aptitude Test and clinical course performance at Nottingham: a prospective cohort study. *BMC Med Educ* 13:32
 32. Harris BH, Walsh JL, Lammy S (2015) UK medical selection: lottery or meritocracy? *Clin Med (Lond)* 15:40–46
 33. McManus IC, Ferguson E, Wakeford R, Powis D, James D (2011) Predictive validity of the biomedical admissions test: an evaluation and case study. *Med Teach* 33:53–57
 34. Emery JL, Bell JF (2009) The predictive validity of the BioMedical Admissions Test for pre-clinical examination performance. *Med Educ* 43:557–564
 35. Groves MA, Gordon J, Ryan G (2007) Entry tests for graduate medical programs: is it time to re-think? *Med J Aust* 186:120–123
 36. Niessen AS, Meijer RR (2016) Selection of medical students on the basis of non-academic skills: is it worth the trouble? *Clin Med (Lond)* 16:339–342
 37. Levashina J, Hartwell CJ, Morgeson FP, Campion MA (2014) The structured employment interview: narrative and quantitative review of the research literature. *Pers Psychol* 67:241–293
 38. Eva KW, Rosenfeld J, Reiter HI, Norman GR (2004) An admissions OSCE: the multiple mini-interview. *Med Educ* 38:314–326
 39. Knorr M, Hissbach J (2014) Multiple mini-interviews: same concept, different approaches. *Med Educ* 48:1157–1175
 40. Eva KW (2003) On the generality of specificity. *Med Educ* 37:587–588
 41. Hissbach JC, Sehner S, Harendza S, Hampe W (2014) Cutting costs of multiple mini-interviews – changes in reliability and efficiency of the Hamburg medical school admission test between two applications. *BMC Med Educ* 14:54
 42. Hampe W, Klusmann D, Buhk H, Münch-Harrach D, Harendza S (2008) Reduzierbarkeit der Abbrecherquote im Humanmedizinstudium durch das Hamburger Auswahlverfahren für Medizinische Studiengänge – Naturwissenschaftsteil (HAM-Nat). *GMS Z Med Ausbild* 25:Doc82
 43. McDaniel MA, Hartman NS, Whetzel DL, Grubb WL (2007) Situational judgement tests, response, instructions, and validity: a meta-analysis. *Pers Psychol* 60:63–91
 44. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P (2012) Evaluations of situational judgement tests to assess non-academic attributes in selection. *Med Educ* 46:850–868
 45. Lievens F, Motowidlo SJ (2016) Situational judgment tests: from measures of situational judgment to measures of general domain knowledge. *Ind Organ Psychol* 9:3–22
 46. Patterson F, Cousans F, Edwards H, Rosselli A, Nicholson S, Wright B (2017) The predictive validity of a text-based situational judgment test in undergraduate medical and dental school admissions. *Acad Med* 92:1250–1253
 47. Consortium U (2017) Test statistics released. <https://www.ukcat.ac.uk/news/2017-test-statistics-released/>. Zugegriffen: 12. Okt. 2017
 48. Lievens F (2013) Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Med Educ* 47:182–189
 49. Dore KL, Reiter HI, Eva KW et al (2009) Extending the interview to all medical school candidates – Computer-Based Multiple Sample Evaluation of Noncognitive Skills (CMSENS). *Acad Med* 84:S9–S12
 50. Dore KL, Reiter HI, Kreuger S, Norman GR (2017) CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. *Adv Health Sci Educ Theory Pract* 22:327–336
 51. Husbands A, Rodgerson MJ, Dowell J, Patterson F (2015) Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC Med Educ* 15:144
 52. Rostom H, Watson R, Leaver L (2013) Situational judgement tests: the role of coaching. *Med Educ* 47:219
 53. Fröhlich M, Kahmann J, Kadmon M (2017) Development and psychometric examination of a German video-based situational judgment test for social competencies in medical school applicants. *Int J Sel Assess* 25:94–110
 54. Hampe W, Hissbach J, Kadmon M (2017) Medizinstudium: Sozial kompetente Bewerber. *Dtsch Arztebl Int* 114:1478–1479
 55. MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e.V. (2015) Nationaler Kompetenzbasierter Lernzielkatalog Medizin. http://www.nklm.de/files/nklm_final_2015-07-03.pdf. Zugegriffen: 22. Aug. 2017
 56. Flanagan JC (1954) The critical incident technique. *Psychol Bull* 51:327–358
 57. Bundesvertretung der Medizinstudierenden in Deutschland e.V. (bvmd), Medizinischer Fakultätentag e.V. (MFT) (2017) Vorschlag für ein neues Modell der Studierendenauswahl in der Medizin. http://www.mft-online.de/files/vorschlag_f_r_ein_neues_auswahlverfahren_von_mft_und_bvmd.pdf. Zugegriffen: 22. Aug. 2017
 58. Ballejos MP, Rhyne RL, Parkes J (2015) Increasing the relative weight of noncognitive admission criteria improves underrepresented minority admission rates to medical school. *Teach Learn Med* 27:155–162