
Aus dem Studiendekanat der Medizinischen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Kompetenzorientierte Prüfungen und
Lehr-/Lernklima an einer Medizinischen Fakultät
—
Diskussionsbeitrag zur Entwicklung
kompetenzorientierter Curricula

Habilitationsschrift zur Erlangung der Venia legendi für das Fach Medizindidaktik
an der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Thomas Rotthoff

Düsseldorf, 2017

Inhaltsverzeichnis

1	Einleitung.....	5
1.1	Forderung nach kompetenzorientierter Ausbildung.....	5
1.2	Definition des Kompetenzbegriffs in der medizinischen Ausbildung.....	6
1.3	Entwicklung der Outcome bzw. Competency Based Education.....	8
1.4	Evidenz für die Effektivität der outcome- bzw. kompetenz-orientierten Ausbildung .	11
1.5	Prüfung von Kompetenzen - wie können Kompetenzen messbar gemacht werden?...	12
1.6	Operationalisierung von Kompetenzen.....	14
1.7	Einfluss des Lehr-/Lernklimas auf die studentischen (Prüfungs-) Leistungen und die Feedbackkultur in einem kompetenzbasierten Curriculum.....	17
1.8	Feedback und Lehr-/Lernklima.....	18
2	Zielsetzung der Arbeit.....	19
2.1	Schaffung von Evidenz für die Validität und Reliabilität von kompetenzorientierten Prüfungen	19
2.2	Entwicklung und Validierung von Messinstrumenten zur Erfassung des Lehr-/Lernklimas.....	19
3	Eigene Untersuchungen	22
3.1	Schaffung von Evidenz für Validität und Reliabilität von kompetenz-orientierten Prüfungen.	22
3.1.1	Verknüpfung des gemessenen Wissens und/oder der gemessenen Fertigkeiten mit definierten Kompetenzen.....	22
3.1.2	Beziehung der Messung einer Kompetenz in verschiedenen Messverfahren und prädiktive Validität von Kompetenzmessungen.....	30
3.2	Erstellung und Validierung von Messinstrumenten zur Erfassung des Lehr- und Lernklimas.....	35
4	Diskussion.....	44
4.1	Geschaffene Evidenz zur Validität und Reliabilität von kompetenzorientierten Prüfungen	44
4.2	Lösungsansätze für die integrative Prüfung von Kompetenzen.....	53
4.3	Bedeutung und Messung des Lehr-/Lernklimas	57
5	Zusammenfassung.....	60
6	Literatur.....	64
7	Anhang.....	77
7.1	Danksagung	77
7.2	Verzeichnis der wissenschaftlichen Veröffentlichungen und Vorträge	78
7.3	Dieser Habilitationsschrift wesentlich zu Grunde liegende Originalarbeiten	86

Erklärungen

Hiermit erkläre ich ehrenwörtlich, dass bei den wissenschaftlichen Untersuchungen, die Gegenstand der schriftlichen Habilitationsordnung sind, ethische Grundsätze und die jeweils gültigen Empfehlungen zur Sicherung guter wissenschaftlicher Praxis beachtet wurden.

Diese Habilitationsschrift wurde weder in gleicher noch in ähnlicher Form in einem anderen Prüfungsverfahren vorgelegt. Außerdem erkläre ich, dass ich bisher noch keine weiteren Habilitationsverfahren eingeleitet oder erfolglos beendet habe.

Düsseldorf, den 2. April 2017

Abkürzungen

AC	Assessment Center
APT	Anvertraubare Professionelle Tätigkeiten
ÄAppO	Ärztliche Approbationsordnung
BGR	Berliner Global Rating
CBE	Competency Based Education
CoMeD	Communication in Medical Education Düsseldorf
DREEM	Dundee Ready Education Environment Measure
ebd.	ebenda
EPA	Entrustable Professional Activities
OBE	Outcome Based Education
OSCE	Objective Structured Clinical Examination
NKLM	Nationaler Kompetenzbasierter Lernzielkatalog Medizin
UCAN	Umbrella Consortium for Assessment Networks
USMLE	United States Medical Licensing Examination
WR	Wissenschaftsrat

1 Einleitung

1.1 Forderung nach kompetenzorientierter Ausbildung

„Die ärztliche Ausbildung erfordert auch künftig ein solides Fundament aus grundlegendem medizinischem, psychosozialen und naturwissenschaftlichem Wissen. An die Stelle der traditionellen Orientierung an Fächern sollte eine an den ärztlichen Rollen und ihren Kompetenzen orientierte Ausbildung treten“ (Wissenschaftsrat 2014 – Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland).

Der Wissenschaftsrat sieht den „Wechsel des Paradigmas von einer inputorientierten Wissensvermittlung“ hin zu einer „Fokussierung auf den Outcome der Studierenden“ dabei als wesentliches Element zur Weiterentwicklung des Medizinstudiums in Deutschland an (ebd. S.70). ‚Verbunden mit der Kompetenzorientierung sind (gemäß der Empfehlung) die Förderung einer studierendenzentrierten Lehre sowie eine Akzentuierung von Fertigkeiten und Haltungen in Kombination zum Wissen – also eine Stärkung des berufsspezifischen, praxisorientierten Spektrums des Medizinstudiums‘ (ebd. S.70). Diese Empfehlungen basieren auf den Betrachtungen der deutschen Modell- und reformierten Regelstudiengänge, den „Erfahrungen des Wissenschaftsrates aus seinen Begutachtungen universitätsmedizinischer Standorte der vergangenen Jahre sowie auf aktuellen nationalen und internationalen Diskussionen, internationalen „Standards“ und in anderen Ländern gewonnene Erfahrungen“ (ebd. S.6).

Wahrgenommene Defizite in der Medizinischen Ausbildung

National und international werden Defizite in der medizinischen Aus- und Weiterbildung mit Blick auf die Anforderungen der heutigen Gesundheitssysteme wahrgenommen. Exemplarisch seien hier eine bereits im Jahr 2008 in Deutschland durchgeführte Befragung von Absolventinnen und Absolventen für das Fach Medizin genannt, nach der diese ihr Niveau zum Studienabschluss hinsichtlich vieler Kompetenzen als niedriger einschätzten als im Beruf gefordert wurde. Besonders deutlich betraf dies die Fähigkeiten zur ärztlichen Gesprächsführung, die effektive Kommunikation im Team, die Erstellung differentialdiagnostischer Überlegungen oder Therapieplanerstellung anhand von Leitlinien sowie mangelnde Fähigkeiten bezüglich Strategien zum Zeitmanagement (Janson 2010). Auch eine Befragung von Verantwortungsträgern im Gesundheitssystem der USA im Jahr 2011 ergab wahrgenommene Defizite bei Ärztinnen und Ärzten z.B. in den Bereichen Teamarbeit, kommunikative Fähigkeiten und Patientensicherheit (vgl. Combes 2014). Die Kompetenzorientierung wird auch als Möglichkeit angesehen, traditionelle Grenzen zwischen

verschiedenen Professionen in den Gesundheitsbereichen zu überwinden (Frenk et al. 2010; Norman 2006; Hawkins et al. 2015).

Steigerung der Nachhaltigkeit durch Kompetenzorientierung

Der Wissenschaftsrat kommt zu der Einschätzung, dass „eine kompetenzorientierte Ausbildung durch die Verknüpfung von Wissen, praktischen Fertigkeiten und Haltungen im Lernprozess eine größere Nachhaltigkeit der vermittelten Lerninhalte sowie eine Konzentration auf das übergeordnete Ausbildungsziel ermöglicht“ (WR 2014 S. 22). Übergeordnetes Ziel der ärztlichen Ausbildung ist gemäß der aktuell gültigen Approbationsordnung (ÄAppO) „der wissenschaftlich und praktisch in der Medizin ausgebildete Arzt, der zur eigenverantwortlichen und selbständigen ärztlichen Berufsausübung, zur Weiterbildung und zu ständiger Fortbildung befähigt ist.“ (ÄAppO 2002).

Die Erkenntnis des WR resultiert wesentlich aus Aussagen von Übersichtsarbeiten, welche wiederum auf den Originalarbeiten von Caraccio et al. 2002; Carraccio & Englander 2013 und Frank et al. 2010 aufsetzen (WR 2014 S. 70).

Der WR stellt seine Empfehlungen auch in den Kontext zum Nationalen Kompetenzbasierten Lernzielkatalog Medizin (NKLM) (WR 2014 S. 70), der sich zum Zeitpunkt der Erstellung der Empfehlungen noch in Entwicklung befand und seit dem 1.7.2015 publiziert ist (NKLM 2015). Der NKLM „beschreibt das Absolventenprofil von Ärztinnen / Ärzten im Sinne eines Kerncurriculums für das Studium der Medizin“ (ebd. S. 6) und orientiert sich dabei an den gesetzlichen Vorgaben der ärztlichen Approbationsordnung (ÄAppO). Auch gemäß des Masterplans Medizinstudium 2020¹ soll das Medizinstudium künftig ‚konsequent kompetenzorientiert ausgerichtet‘ sein und der weiterentwickelte NKLM dabei "verbindlicher Bestandteil der Approbationsordnung" werden. (Beschlusstext - Masterplan Medizinstudium 2020 vom 31.3.2017)

1.2 Definition des Kompetenzbegriffs in der medizinischen Ausbildung

Im NKLM ist „Kompetenz“ definiert als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die

¹ Auf Basis des Koalitionsvertrages von CDU und SPD wurde im Frühjahr 2015 eine Konferenz von Gesundheitsminister- und Wissenschaftsminister des Bundes und der Bundesländer initiiert, die einen Masterplan Medizinstudium 2020 entwickeln sollte. Der Beschlusstext wurde in einer gemeinsamen Pressemitteilung des Bundesministeriums für Gesundheit und des Bundesministeriums für Bildung und Forschung am 31.3.2017 veröffentlicht (www.bundesgesundheitsministerium.de). Eine Expertenkommission wird nachfolgend innerhalb eines Jahres eine Änderung der Approbationsordnung erarbeiten.

damit verbundenen motivationalen, volitionalen² und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“; und geht zurück auf Weinert (Weinert 2002). Eine kompetenzorientierte Ausbildung wird im internationalen Sprachgebrauch als „Competency Based Education“ (CBE) oder „Outcome Based Education“ (OBE) bezeichnet. Die Begriffe werden in der Literatur und international dabei nicht einheitlich, häufig auch synonym genutzt. Beide Konzepte haben zum Ziel, die (medizinische) Ausbildung an zuvor definierten Zielen auszurichten. International einheitliche und allgemein anerkannte Definitionen existieren für diese Begriffe bis heute nicht, jedoch gibt es Versuche, den Begriff der „Competency Based Medical Education“ zu definieren: „*Competency-based education (CBE) is an approach to preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and organized around competencies derived from an analysis of societal and patient needs. It deemphasizes time-based training and promises greater accountability, flexibility, and learner-centredness.*“ (Frank et al. 2010). Nach einer Definition von Albanese et al. (2008) besteht der Unterschied zwischen einer „Outcome Based Education“ und „Competency Based Education“ darin, dass der Begriff „Outcome“ die Fähigkeiten und Qualitäten definiert, die wir uns von den Studierenden erwarten, wohingegen Kompetenzen festlegen, welche Fähigkeiten und Qualitäten Ärztinnen und Ärzte für die Behandlung von Patienten benötigen (Albanese et al. 2008).

Gesellschaftliche Bedarfe als Grundlage einer kompetenzorientierten Ausbildung

Frenk et al. (2010) machen die Bedarfe der Gesellschaft an die Gesundheitsversorgung zum Bezugspunkt einer kompetenzorientierten Ausbildung, wobei diese Bedarfe die Kompetenzen und letztlich auch Outcomes definieren, welche die Konzeption von Curricula mit ihren Prüfungsinstrumenten steuern. Neben berufsspezifischen Kompetenzen (z.B. diagnostische Kompetenz), gelten übergeordnete Kompetenzen, wie beispielsweise die Kommunikative Kompetenz, als essenziell für eine optimale spätere Berufsausübung (van der Vleuten & Dannefer 2012). Bei der ärztlichen Tätigkeit geht es um die Fähigkeit, verschiedene Kompetenzen für eine optimale Patientenversorgung situationsbezogen zu integrieren (Ten Cate et al. 2010). Kompetenz ist somit immer im Kontext eines spezifischen Handlungsumfeldes zu betrachten. In der medizinischen Ausbildung soll die Kompetenzentwicklung von Beginn an kontinuierlich rückgemeldet, gefördert werden und soweit wie möglich in das spätere Arbeitsumfeld eingebunden sein (Holmboe et al. 2010).

Die mit der Kompetenzorientierung intendierte Stärkung des „berufsspezifisch, praxisorientierten Spektrums des Medizinstudiums“ (WR 2014, S. 70) ist argumentativ zunächst nachvollziehbar.

² Bedeutung: vom Willen bestimmt

Gleichzeitig stellt sich die Frage, welche Evidenz es für die Wirksamkeit oder den Erfolg einer solche Kompetenzorientierung gibt. Beginnend mit einem Blick auf die historische Entwicklung der „Outcome based Education“ soll dieser Frage in den nächsten drei Kapiteln nachgegangen werden.

1.3 Entwicklung der Outcome bzw. Competency Based Education

Die Grundidee der outcome- oder kompetenzbasierten Ausbildung basierte auf der Formulierung expliziter Lernziele und geht zurück auf Ralph W. Tyler (Tyler 1949). Von seinem Schüler Benjamin Bloom wurde der Ansatz dann als die Bloom´sche Taxonomie für Wissen, Fertigkeiten und Haltungen weiterentwickelt (Bloom 1956).

Die Konzepte von Tyler und Bloom basieren wesentlich auf der Theorie des Behaviorismus, welche durch die Arbeiten experimenteller Psychologen wie Watson, Pavlov, Thorndike und Skinner repräsentiert wurden (Morcke et al. 2013). Bereits in den 1970er Jahren wirkte das US-amerikanische Ministerium für Gesundheit, Erziehung und Soziales darauf hin, dass einige Colleges und Universitäten kompetenzbasierte Curricula entwickelten (Ashworth PD & Saxton J 1990) und ihre Curricula an Lernzielen ausrichteten (Morcke et al. 2013; Mager 1994).

Kompetenzorientierung zur Individualisierung des Trainings

Der zentrale Ansatz einer kompetenzorientierten Ausrichtung beruht dabei auf der Annahme, dass Berufsrollen in einzelne Elemente aus definiertem Wissen oder Fertigkeiten zerlegt werden können, welche - separat erworben - dann zur umfassenden Kompetenz führen (Brightwell & Grand 2013). Im Vereinigten Königreich wurden 1986 sog. NVQs (National Vocational Qualifications) eingeführt, welche sich an Kompetenzen ausrichteten, die jeweils Bestandteil verschiedener Berufsrollen waren (Hyland 1993). Dieser Bewegung lag wesentlich die Intention zugrunde, den Arbeitsmarkt in einer zunehmend globalisierten Welt wettbewerbsfähig zu halten (Leung 2002). Die Vorteile einer Ausrichtung an Kompetenzen wurden in der Individualisierung des Trainings und der Transparenz über die Definition zu erreichender Ziele gesehen (ebd.). Über eine Competency Based Education (CBE) sollen Absolventinnen und Absolventen bestmöglich auf die Berufsanforderungen von heute vorbereitet werden. Wie bereits Frank et al. 2010 in ihrer Definition von CBE die abnehmende Bedeutung der Ausbildungsdauer aufnehmen (Frank et al 2010), so wird auch nach einer Analyse von Brightwell und Grand (2013) CBE am Fortschritt des individuellen Kompetenzerwerbs festgemacht und weniger an der rein zeitlichen Dauer einzelner Trainingseinheiten. Hiernach könnten Studierende mit guter Performanz schneller im Studium vorankommen als Studierende mit unzureichender Performanz (Brightwell & Grand 2013). In

Europa ist ein solches Vorgehen für das Medizinstudium aktuell nicht möglich, da der zeitliche Umfang des Medizinstudiums durch eine EU-Directive reglementiert ist, wodurch ein Zeitrahmen von insgesamt sechs Jahren und 5500 Stunden vorgegeben ist (Richtlinie 2005/36/EG über die Anerkennung von Berufsqualifikationen). Bis heute existieren auch keine „best practice“ Beispiele oder Modelle, wie solch individualisierte Ausbildungswege aussehen können (Hawkins et al. 2015).

Kritik an der Outcome bzw. Competency Based Education

Nach Weinert beinhaltet der Kompetenzbegriff die erfolgreiche Problemlösung in variablen Situationen (Weinert 2002). Gerade die Inhalt- und Kontextspezifität des klinischen Handelns in variablen Situationen werden jedoch auch als Widerspruch zu einem isolierten Training verschiedener Kompetenzen angeführt, da Studierende beispielweise in einem Fall gut und in einem anderen Fall mit abweichenden Kontext eine schlechte Performanz zeigen können (Lurie 2012). Die Zergliederung des Handelns in kleinere Häppchen entlang von einzelnen Lernzielen, ermöglicht in der Einschätzung von Ende und Davidoff (1992) keine Aussage darüber, ob und wann von diesen kleinen Elementen in realen (variablen) klinischen Situationen Gebrauch gemacht werden kann (Ende & Davidoff 1992). In der Ausrichtung auf die Ziele und weniger auf das tatsächliche Lehren wird außerdem eine Limitierung der curricularen Inhalte und der Lehre auf ausschließlich definierbare und messbare Inhalte gesehen (Winter 1995; Norman 2006; Malone & Supri 2012). Einige Autoren sehen darin die Gefahr einer Atomisierung der Lerninhalte, eine Schwerpunktsetzung auf einfache routinierte Fertigkeiten und ein „teaching-to-the-test“ (Hodges & Lingard 2012 S.6; Tekian et al. 2015). Lehre und Lernen - so die Kritiker - laufen damit Gefahr, oberflächlich zu werden, indem lediglich ein Minimalstandard angestrebt wird, anstatt sich kritisch hinterfragend um ein umfänglicheres Verständnis des Stoffes zu bemühen (Malone & Supri 2012). Lernziele für Wissen und einfache Fertigkeiten können sinnvoll sein, wohingegen Werte und Haltungen wie affektives, soziales, kulturelles oder ethisches Lernen aus Erfahrungen im Verlauf des Lernprozesses resultiert und mit einer Outcome Based Education als nur schwer abbildbar und insbesondere auch als nicht überprüfbar angesehen wurden (Stenhouse 1975). Erridge & Perry (1994) sind der Ansicht, dass eine „Competence Based Education“ sich besser für den Bereich des niedrig qualifizierten Sektors eignet, bei dem einfachere routinierte Aufgaben und einfach zu überprüfende Fertigkeiten verlangt werden. Für hochqualifizierte Berufe, wie dem des Arztes / der Ärztin, wird das Konzept dagegen als nur wenig geeignet angesehen, da hierfür hochkomplexe Fähigkeiten wie Analyse, Urteils- und Reflexionsvermögen, Professionalität und Empathie erforderlich sind, die mit einem rein behavioristischen Ansatz nicht hinlänglich dargestellt werden können (Huddle & Heudebert 2007; Malon & Supri 2012; Brightwell & Grant 2013).

Ungeachtet der Kritiken begann in den 1980er Jahren mit den Arbeiten von William Spady in den USA eine Renaissance der Outcome Based Education (OBE). Spady versuchte das Problem dadurch zu lösen, dass er Haltungen als eine Voraussetzung für Outcomes ansah und nicht als Outcomes selber. Hierdurch wurden Haltungen aus der OBE ausgeklammert (Spady 1988; Spady 1994).

Für die Medizin begann die OBE Orientierung maßgeblich 1999 mit der Publikation von Harden et al., die auf den Arbeiten von Spady aufsetzte (Morcke et al. 2013; Harden 1999). Nach Harden erfolgt die Gestaltung eines kompetenzbasierten Curriculums demnach auf Basis definierter Ziele für Wissen, Fertigkeiten und Haltungen und beinhaltet passende Prüfungsformate sowie für alle Studierenden eine individuelle und regelmäßige Beurteilung des Fortschrittes im Kompetenzerwerb. Dabei sollen neben fachlichen Inhalten auch übergeordnete Aspekte wie beispielsweise kommunikative Fähigkeiten und ethische Herangehensweise in Bezug auf den Arztberuf berücksichtigt werden (Harden 1999). Auch in dieser Definition fehlt die Trennschärfe zwischen outcome- und kompetenzbasierter Ausbildung.

Das Jahr 1999 war darüber hinaus bedeutsam für die zunehmende Verwendung und Akzeptanz der Outcome bzw. Competence Based Education (Albanese et al. 2008), da weitere Rahmenkonzepte wie „The Medical School Objectives Project (MSOP)“ der Association of American Medical Colleges (AAMC) sowie des Accreditation Council for Graduate Medical Education (ACGME) und American Board of Medical Specialties vorgelegt wurden. Letztere fokussierte auf die Definition von sechs Kompetenzen für den Zertifizierungsprozess in der Postgraduierenausbildung und sollte gleichzeitig als Orientierung und Richtlinie für die Ausbildungsprogramme von Ärztinnen und Ärzten dienen. Ein Jahr zuvor wurde bereits das CanMEDs Projekt in Canada publiziert (Neufeld et al. 1998). Dieses von der Society of Physicians and Surgeons of Canada ursprünglich für die Postgraduierenausbildung entwickelte und implementierte Modell beschreibt mit sieben Berufsrollen Fähigkeiten, die Ärzte benötigen, um die Bedürfnisse der zu versorgenden Menschen effektiv erfüllen zu können. Die (inzwischen modifizierten) Rollen sind: Medical Expert (Medizinischer Experte), Communicator (Kommunikator), Collaborator (Interprofessioneller Partner), Leader (*normals Manager*) (Verantwortungsträger), Health Advisor (Gesundheitsberater und Fürsprecher), Scholar (Wissenschaftlicher und Lehrer) und Professional (Professionelles Vorbild) (www.royalcollege.ca). Für jede dieser Rolle gibt es neben einer Rollendefinition und einer inhaltlichen Beschreibung mehrere formulierte Schlüsselkompetenzen, denen wiederum weitere präzisierte und operationalisierte Lernziele folgen. Dieses Konzept der CBE kommt damit der von Frank et al.

2010 vorgeschlagenen CBE-Definition, sich an den Bedürfnissen der Gesellschaft und der Patienten auszurichten, sehr nahe (Frank et al. 2010).

Um das Jahr 2000 wurde das ursprünglich für die ärztliche Fort- und Weiterbildung angedachte CanMEDS Konzept in den Bereich der medizinischen Ausbildung übertragen und wird heute international - auch in Europa - vielfach in Frameworks (Rahmenkonzepten) verwendet; so z.B. in Deutschland mit dem NKLM, in den Niederlanden mit dem Rahmenkonzept für die „undergraduate medical education in the Netherlands“ und in der Schweiz mit dem Schweizer Lernzielkatalog (NKLM 2015; Van Herwaarden, Laan & Leunissen 2009; Bürgi et al. 2008). In anderen Konzepten werden die Berufsrollen nicht oder nur teilweise explizit aufgeführt, die dahinterliegenden Kompetenzen jedoch gleichermaßen operationalisiert dargelegt. Als Beispiele sind hier das Tuning Project for Medicine zur Entwicklung eines paneuropäischen Rahmenkonzeptes für die Lernziele im Medizinstudium (Cumming & Ross 2007), „Tomorrows doctors“ des General Medical Council (General Medical Council 2009) oder auch „The Scottish doctor“ (Simpson et al. 2002) zu nennen. Aber auch einzelne universitäre Curricula wie z.B. das Düsseldorfer Curriculum Medizin (Modellstudiengang) wurden so konzipiert (Düsseldorfer Curriculum Medizin).

1.4 Evidenz für die Effektivität der outcome- bzw. kompetenz-orientierten Ausbildung

Im letzten Jahrzehnt haben sich die Competency Based bzw. Outcome Based Education nahezu zu einem Paradigma mit dem Status eines „god term“ in der medizinischen Aus-, Fort- und Weiterbildung entwickelt (Hodges & Lingard 2012, S.1). Zur sprachlichen Vereinfachung wird nachfolgend ausschließlich der Begriff der Competency Based Education (CBE) verwendet. Obwohl der Ansatz einer CBE bezogen auf die späteren beruflichen Anforderungen verständlich und nachvollziehbar erscheint, wird die Entwicklung von CBE bis heute als wesentlich von politischen, sozialen und regulatorischen Motiven beeinflusst und weniger auf wissenschaftlicher Evidenz basierend angesehen. (Brightwell & Grant 2013; Harden 2002; Harden 2007; Hawkins et al. 2015). Exemplarisch können für diese Motive wirtschaftliche Gründe in den USA angeführt werden, die eine Steigerung der Effizienz und Verkürzung der medizinischen Ausbildung fordern; so führt eine hohe finanzielle Verschuldung der Medizinstudierenden während ihres mehrjährigen Studiums zu einer Schieflage bei der späteren Wahl der Fachdisziplin, die sich demnach mehr an den Verdienstmöglichkeiten als an den Bedürfnissen der Gesellschaft orientiert (Cooke, Irba & O'Brien 2010, S. 28; Hodges 2010). Brightwell und Grant (2013) sehen im CBE-Ansatz eine Form der Selbstvergewisserung einer Gesellschaft, die ihre Erwartungen an das ärztliche Handeln

definiert und sicherstellen möchte, dass die ausgebildeten Ärztinnen und Ärzte in der Lage sind, diese Erwartungen mit ihrem Handeln zu erfüllen. Kritiker kommen zu dem Schluss, dass über einen Zeitraum von 40 Jahren bis heute kaum stichhaltige Evidenz hinsichtlich der Validität und Effektivität einer kompetenzorientierten medizinischen Ausbildung vorliegt und komplexere kognitive Ausbildungsziele Gefahr laufen, nicht mehr ausreichend berücksichtigt zu werden (Malone & Supri 2012).

Mit den Empfehlungen des Wissenschaftsrates, an die Stelle der traditionellen Orientierung an Fächern eine an den ärztlichen Rollen und ihren Kompetenzen orientierte Ausbildung zu stellen (Wissenschaftsrat 2014, S. 7), geht auch die Frage nach der Evidenz und Kongruenz von Lehre und Prüfungen einher. Gerade für die 'Fragen nach der Validität, Reliabilität und Objektivität von handlungsnäheren, kompetenzorientierten Prüfungsformaten sieht auch der Wissenschaftsrat weiteren Forschungsbedarf' (ebd. S. 9) und verweist auf die Stärkung einer „evaluativen Ausbildungsforschung“ (ebd. S. 47). Auch andere Autoren kommen zu dem Schluss, dass die Entwicklung der CBE zum internationalen Paradigma nicht durch eine substantielle Lehr-/Lernforschung begleitet wurde, welche den Einfluss von Lernzielen bzw. Learning Outcomes auf das Lehren und Lernen (und Prüfen) untersucht (Morcke, Dornan & Eika 2013). CBE Programmen wird eine longitudinale Forschung nahegelegt, welche die Prüfungsergebnisse mit relevanten Outcomes verknüpft. Diese Form der Evidenz erscheint unerlässlich für den Nachweis der Validität von Prüfungen im Rahmen von CBE sowie der Entscheidung darüber, wer als kompetent einzustufen ist und wer nicht (McClarty & Gaertner 2015).

1.5 Prüfung von Kompetenzen - wie können Kompetenzen messbar gemacht werden?

Seit der Entwicklung von CBE Programmen gilt der Prüfungsqualität eine hohe Aufmerksamkeit. Harris und Keller hoben bereits 1976 hervor, dass *“the major development effort in competency-based education should not lie in design of instructional materials but in design of appropriate performance assessments. Furthermore, institutions should not commit themselves to competency-based curricula unless they possess means to directly assess students’ performance.”* (Harris & Keller 1976). Trotz vieler Jahre der Entwicklung und Implementierung von kompetenzbasierten Curricula, kratzen in der Einschätzung von Hodge & Lingard anspruchsvolle Prüfungskonzepte zur Messung von Kompetenzen bis heute nur an der Oberfläche (Hodge & Lingard 2012, S. 8). Bei der Entwicklung kompetenzorientierter Prüfungen werden drei Schritte als besonders bedeutsam angesehen: erstens, inwieweit Prüfungsergebnisse tatsächlich das abbilden, was mit dem Prüfungsformat gemessen werden soll (Validität); zweitens, wie die Grenze definiert wird, an der kompetente von gerade inkompetenten Kandidaten

unterschieden werden und drittens, inwieweit bessere Ergebnisse in den Prüfungen auch zu einer besseren Performanz im späteren Beruf führen (McClarty & Gaertner 2015.) Noch im Jahr 2002 zeigten CBE Programme in der US-amerikanischen Higher Education kaum robuste Daten zur Reliabilität und Validität der Prüfungen (Jones, Voorhees & Paulson 2002).

Die Umsetzung kompetenzorientierter Curricula bringt also höhere Erwartungen an die Beurteilung und Dokumentation der individuellen Lernfortschritte und damit an die Prüfungen mit sich. Für die medizinische Ausbildung wird die Verwendung mehrdimensionaler Prüfungsinstrumente gefordert, über die sichergestellt werden kann, dass die Studierenden eine definierte Kompetenzstufe erreicht haben (Hawkins et al. 2015). Allerdings werden das wahrgenommene Fehlen zuverlässiger Prüfungsformate sowie das Fehlen einer übergeordneten Prüfungsstrategie als wesentliche Gründe für die bisher nur schleppende, tatsächliche Umsetzung von kompetenzbasierten Ausbildungsmodellen angesehen (Hawkins et al. 2015). Um tragfähig zu werden, sollten „Competence Based Education“-Programme Evidenz für die Validität ihrer Prüfungen sammeln, wie sie in den „best practices for assessment in competency-based education“ des Centers on Higher Education Reform, USA, formuliert wurden (McClarty & Gaertner 2015):

1. Eindeutige Kompetenzdefinitionen
2. Explizite Verknüpfung der in den Prüfungen gemessenen Fertigkeiten mit definierten Kompetenzen.
3. Das Verhalten oder die gedanklichen Prozesse der Studierenden spiegeln die Kompetenzen wider.
4. Die in einer Prüfung gemessene Kompetenz steht in Beziehung zu anderen Messergebnissen in Bezug auf dieselbe Kompetenz.
5. Der Zusammenhang von Prüfungsergebnissen und zukünftigen Outcomes wie z.B. dem Erfolg am Arbeitsplatz oder der Weiterentwicklung von Kompetenz wird empirisch belegt.

Hawkins (2015) sieht in der Prüfungsentwicklung die größte Herausforderung an eine CBE, zumal den Ergebnissen von Prüfungen zu verschiedenen Kompetenzdomänen momentan nur eine sehr unterschiedliche Vertrauenswürdigkeit attestiert wird. Die Inhalt- und Kontextspezifität des klinischen Handelns werden gar als Widerspruch zu einem isolierten Training verschiedener Kompetenzen angeführt (Lurie 2012). Bieten beispielsweise Prüfungen mit Multiple Choice Fragen für theoretisches Wissen messtechnisch reliable Ergebnisse, so ist dieses für die Überprüfung von Dimensionen wie Teamfähigkeit oder Professionalität nicht so einfach möglich. In der Vergangenheit wurde der Versuch, komplexere Kompetenzen mit objektiven Instrumenten zu erfassen, häufig verfehlt (Huddle & Heudebert 2007). Das objektiv Messbare läuft nach Huddle und Heudebert (2007) Gefahr, sich wesentlich auf Wissen und einfache Fertigkeiten zu reduzieren

(ebd.). Auch Hodges (2010) sieht die Gefahr, dass schwieriger zu prüfende Fragen erst gar nicht gestellt werden.

In der realen Welt ist Kompetenz ein komplexes und vielschichtiges Phänomen, welches nach Einschätzung einiger Autoren nicht einfach auf eine Anzahl messbarer Teilstücke heruntergebrochen werden kann, da die gesamte Kompetenz mehr ist als die Summe seiner Einzelteile (Malone & Supri 2012; Hyland T 1997, Talbot 2004). „Arzt / Ärztin sein“ ist demnach mehr als die Summe einzeln erfolgreich abgelegter Aufgaben in Bezug auf Wissen oder Fertigkeiten (Ashworth & Saxton 1990; Brightwell & Grant 2013). Ziel sollte demnach die Messung der umfassenden klinischen Fähigkeit sein, abstraktes Wissen und einzelne Fertigkeiten passend im klinischen Kontext auszuüben. In der Zeit nach der Approbation (post graduate) wird dieses zunehmend mit dem Format der „Entrustable Professional Activities“ (EPA) - deutsch: „Anvertraubaren Professionellen Tätigkeiten“ (APT) - versucht, um verschiedene Kompetenzen integrativ und eingebunden in konkrete Aufgabenstellungen des ärztlichen Handelns zu beurteilen (Ten Cate et al. 2010; Hawkin et al. 2015). Auch für dieses Konzept fehlt bisher noch die wissenschaftliche Evidenz zur Wirksamkeit. Für die Zeit der medizinischen Ausbildung, und hier insbesondere für die früheren Phasen des Studiums, in denen die Studierenden noch nicht kontinuierlich in den klinischen Kontext eingebunden sind, bedarf es hingegen anderer Prüfungskonzepte und einer Hinführung der Studierenden zum Kompetenzerwerb.

1.6 Operationalisierung von Kompetenzen

Ausgehend von der Annahme, dass Berufsrollen zunächst in einzelne Elemente aus definiertem Wissen oder Fertigkeiten zerlegt werden können, stellt die Ausrichtung der Lehr- und Prüfungskonzepte am Cambridge Modell einen möglichen Ansatz für die Operationalisierung von Kompetenzen im Rahmen der medizinischen Ausbildung dar (Rethans et al. 2002). Hierbei handelt es sich um eine Weiterentwicklung des Kompetenzstufenmodells nach Miller (Miller 1990). Das Cambridge Modell definiert vier Lern- und Prüfungsebenen („Wissen“, „Wissen Wie“, „Zeigen Wie“ und „Tun“). Kompetenz wird in diesem Modell als Fähigkeit beschrieben, eine direkt beobachtbare Aufgabe in einer definierten, kontrollierten Prüfungsumgebung z.B. in Simulation erfolgreich durchzuführen und entspricht der Ebene des „Zeigen Wie“ (Ramani & Leinster 2008; Rethans et al. 2002). Die Ebenen des „Wissen“ und „Wissen Wie“ sind dabei bereits integrierter Bestandteil dieser Ebene. Die Beurteilung des Handelns in der realen klinischen Umgebung wird über die höchste Stufe des Stufenmodells abgebildet und als Performanz bezeichnet. Die Performanz integriert damit neben den beiden Stufen des Wissens auch die Stufe der Kompetenz, also des „Zeigen Wie“. Kompetenzbasierte Prüfungsformate messen entsprechend dieses Modells

unter ausgewählten Prüfungsbedingungen, wohingegen die Performanz die Beobachtung des realen Handelns in der unmittelbaren klinischen Umgebung meint (Rethans et a. 2002). Auch der Nationale kompetenzbasierte Lernzielkatalog Medizin (NKLM) definiert für die ärztlichen Berufsrollen solch verschiedenen Kompetenzebenen: „Wissen“ (Faktenwissen), „Wissen Wie“ (Handlungs- und Begründungswissen), „Zeigen Wie“ (Unter Anleitung durchführen und demonstrieren) und „Performanz“ (selbständig situationsadäquat und in Kenntnis der Konsequenzen durchführen) (NKLM 2015).

Aufgabe und Messkriterien von Prüfungen

Prüfungen erfüllen im Studium unterschiedliche Funktionen. Es werden ergebnisorientierte (= summative) von prozessorientierten (= formative) Prüfungen unterschieden. Summative Prüfungen dokumentieren in der Summe den Erwerb bestimmter Qualifikationen über eine Differenzierung der erbrachten Leistungen, die dazu berechtigen, im Studium voranzukommen und/oder einen bestimmten Abschlussgrad und Titel zu führen (Ruschin 2010). Formative Prüfungen werden als wirksames Instrument für eine Rückmeldung zum individuellen Lernprozess, Lernfortschritt und Kompetenzerwerb eingesetzt (ebd.). Summative Prüfungsformate werden nach drei Hauptgütekriterien eingestuft: Objektivität, Reliabilität und Validität.

Objektivität bezeichnet das Maß, wie weit in der diagnostischen Situation eine Standardisierung des gesamten Testvorganges gelingt und demnach eine Unabhängigkeit der Testergebnisse vom Anwender besteht (Fisseni 2004, S.47). „Bei der Durchführung und Auswertung desselben Tests soll das gleiche Verhalten eines Probanden immer in gleicher Weise quantifiziert und die quantifizierten Ergebnisse (Test-Scores) immer in gleichem Sinne interpretiert werden“ (ebd. S. 47). Es werden drei Arten der Objektivität unterschieden: Durchführungs-, Auswertungs- und Interpretationsobjektivität.

Reliabilität bezeichnet die Zuverlässigkeit eines Tests, d.h. den Grad der Genauigkeit, mit dem er ein bestimmtes Merkmal misst, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht (Lienert & Raatz 1998, S.9). Der Grad der Reliabilität wird durch einen Reliabilitätskoeffizienten bestimmt, der angibt, „in welchem Maße unter gleichen Bedingungen gewonnene Messwerte über ein und denselben Probanden übereinstimmen, in welchem Maße also die Testergebnisse reproduzierbar sind“ (ebd. S. 9). Zur Bestimmung der Reliabilität gibt es verschiedene Zugangswege: bei der Konsistenzanalyse werden die Elemente eines Tests als multiple halbierte Testteile analysiert (ebd. S. 10). Ein häufiges Maß für die interne Konsistenz ist Cronbachs α (Alpha-Koeffizient).

Die *Validität* oder Gültigkeit eines Tests ist das wichtigste Testgütekriterium und gibt den Grad der Genauigkeit an, mit dem dieser Test tatsächlich das misst, was er messen soll. (Fisseni 2004, S. 62; Lienert & Raatz 1998, S. 10). „Ein Test ist demnach vollkommen valide, wenn seine Ergebnisse einen unmittelbaren und fehlerfreien Rückschluss auf den Ausprägungsgrad des zu erfassenden Merkmals zulässt“ (Lienert & Raatz 1998, S. 10). Auch bei der Validität können grundsätzlich drei verschiedene Aspekte unterschieden werden: *Inhaltliche Validität*, *Konstruktvalidität* und *Kriteriumsvalidität*.

Ein Test bzw. seine Elemente sind dann *inhaltsvalide*, wenn sie eine optimale und repräsentative Stichprobe des zu messenden Inhalts darstellen. Beispielsweise würde eine Klausur für das Fach Innere Medizin inhaltlich valide sein, wenn die Aufgaben inhaltlich eine repräsentative Auswahl aus allen Teilgebieten der Inneren Medizin darstellen. (ebd. S. 10). „Inhaltliche Validität wird einem Test in der Regel durch ein Rating von Experten als „Konsens von Kundigen“ zugebilligt“ (ebd. S. 11). Je klar umrissener ein zu messendes Konstrukt ist, desto leichter kann dieses inhaltsvalide geprüft werden. Mit einer MC-Klausur im Fach Innere Medizin kann Faktenwissen leichter inhaltlich valide geprüft werden als dieses beispielsweise für eine Prüfung möglich ist, mit der sozial-ethische Kompetenz gemessen werden soll.

Die *Kriteriumsvalidität* lässt sich ermitteln, indem „die Testergebnisse einer Stichprobe von Probanden mit einem oder mehreren Außenkriterien korreliert werden, welche vom Test unabhängig erhoben werden“ (ebd. S. 11) „Damit lässt sich dann direkt von der Testleistung auf die Kriterien-Leistung schließen, oder man kann das Kriterium als einen ausreichend validen Repräsentanten für das zu erfassende Merkmal ansehen und so indirekt Aussagen über dieses Merkmal machen“ (ebd.). „Je größer die Gemeinsamkeit des von Test und Kriterium erfassten Merkmalsanteils, umso größer ist die kriteriumsbezogene Validität eines Tests.“ (ebd.). Dabei wird das Kriterium entweder gleichzeitig mit dem Test erhoben (*konkurrente oder Übereinstimmungsvalidität*) oder zu einem späteren Zeitpunkt als der Test (*prädiktive oder Vorhersagevalidität*) (Fisseni 2004).

Die *Konstruktvalidität* schließt inhaltliche und kriteriumsbezogene Validität ein und umschreibt die Übereinstimmung zwischen einem Testergebnis und einem Netz anderer Scores oder anderer Aussagen (Fisseni 2004; S. 68). „Ein Test etwa, der den individuellen Ausprägungsgrad von Angst messen soll, hätte dann eine hinreichende Konstruktvalidität, wenn nachgewiesen wurde, dass das vom Test erfasste Merkmal in genügender Übereinstimmung mit dem theoretischen Konstrukt „Angst“ steht.“ (Lienert & Raatz 1998, S. 11) Ein klassisches Instrument der Konstruktvalidierung für Fragebögen ist die Faktorenanalyse. „Sie kann Items identifizieren, die auf demselben Faktor hoch laden. Solche „gleichartigen“ Items lassen sich (möglicherweise) als Indikatoren einer gemeinsamen Eigenschaftsdimension interpretieren.“ (Fisseni 2004, S. 68). „Das Konstrukt,

dessen Indikator das Testverhalten ist, wird eingebettet in ein „nomologisches³ Netz“ inhaltlich verwandter, ähnlicher Konstrukte (*konvergente Konstruktvalidität*) oder theoretisch entfernter Konstrukte, die niedrig oder nicht mit dem Test korrelieren sollen (*diskriminante Konstruktvalidität*)“ (ebd. S. 68).

Für Fragen nach der Validität, Reliabilität und Objektivität von handlungsnäheren, kompetenzorientierten Prüfungsformaten sieht auch der Wissenschaftsrat Forschungsbedarf (Wissenschaftsrat 2014, S. 9).

1.7 Einfluss des Lehr-/Lernklimas auf die studentischen (Prüfungs-) Leistungen und die Feedbackkultur in einem kompetenzbasierten Curriculum

Prüfungen (summativ und formativ) als Bestandteil eines Curriculums sollen die Leistungen von Studierenden erfassen und rückmelden. Die curriculare Implementierung von Prüfungen bedeutet dabei jedoch nicht zwangsläufig, dass diese auch zielgerichtet umgesetzt werden. Deren Einsatz und Nutzung hängt auch davon ab, wie die Menschen in einer Bildungseinrichtung das Curriculum leben und erleben.

Dijkstra et al. (2015) weisen auf den Einfluss der Lernumgebung hinsichtlich der erzielten Ausbildungsergebnisse für kompetenzorientierte Curricula hin. Auch andere Autoren haben zeigen können, dass die individuelle Wahrnehmung des Lehr-/Lernklimas einen wichtigen Einfluss auf das Verhalten und die in dieser Lernumgebung erzielten Ausbildungsergebnisse hat und ein positives Klima durchschnittlich zu besseren Leistungen führt als ein schlechtes Klima (Kusurkar et al. 2013; Pimparyon et al. 2000; Genn 2001b; Haertel 1981; Thistlethwaite 1959). Obwohl sich unter dem Begriff Lehr- und Lernklima fast jeder etwas vorstellen kann, gestaltet sich die Suche nach einer verbindlichen Definition äußerst schwierig und macht deutlich, dass es sich hierbei nicht um ein scharf umschriebenes Konstrukt handelt (Genn & Harden, 1986; Ostapczuk 2014, S.3). Das Konzept des Lehr- und Lernklimas enthält sowohl externe Anteile i.S. einer objektiven Umgebung als auch interne Anteile i.S. einer Persönlichkeitseigenschaft, da die subjektive Wahrnehmung ein- und derselben objektiven Tatsache interindividuell stark variieren kann (Ostapczuk 2014, S.3; Dornic & Ekehammar, 1990). Für die Bedeutung der externen Anteile spricht die meistens recht einvernehmliche Wahrnehmung des Klimas an einer Bildungseinrichtung als insgesamt eher gut bzw. eher schlecht (Ostapczuk 2014, S 4; Genn 2001a). Genn definiert das Lehr-/Lernklima als alles, was an einer Bildungseinrichtung passiert. Dazu

³ Bedeutung: durch Verknüpfung von Aussagen herleitbar

zählen das geplante Curriculum genauso wie die menschlichen Umgangsformen, die Atmosphäre und die gelebten Werte (Genn 2001a). Auch andere Autoren heben hervor, dass Lernen immer in einer spezifischen Umgebung stattfindet und durch die dort gelebten Werte, Stimmungen und Umgangsformen beeinflusst wird (Stacy 2003). Wie Studierende oder die Fakultätsmitglieder ihre Umgebung wahrnehmen kann somit als Klima bezeichnet werden (Till 2005). Diese gelebten Werte und Umgangsformen erscheinen besonders für formative und prozessorientierte Prüfungen (Feedback) von Bedeutung, über die der individuelle Lernprozess und Kompetenzerwerb der Studierenden konstruktiv gefördert werden soll. Hierauf wird im Verlauf der Arbeit näher eingegangen.

1.8 Feedback und Lehr-/Lernklima

Die Fähigkeit, überdurchschnittliche Leistungen zu erbringen, wird in der Regel nicht darauf zurückgeführt, dass eine überdurchschnittliche Begabung vorliegt oder Inhalte übermäßig häufig wiederholt werden, sondern eine stetige Verbesserung der Leistung resultiert wesentlich aus Feedback (Ericsson 2004; Ericsson & Charness 1994). Feedback ist entsprechend als formatives Prüfungselement international wichtiger Bestandteil in den kompetenzorientierten Rahmenkonzepten für die Ärztliche Aus-, Weiter- und Fortbildung (Issenberg et al. 2005). Die Aufnahme von Feedback in Rahmenkonzepte, Curricula und Verordnungen bedeutet aber noch nicht automatisch dessen Durchführung und gelingende Umsetzung in dem Sinne, dass Feedback konstruktiv gegeben oder angenommen wird und zu einer Verhaltensänderung führt. Archer weist darauf hin, dass effektives Feedback im Rahmen der medizinischen Ausbildung nicht alleine über eine vorgeschriebene Anzahl von Feedbacks gelingt, sondern von den beteiligten Personen verinnerlicht und im Sinne einer Feedbackkultur gelebt werden muss (Archer 2010). Entsprechend der Literatur sind hierfür neben einem vertrauenswürdigem Verhältnis zwischen Feedbackgeber und -nehmer (Crommelinck & Anseel 2013) eben auch die Lernumgebung (d.h. das Lehr-/Lernklima) entscheidend, in der das Feedback stattfindet (Kinicki et al. 2004).

2 Zielsetzung der Arbeit

Die vorliegende Arbeit hat zwei wesentliche Ziele:

2.1 Schaffung von Evidenz für die Validität und Reliabilität von kompetenzorientierten Prüfungen

2.2 Entwicklung und Validierung von Messinstrumenten zur Erfassung des Lehr-/Lernklimas

Ad 2.1. Schaffung von Evidenz für die Validität und Reliabilität von kompetenzorientierten Prüfungen

Die Schaffung von Evidenz für die Validität und Reliabilität von kompetenzorientierten Prüfungen erfolgt methodisch in Ausrichtung auf die Kriterien für Testaufbau und Testanalyse (Lienert & Raatz 1998) sowie den „Standards for Educational and Psychological Testing“ (American Educational Research Association 1999 u. 2014).

Insbesondere die Schaffung von Evidenz für die Validität orientiert sich dabei an den Empfehlungen zu „best practices for assessment in competency-based education“ des Centers on Higher Education Reform, USA (McClarty & Gaertner 2015); (vgl. hierzu S. 13. Nr. 2, 4 und 5):

2.1.1. Verknüpfung des in einer Prüfung gemessenen Wissens und/oder der gemessenen Fertigkeiten mit definierten Kompetenzen

Messung von „Wissen“ bei der diagnostisch-therapeutischen Kompetenz in einer fallbasierten Online-Prüfung:

- **Rotthoff T**, Baehring T, Dicken HD, Fahron U, Fischer MR, Adler M, Hege I, Scherbaum WA. Fallbasierte Online-Prüfungen im Medizinstudium – Zielsetzung, Implementierung und praktische Erfahrungen. *GMS Medizinische Informatik, Biometrie und Epidemiologie* (MIBE) 2006; 2(3): Doc11
- **Rotthoff T**, Baehring T, Dicken HD, Fahron U, Richter B, Fischer MR, Scherbaum WA. Comparison between Long Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial. *BMC Med Educ* 2006, 6:50

Messung von kommunikativer Kompetenz in einer Objective Structured Clinical Examination (OSCE):

- Mortsiefer A., Immecke J., **Rotthoff T.**, Karger A., Schmelzer R., Raski B., in der Schmitten J, Altiner A, Pentzcek M. Summative assessment of undergraduates' communication competence in challenging doctor–patient encounters. Evaluation of the Düsseldorf CoMeD-OSCE. *Pat Educ Couns* 2014; 95: 348-355

2.1.2. Beziehung der in einer Prüfung gemessene Kompetenz zu anderen Messergebnissen in Bezug auf dieselbe Kompetenz sowie die Beziehung von Prüfungsergebnissen und zukünftigen Outcomes

Ermittlung der Kriteriumsvalidität eines kompetenzbasierten Assessment Centers mit einem Follow-Up über 4 Jahre:

- **Rotthoff T.**, Ostapczuk M, Kröncke KD, Zimmerhofer A, Decking U, Schneider M, Ritz-Timme S. Criterion validity of a competency-based assessment center in medical education – a 4-year follow up study. *Med Educ Online* 2014; 19:252-54

Ad 2.2. Die Erstellung und Validierung von Messinstrumenten zur Erfassung des Lehr-/Lernklimas

Da die individuelle Wahrnehmung des Lehr- /Lernklimas einen wichtigen Einfluss auf das Verhalten und die in der Lernumgebung erzielten Ausbildungsergebnisse/Outcomes hat und ein positives Klima durchschnittlich zu besseren Leistungen führt als ein schlechtes Klima (s.o.), wurden Messinstrumente zur Erfassung des Lehr-/Lernklimas zum internationalen Gebrauch erstellt und für den deutschsprachigen Raum in Bezug auf die geforderten Testkriterien (s.o.) validiert.

- **Rotthoff T.**, Ostapczuk M, de Bruin J, Decking U, Schneider M, Ritz-Timme S. Assessing the learning environment of a Faculty. Validation of the Dundee Ready Education Environment Measure (DREEM) with Students and Teachers. *Med Teach.* 2011; 33(11):e624-36.
- Ostapczuk M, Hugger A, de Bruin J, Ritz-Timme S, **Rotthoff T.** DREEM on, dentists! Students' perception of the educational environment in a German dental school as

measured by the Dundee Ready Education Environment Measure. *Eur J Dent Educ.* 2012; 16(2):67-77

- **Rotthoff T**, Ostapczuk M, de Bruin J, Kröncke K, Decking U, Schneider M, Ritz-Timme S. Development and evaluation of a questionnaire to measure the perceived implementation of the mission statement of a competency based curriculum. *BMC Med Educ* 2012; 12:109

3 Eigene Untersuchungen

Die Darstellung der eigenen Untersuchungen erfolgt entlang der Systematik in der Zielsetzung (Kapitel 2). Sie beginnt mit der Präsentation eigener Arbeiten zur geschaffenen Evidenz für die Validität und Reliabilität von kompetenzorientierten Prüfungen und schließt mit der Erstellung und Validierung von Messinstrumenten zur Erfassung des Lehr-/Lernklimas ab.

3.1 Schaffung von Evidenz für Validität und Reliabilität von kompetenzorientierten Prüfungen.

3.1.1 Verknüpfung des gemessenen Wissens und/oder der gemessenen Fertigkeiten mit definierten Kompetenzen

Ziel der beiden nachfolgend zusammengefassten eigenen Arbeiten war es, über eine online und fallbasierte Prüfung einen höheren Realitäts- bzw. Praxisbezug herzustellen und dabei mit sog. Long Menu Fragen die klinische „Entscheidungsfindung“ als Teilaspekt der diagnostisch-therapeutischen Kompetenz in einer summativen Prüfung zu erfassen.

Der Kompetenzbegriff beinhaltet Wissen, Fertigkeiten und Haltungen. Wissen ist somit ein grundlegender Baustein jeder Kompetenz und auch bei der diagnostisch-therapeutischen Kompetenz wesentlich. Untersuchungen zeigen, dass ein Lernen an Patienten im klinischen Kontext nachhaltiger ist, wenn das hierbei erworbene Wissen später besser reproduzierbar und auf andere klinische Situationen übertragbar ist (Koens, Ten Cate & Custers 2003; Spencer 2003). Den Studierenden fehlt oftmals noch diese Fähigkeit, Wissen situationsbezogen anzuwenden und Strategien zur Lösung von Problemen mittels geeigneter diagnostischer und therapeutischer Maßnahmen zu entwickeln. Fallstudien bieten sich für eine anwendungsorientierte Wissensüberprüfung besonders an und werden daher als Prüfungsformat auch gemäß Vorgabe der ÄAppO im zweiten Abschnittes der ärztlichen Prüfung verwendet (ÄAppO § 28 (1)). Ergänzend zur Behandlung realer Patienten, soll der Einsatz fallbasierter und realitätsnah gestalteter E-Learning Fälle mittels hoher Anwendungsorientierung diesen Lernprozess simulieren und zusätzlich unterstützen (Leong, Baldwin & Adelman 2003; Simonsohn & Fischer 2004).

Publikation 1

Rotthoff T, Baehring T, Dicken HD, Fahrion U, Fischer MR, Adler M, Hege I, Scherbaum WA. **Fallbasierte Online-Prüfungen im Medizinstudium – Zielsetzung, Implementierung und praktische Erfahrungen.** *GMS Medizinische Informatik, Biometrie und Epidemiologie (MIBE)* 2006; 2(3): Doc11

Publikation 2

Rotthoff T, Baehring T, Dicken HD, Fahrion U, Richter B, Fischer MR, Scherbaum WA. **Comparison between Long Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial.** *BMC Med Educ* 2006, 6:50

Prüfung von diagnostisch-therapeutischer Kompetenz

Vor dem Hintergrund, in Prüfungen nicht nur „Wissen“, sondern auch Teilkompetenzen wie diagnostisches Denken und Handeln zu prüfen, wurde an der Medizinischen Fakultät der HHU erstmalig in Deutschland mit CASUS® (www.casus.eu) ein fall- und computerbasiertes Lernsystem zum Lernen mit anschließender internetbasierter Online-Prüfung verbindlich und als summative Prüfung eingesetzt.

Die Wissensüberprüfung erfolgt im Medizinstudium und in den Staatsexamina zu einem großen Teil mit Fragen im Antwortwahlverfahren (z.B. Multiple Choice (MC) Fragen. Damit kommt gerade den MC Fragen wegen ihrer Dominanz in den Prüfungen des Medizinstudiums eine ganz besondere Bedeutung für die Lernsteuerung zu. Vielfach prüfen die im Medizinstudium verwendeten MC-Fragen bisher überwiegend faktisches und nur in geringerem Maße anwendungsorientiertes Wissen. Das zeigt auch ein aktueller Blick in die Prüfungsfragendatenbank (Item Management System) des Umbrella Consortium for Assessment Networks (UCAN), dem zurzeit 20 deutsche Fakultäten, einschließlich der Düsseldorfer Fakultät, angehören (UCAN 2016). MC Fragen bleiben damit häufig unter ihren Möglichkeiten, auch tiefergehendes Verständnis und analytisches Denken zu prüfen (Morrison, Ross & Kemp 2004, S.271). Für Prüfungen hat sich der Stimulus einer Frage z.B. über deren Einbindung in einen medizinischen Kontext als bedeutsam erwiesen (Schuwirth & Van der Vleuten 2004; Schuwirth et al. 2001). Durch einen höheren Realitäts- bzw. Praxisbezug kann die Validität der Prüfung verbessert werden, weshalb auch vermehrt MC Fragen verwendet werden sollen, die in einen klinischen Kontext eingebettet sind (Van der Vleuten & Schuwirth 2005). Über die Einbindung in einen klinischen Kontext lässt sich auch „Entscheidungsfindung“ als Teilaspekt der diagnostisch therapeutischen Kompetenz sehr gut über Fallstudien abbilden. Wurde anfänglich angenommen, dass Entscheidungskompetenz eine

generelle Fähigkeit und weitgehend unabhängig vom Faktenwissen ist, stellten Bordage und Page 1987 die Bedeutung von Wissen in diesem Prozess heraus und leiteten daraus das Konzept der Key-Feature Fälle ab. Statt der vollständigen Bearbeitung eines einzelnen langen Falles, reduzierten sie Fälle und die klinischen Probleme auf ihre kritischen Schritte (Key Features) (Bordage & Page 1987, S. 57-75; Kopp, Möltner & Fischer 2006). Das Hauptaugenmerk wurde bei der Fallbearbeitung auf die Schritte gelegt, bei denen im klinischen Alltag häufig Fehler gemacht werden. (Page & Bordage 1995; Page, Bordage & Allen 1995).

MC Fragen sind weniger gut geeignet, spontan eine Diagnose zu generieren, was mit Freitextfragen (FT) besser möglich ist. Freitextfragen erschweren jedoch bei großen Studierendenzahlen eine automatisierte Auswertung, weshalb Long Menu (LM) Fragen Anwendung finden können. Bei diesem Fragenformat müssen die Studierenden auf eine offen gestellte Frage aktiv einen Begriff in ein Textfeld eingeben. Der eingegebene Begriff wird mit einer im Computerprogramm hinterlegten langen Liste („Long Menu“) abgeglichen und daraus erkannt. Damit wird eine automatisierte Auswertung möglich und aufgrund der Länge dieser Liste verliert die Wiedererkennung von richtigen Antworten - wie bei MC Fragen - an Bedeutung. Die Antwort muss vom Prüfling aktiv generiert werden. Mit diesem Fragenformat kann der Stimulus für eine aktive Antwort im Zusammenhang mit einer klinischen Entscheidungsfindung bei Diagnostik oder Therapie gesetzt werden (Veloski, Rabinowith & Robeson 1993) und entspricht der Ebene des „Wissen wie“ im zuvor dargelegten Cambridge Modell (Rethans et al. 2002). Computerbasiert wurden sie erstmalig von Schuwirth et al. verwendet (Schuwirth, Stoffers, Peperkamp 1996) und von Fischer et al. in einer Key Feature Prüfung reliabel eingesetzt (Fischer et al. 2005). Die Verwendung der Long Menu Fragen/Antworten erfolgte alternativ zu Freitextfragen, allerdings fehlten bis dahin vergleichende Untersuchungen dieser beiden Fragen/Antwortformate.

Methode

In zwei fächerübergreifend konzipierten, einwöchigen Modulen des 4. Studienjahres (Modul Diabetes, Modul Anämie) wurden CASUS®-Fälle als e-Learning Einheit curricular eingebunden. Die Abschlussprüfung beider Module erfolgte ebenfalls mit CASUS® Fällen in Form einer summativen online Key Feature (KF)-Prüfung am Ende des Semesters. Die Prüfungen waren verpflichtend und die Studierenden konnten darüber Bonuspunkte für die spätere Abschlussklausur im Fach Innere Medizin erwerben. Die Prüfungen wurden regelmäßig am Ende des 7. Semesters mit ca. 140 Studierenden durchgeführt. In einer dieser Prüfungen wurde untersucht, ob sich Long Menu Fragen in Bezug auf ihre Schwierigkeit, Anwendung und Antwortzeit von Freitextfragen unterscheiden und welche Inhalte sich für Long Menu Fragen eignen.

Es wurden 146 Studierende des 4. Studienjahres in 2 Gruppen (A und B) randomisiert, 142

Studierende (76 Frauen und 66 Männer) nahmen an der Studie teil.

Für die Prüfung wurden 7 Key Feature Fälle erstellt mit insgesamt 25 Fragen im Peer-Review Verfahren zu den Themengebieten Hämatologie und Diabetologie. Die inhaltliche Validität kann dem Test über das Experten-Rating zugebilligt werden (Lienert & Raatz 1998 S.11) In beiden Gruppen hatten alle Fragen denselben Inhalt, neun Fragen unterschieden sich jedoch in ihrem Antwortformat. Gruppe A beantwortete 9 Fragen im Long Menu und Gruppe B diese Fragen im Freitextformat. Da es sich um eine reale Prüfung handelte, konnten die Studierenden der Gruppe A diese Fragen zusätzlich mit einer Freitextantwort beantworten, sofern sich die von den Studierenden gesuchte Antwort nicht im vorgehaltenen Long Menu befand. Die übrigen 16 Fragen und Antwortformate waren für beide Gruppen identisch. Mit Beginn der Texteingabe der Long Menu Antwort glich der Computer die Eingaben alphabetisch mit einer hinterlegten Liste ab, die von den Studierenden über ein Pop-up und scroll-down Menü eingesehen werden konnte. Wurde beispielsweise als Antwortbegriff „Ketoazidose“ gesucht und die Buchstabenfolge „Keto“ eingegeben fanden sich im Long Menu die Begriffe Ketoazidose, Ketokonazol, Ketolides, Ketotifen. Die Studierenden mussten die gewählte Antwort bestätigen und damit wurde die Antwort in die Datenbank geschrieben. Die Long Menu Liste enthielt 500 Begriffe aus dem Bereich Innere Medizin zu Diagnosen, diagnostischem Vorgehen und Therapieverfahren. Die Liste wurde aus 8000 MC Fragen des 2. Staatsexamens erstellt. Alle Freitextantworten wurden von zwei Personen anhand eines zuvor definierten Antwortschlüssels ausgewertet. Die Antwortzeiten der Studierenden für die Fragen wurden in Sekunden gemessen. Zum Bestehen der Prüfung waren 60% der zur erreichenden Gesamtpunktzahl erforderlich.

Basierend auf der Null Hypothese gleicher Mittelwerte für richtig beantwortete Long Menu und Freitextfragen untersuchten wir auf Unterschiede. Der Alpha-Fehler wurde mit $p = 0.05$ festgelegt. Unterschiede zwischen den Items wurden zusätzlich analysiert und mit dem Chi-square und Fischer's Exact Tests berechnet (StatsDirect Statistical Software (Version 2.2.6.)). Die Antwortzeiten für die Long Menu und Freitext-Fragen wurden elektronisch erfasst und mit dem t-test für unabhängige Stichproben geprüft.

Ergebnisse

Insgesamt stieg mit der Einführung der Online-Prüfung der Nutzungsgrad des CASUS® Lernfälle durch die Studierenden während des Semesters mit zunehmender Prüfungsnähe steil an: von $n = 12$ Studierende zu Semesterbeginn (Oktober 2004) auf $n = 132$ Studierende zum Semesterende unmittelbar vor der Prüfung (Februar 2005). In der freiwilligen Evaluation zeigten die Prüfungsfälle im Vergleich zu den Lernfällen auf einer 5-Punkt Likert Skala (1 = mangelhaft bis 5 = sehr gut) eine etwas niedrigere inhaltliche Akzeptanz (Lernfälle $M = 3.6$ ($n = 89$); Prüfungsfälle $M = 3,4$ ($n = 109$), was mit der Prüfungssituation erklärt werden kann. Das fallbasierte computerbasierte

Prüfungsformat selber wurde mit einem Mittelwert von $M = 3.9$ und der technische Ablauf mit einem Mittelwert von $M = 4,5$ gut bis sehr gut bewertet und zeigte eine hohe Akzeptanz bei den Studierenden.

Entsprechend der Null-Hypothese fanden sich keine signifikanten Unterschiede zwischen den Mittelwerten der korrekten Long Menu und korrekten Freitext Antworten ($p = 0,92$). Im Mittel wurden die Fragen über das Long Menu Antwortverfahren mit 73,3% (SD 16,7) und bei der Freitextantwortoption mit 73,5% (SD 19,2) richtig beantwortet. Bis auf ein Item (Item Nr. 3) fanden sich auch zwischen den einzelnen Items keine signifikanten Unterschiede für Long Menu oder Freitextantwort. Bei den MC Fragen bestand hinsichtlich der richtigen Beantwortung zwischen beiden Gruppen kein signifikanter Unterschied (Gruppe A 67,5 % (SD 25,2); (Gruppe B 69,8 % (SD 25,3)). Auch für die im Mittel benötigte Zeit für die Beantwortung der offenen Fragen über das Long Menu Antwortverfahren im Vergleich zum Freitext-Antwortverfahren bestand kein signifikanter Unterschied (LM 101 Sekunden; SD 32; FT 107 Sekunden; SD 31 $p = 0.65$).

Lediglich 20 (3,2 %) von 630 möglichen Antworten in Gruppe A wurden über das Long Menu bezogen auf die Sinnhaftigkeit der Fragestellung falsch eingegeben. Eine Detailanalyse ergab, dass bei 13 dieser 20 Antworten von den Studierenden unplausible Antworten aus dem Long Menu ausgewählt und bestätigt wurden (Beispiel: erwartete richtige Antwort „Ketoazidose“, vom Studierenden ausgewählte Antwort „Perikardtampnade“). Bis auf eine Antwort waren alle von den Studierenden in das zusätzliche Freitextfeld eingegebenen Antworten auch im Long Menu gelistet.

Prüfung von kommunikativer und sozialer Kompetenz

Hinsichtlich der best practice Empfehlungen, das in einer Prüfung gemessene Wissen oder die gemessenen Fertigkeiten mit definierten Kompetenzen zu verknüpfen, stellt die Überprüfung der kommunikativen Kompetenz eine besondere Herausforderung dar, da diese als eine sehr komplexe, für die optimale (ärztliche) Berufsausübung aber essentielle Kompetenz eingestuft wird (van der Vleuten & Dannefer 2012). In der Vergangenheit wurde der Versuch, gerade komplexere Kompetenzen mit objektiven Instrumenten zu erfassen, häufig verfehlt (Huddle & Heudebert 2007). Das Training kommunikativer Fähigkeiten wird auch in der ärztlichen Approbationsordnung und damit für die medizinische Ausbildung definiert (ÄAppO). Im NKLM werden zentrale ärztliche Kompetenzen in Form von Rollen zugeordnet, welche sich ihrerseits wiederum aus Teilkompetenzen und Lernzielen zum medizinischen Wissen, wissenschaftlichen und klinischen Fähigkeiten und Fertigkeiten sowie professionellen Haltungen zusammensetzen. Für die kommunikative Kompetenz ist die eigene Rolle des „Kommunikators“ definiert (NKLM 2015, Kapitel 2). Leitlinien empfehlen im Rahmen der ärztlichen Ausbildung nicht nur

grundlegende kommunikative Fähigkeiten zu vermitteln, sondern auch komplexere Arzt–Patient Situationen, wie beispielsweise den Umgang mit starken Emotionen oder das Überbringen schlechter Nachrichten (Bachmann et al. C, 2013; von Fragstein et al. 2008). Untersuchungen konnten positive Effekte von Kommunikationstrainings hinsichtlich einer Verbesserung von kommunikativer Kompetenz auch für anspruchsvolle Situationen wie dem „Überbringen schlechter Nachrichten“, „Umgang mit häuslicher Gewalt“ oder „Risikokommunikation“ nachweisen (Liénard et al. 2010; Fallowfield & Jenkins 2004; Heron et al. 2010). Eine professionelle Arzt-Patient Kommunikation stärkt nicht nur die Zufriedenheit bei Ärzten und Patienten, sondern hat auch positive Effekte auf das erzielte Behandlungsergebnis von Patienten (Duffy et al. 2004; Simpson et al. 1991; Kurtz, Silverman & Draper 1998; Stewart 1995; Street et al. 2009). Das dem Thema allerdings bisher auch in der ärztlichen Fortbildung nur eine geringe Bedeutung beigemessen wurde, konnte für das Fortbildungsverhalten von Ärztinnen und Ärzten gezeigt werden (Rotthoff et al. 2010).

Nachfolgend werden nun die Hintergründe, die Entwicklung und die Durchführung einer summativen Prüfung zur Erfassung der kommunikativen Kompetenz in speziellen ärztlichen Gesprächssituationen im Rahmen des Düsseldorfer Curriculums dargestellt.

Im Düsseldorfer Studiengang Medizin wurde mit der Entwicklung des CoMeD Programms (Communication in Medical Education Düsseldorf) den Rahmenvorgaben der ÄAppO Rechnung getragen (www.comed-duesseldorf.de). Nach einem kommunikativen Basistraining im zweiten und dritten Studienjahr mit Fokussierung auf der Erhebung einer Anamnese, wurden ab dem Jahr 2008 sieben Trainingseinheiten (CoMeD-Kurse) zu komplexeren kommunikativen Fähigkeiten für Studierende im 4. Studienjahr entwickelt und sukzessive implementiert: 1. Aggressiv fordernder Patient 2. Schuld und Schamgefühle, 3. Überbringen schlechter Nachrichten, 4. Shared decision making, 5. Sexualanamnese, 6. Aufklärungsgespräch, 7. Interkulturelle Kommunikation (Mortsiefer, Rotthoff & Schmelzer et al. 2012) Die Kurse wurden durch begleitende Lehreinheiten (Präsenz– oder E-Learning) vorbereitet und anhand praktischer Übungen mit Schauspielpatienten durchgeführt. Für die Überprüfung dieser Lehr-/Lerneinheiten wurde ein OSCE (Objective Structured Clinical Examination) entwickelt, dessen Umsetzung und erzielten Messergebnisse nun genauer erläutert werden.

Publikation 3

Mortsiefer A., Immecke J., Rothhoff T., Karger A., Schmelzer R., Raski B., in der Schmitten J, Altiner A, Pentzack M. **Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Düsseldorf CoMeD-OSCE.** *Pat Educ Couns* 2014; 95: 348-355

Die OSCE ist als objektives, reliables, valides und auch durchführbares Prüfungsinstrument seit vielen Jahren zur Überprüfung klinischer Fertigkeiten etabliert. Über die Prüfung von kommunikativer Kompetenz im Rahmen eines OSCE wird zwar schon in den 90er Jahren berichtet (Hargie et al. 1998), jedoch stehen häufiger eher die allgemeine Kommunikation und nicht eine schwerpunktmäßige Überprüfung komplexer Gesprächssituationen im Fokus (Newble 2004; van Nuland et al. 2007; Schirmer et al. 2005; Rider, Hinrichs & Lown 2006).

Für die medizinische Ausbildung wurde bisher nicht über eine spezielle OSCE Prüfung zu komplexen Gesprächssituationen berichtet.

Methoden

Die vorliegende Untersuchung beschreibt die Entwicklung, Durchführung und Analyse eines summativen OSCE bezüglich seiner Reliabilität sowie Konstrukt- und Kriteriumsvalidität. Hierbei wurden für OSCE spezifizierte „reporting standards“ berücksichtigt (Patricio et al. 2009). Zur Beurteilung der Reliabilität sowie der Konstrukt- und Kriteriumsvalidität wurden die Ergebnisse von drei Prüfungsdurchgängen (Feb. 2011, Juli 2011, Feb. 2012) kumuliert ausgewertet. Die Prüfungsstationen waren zuvor im Jahr 2010 zweimal pilotiert worden. Alle Prüfungsstationen wurden von didaktischen Experten der CoMeD Steuerungsgruppe aus den Fachbereichen Allgemeinmedizin, Innere Medizin, Psychosomatische Medizin und Medizinischen Psychologie entwickelt. Sämtliche Prüfungsinhalte wurden den Studierenden im Rahmen von CoMeD Kursen während des Semesters vermittelt. Die Prüfungen fanden mit standardisierten Schauspielpatienten und trainierten Prüfern (inkl. Durchführung eines Standard-Setting) statt.

Tabelle 1 (OSCE Stationen zur Kommunikation)

<i>OSCE Station</i>	<i>Kommunikationsthema</i>	<i>Leitfaden /Modell</i>	<i>Körperlicher Befund</i>	<i>Einsatzort</i>
Schuld und Scham	Häusliche Gewalt	NURSE*	Multiple Prellungen	Notaufnahme
Überbringen schlechter Nachrichten	Erstmalige Mitteilung einer schlechten Diagnose	SPIKES*	Rheumatoide Arthritis	Hausarztpraxis
Aggressive(r) Patient(in)	Beschwerdemanagement	CALM***	vorzeitig einsetzende Wehen	Entbindungsstation
Shared Decision Making	Risikokommunikation	OPTION****	primäre kardiovaskuläre Prävention	Hausarztpraxis

*Back et al. (NURSE), ** Baile et al. (SPIKES), ***Howard (CALM), **** Elwyn et al. (OPTION)

Als Messinstrument wurde eine validierte Global Rating Scale (GR) verwendet (Scheffer et al. 2008). Ratingskalen bestehen im Unterschied zu Checklisten meist aus weniger Items, die Verhalten auf einem globaleren Niveau beschreiben (Scheffer 2009, S. 20.) Das hier verwendete globale Rating Scale besteht aus vier Dimensionen, die auf einer 5er Likert Skale (ebd.) bewertet werden:

- Eingehen auf die Gefühle und Bedürfnisse des Patienten (Empathie)
- Logischer Zusammenhang des Gespräches (Struktur)
- verbaler Ausdruck und
- nonverbaler Ausdruck

Statistik

Zur Berechnung der Gesamt-Reliabilität des OSCE sowie der Stations- und Itemreliabilität wurde Cronbachs' alpha berechnet. Die Interrater-Reliabilität wurde mittels Intraklassen-Korrelationskoeffizienten bestimmt. Zur Analyse der Konstruktvalidität wurden Parallelmessungen mit einem anderen Bewertungsbogen (Cologne Evaluation-Score for Communication - CEC) (Koerfer et al. 2008) durchgeführt und Pearson Korrelationen berechnet.

Zur Bestimmung der Kriteriumsvalidität des Kommunikations-OSCE wurden Unterschiede zwischen Männern und Frauen in der Prüfung mit dem Mann-Whitney U-Test ermittelt, da Geschlechterunterschiede bei der Kommunikation belegt sind.

Ergebnisse

Insgesamt konnten Daten von 456 Studierenden aus drei aufeinanderfolgenden Prüfungen analysiert werden. Im Mittel wurden 59,4 Punkte (74,25 %) der maximal zu erreichenden Punkte erworben (Min = 20, Max = 79, STD 9,04). Über das zuvor von den Prüfern durchgeführte Standardsetting wurde ein Minimum-Score von 50 Punkten für das Bestehen der Prüfung ermittelt. Die Gesamtreliabilität des OSCE lag bei $\alpha = 0,64$ und damit niedriger als mit 0,8 für summative, scheinrelevante Prüfungen gefordert (Möltner, Schellberg & Jünger). Sowohl die Reliabilität der Prüfungsstationen ($> 0,8$) als auch die Diskrimination der einzelnen Prüfungsstationen waren gut. Die konvergente (Konstrukt-) Validität kann aufgrund der hohen Korrelation zwischen der Global Rating Scale mit dem Cologne Evaluation-Score for Communication – CEC als gut eingestuft werden. Für die Kriteriumsvalidität fand sich ebenfalls ein gutes Ergebnis. Frauen erzielten signifikant bessere Leistungen im Kommunikations-OSCE als die Männer ($p < 0,001$). Frühere Untersuchungen bescheinigen Frauen im Vergleich zu Männern im Gespräch einen stärkeren emotionalen Zugang zum Gesprächspartner (Roter, Hall & Aoki 2002). Das Eingehen auf die Gefühle des Patienten wurde mit der hier verwendete Rating Scale erfasst.

Mit dem Kommunikations-OSCE konnte gezeigt werden, dass eine einzelne komplexe Kompetenz wie die kommunikative Kompetenz mit einem bereits etablierten Verfahren zufriedenstellend reliabel und valide gemessen werden kann.

3.1.2 Beziehung der Messung einer Kompetenz in verschiedenen Messverfahren und prädiktive Validität von Kompetenzmessungen

Auch der Vergleich von Messergebnissen derselben Kompetenz mit verschiedenen Mess-/Prüfungsverfahren wird für die Validierung der CBE empfohlen. Gemeint ist damit, inwieweit unterschiedliche Messinstrumente zum gleichen Messergebnis bei der Bewertung derselben Kompetenz eines Prüflings führen. Wenn einem Prüfling Kompetenz unter Verwendung eines Messinstrumentes bescheinigt wird, stellt sich außerdem die Frage, ob die bescheinigte Kompetenz dann auch zu kompetentem Verhalten in einer anderen Situation führt (= Outcome). Es geht also um den prädiktiven Wert (= prädiktive Validität) der erfolgten Messung. Vor diesem Hintergrund wurde mit der nachfolgenden Studie untersucht, inwieweit sich ein kompetenzbasiertes

Assessment Center (AC) für eine integrative Prüfung und Rückmeldung von übergeordneten Kompetenzen bei Medizinstudierenden im 2. Semester - also in einer frühen Phase des Medizinstudiums - eignet. Neben der Ausbildung und Überprüfung einzelner separater Kompetenzen, geht es bei der ärztlichen Tätigkeit um die Fähigkeit, verschiedene Kompetenzen für eine optimale Patientenversorgung situationsbezogen zu integrieren.

Kompetenz ist immer im Kontext eines spezifischen Handlungsumfeldes zu betrachten. Welche Möglichkeiten bieten sich also für Medizinstudierende, wenn sie noch nicht intensiv in den klinischen Kontext eingebunden sind? Für diese Gruppe sind neue Konzepte zur Beurteilung, Rückmeldung und Weiterentwicklung von übergeordneten Kompetenzen auf der Ebene des „shows how“ in einer integrativen und umgebungstypischen Weise erforderlich.

Die nachfolgende Arbeit beschreibt die Entwicklung, Durchführung und Ergebnisse eines AC, der integrativ verschiedene Kompetenzen messen will.

Publikation 4

Rotthoff T, Ostapczuk M, Kröncke KD, Zimmerhofer A, Decking U, Schneider M, Ritz-Timme S. **Criterion validity of a competency-based assessment center in medical education – a 4-year follow up study.** *Med Educ Online* 2014; 19:252-54

Mit der vorliegenden Arbeit wurde die Eignung eines Assessment Centers (AC) mittels psychometrischer Analyse der AC-Elemente untersucht. Darüber hinaus wurde die Kriteriumsvalidität des AC durch einen prädiktiven Vergleich der AC Testergebnisse von Teilnehmerinnen und Teilnehmern mit deren Ergebnissen in späteren kompetenzbasierten Prüfungsformaten über einen Zeitraum von 4 Jahren analysiert. Entsprechend des o.g. Cambridge Modells erfolgt im AC die Kompetenzmessung wie auch beim Kommunikations-OSCE auf der Ebene des „Zeigen wie“ (Kompetenz) durch Beobachtung in einer definierten, kontrollierten Prüfungsumgebung (Rethans et al. 2002; Ramani & Leinster 2008). Für AC ist die Beurteilung von beobachtbarem Verhalten ein wesentliches Kriterium, weshalb diese Methode gewählt wurde. Meist geht es in einem AC um die Auswahl von geeigneten Bewerbern für eine berufliche Tätigkeit, d. h. um die Vorhersage zukünftiger „Job Performance“ zum Zwecke der Personalauswahl. Die Forschung bescheinigt dieser Methode eine gute Kriteriums- bzw. prädiktive Validität für die spätere Berufsausübung (Bieri & Schuler 2011; Melancon & Williams 2006; Arthur et al. 2003; Gaugler et al 1987), wengleich die Ergebnisse aufgrund sehr unterschiedlicher Konzeptionen der AC, fehlender methodischer Standardisierung und überwiegend kleiner Teilnehmerzahlen ziemlich

heterogen sind. Trotz der umfangreichen Literatur zu AC allgemein gibt es nur wenig Literatur zur Entwicklung von Competency-Based AC (Melanon et al. 2006), obwohl ein solcher Ansatz insbesondere für den Bereich des „Human Resource Development“ als sinnvoll angesehen wird, um bestimmte Verhaltensänderungen zu erfassen (Chen & Naquin 2006). Das Düsseldorfer Curriculum Medizin der Medizinischen Fakultät der Heinrich-Heine-Universität ist kompetenzorientiert und richtet sich an acht Kompetenzen (K) aus: (Humanbiologische K, Präventiv-diagnostisch-therapeutische K, Kommunikative K, Sozial-ethische K, Wissenschaftliche K, Lehr-K, Wirtschaftliche K und Selbst-K). Für jede dieser acht Kompetenzen wurden übergeordnete Lernziele formuliert, die in unterschiedlicher Gewichtung im AC Berücksichtigung fanden.

Methode

In einer Anforderungsanalyse wurden zunächst Qualitäten, Verhaltensweisen und Haltungen definiert, die bei den Studierenden bezüglich einzelner übergeordneter Kompetenzen für das Studium und den späteren Arztberuf wichtig sind. Die Ergebnisse der Anforderungsanalyse fokussierten auf folgenden vier Kompetenzen des Düsseldorfer Curriculums: *Sozial-ethische Kompetenz*, *Kommunikative Kompetenz*, *Selbstkompetenz* und *Lehrkompetenz*.

Für die Konzeption des AC wurde ein „mixed method design“ gewählt, da dieses Design mit einer höheren Validität einhergeht als ein „One-instrument-to-one-competency“ Programm (Schuwirth et al. 2011; Ten Cate et al. 2010, Joiner 2000; International Task Force on Assessment Center Guidelines 2009). Im AC sollten Studierende dann in unterschiedlichen Übungen (s.u.) beobachtet und diese Beobachtungen dann den zuvor definierten Kriterien (s.o.) zugeordnet werden. Als Methoden wurden Fragebogen, Interview, Gruppendiskussion und Rollenspiel ausgewählt und entwickelt.

Durchführung

Die Fragebögen wurden von den Studierenden im Vorfeld elektronisch bearbeitet. Im Anschluss an alle absolvierten Assessmentelemente erhielt jeder Studierende durch einen Assessor ein persönliches Feedback über seine Stärken und Schwächen in den einzelnen Assessmentelementen. Zur Überprüfung der konvergenten und diskriminanten Validität wurden die Ergebnisse der Studierenden im AC über Follow-Up Messungen mit anderen kompetenzorientierten summativen und formativen Prüfungsergebnissen des Düsseldorfer Curriculum Medizin in den Studienjahren 3 und 4 mittels Pearson's Korrelationskoeffizient (r) korreliert. Unterschiede wurden bezüglich Signifikanz mittels z-test nach Meng et al. getestet (Meng et al. 1992).

a) Objective Structured Clinical Examination (OSCE) für ärztliche Basisfertigkeiten (Jahr 3)

Summativer (scheinrelevanter), verpflichtender OSCE mit 9 Prüfungsstationen, wovon in sieben Stationen körperliche Untersuchungstechniken und in jeweils einer die Erhebung einer Anamnese bzw. das Legen einer Venenverweilkanüle geprüft wurden.

b) OSCE für kommunikative Fähigkeiten (Jahr 4) (s.o).

Summativer (scheinrelevanter), verpflichtenden OSCE mit 4 Prüfungsstationen zu speziellen Arzt-Patient Gesprächen unter Einsatz von standardisierten Patienten: „Aggressiv fordernder Patient“, „Überbringung schlechter Nachrichten“, „Schuld und Scham“ und „Shared decision making“.

c) Peer-Feedback (Jahr 4)

Formatives, freiwilliges Assessment. Selbsteinschätzung und wechselseitige Fremdeinschätzung von Kommilitonen einer zugehörigen Gruppe einmal im Semester zu beobachtbaren kommunikativen und sozialen Verhaltensweisen.

Statistik

Es wurde die psychometrische Güte des Fragebogens analysiert und Korrelationen zwischen verschiedenen Messvariablen berechnet. Die Analyse der Korrelationen zwischen den in den AC Elementen erfassten Kompetenzen erfolgte gemäß der Multi-Trait-Multi-Method-Validierung nach Campbell und Fiske (1959).

Ergebnisse

Insgesamt nahmen $n = 26$ Studierenden an allen Elementen des AC teil. Die Ergebnisse der Reliabilitätsanalysen unter Verwendung von Cronbachs α ergaben für den gesamten Fragebogen (79 Items) eine gute Reliabilität mit $\alpha = 0,85$.

Die Studierenden beurteilten sich auf allen Dimensionen – außer bei der Lehrkompetenz – signifikant schlechter als sie von den Assessoren im AC beurteilt wurden, die Effekte waren jedoch moderat. In der Frage, welcher der vier Bestandteile des Assessmenttages (3 AC-Übungen + 1 FB) die Gesamtleistung am besten vorhersagt, zeigte sich, dass das Interview am besten ($r = 0,74, p < 0,01$) und die Gruppendiskussion am wenigsten ($r = 0,54, p < 0,01$) mit der Gesamtleistung korrelierte. Die Messung der Konstruktvalidität war in den Analysen nur schwach und es fanden sich keine signifikanten Korrelationen für die Messungen der einzelnen Kompetenzen mit den verschiedenen Elementen des AC (r 's = -0,29 bis 0,21). Es fand sich eine moderate Korrelation

zwischen der bewerteten sozial-ethischen Kompetenz in der Gruppendiskussion und der im Interview ($r = 0,32$, $p = 0,04$ einseitig).

In Bezug auf prädiktive Effekte des AC (*Tabelle 2*) für ein späteres Verhalten fanden sich keine signifikanten Korrelationen zwischen dem AC und dem OSCE für ärztliche Basisfertigkeiten nach dem 3. Studienjahr, auch nicht mit der kommunikativen Station „Erhebung einer Anamnese“. Es zeigten sich moderate Korrelationen des Ergebnisses aus dem Gesamt AC mit dem Kommunikations-OSCE und ein sehr starker konvergenter Zusammenhang zwischen der im AC gemessenen sozial-ethischen Kompetenz und dem Ergebnis im Kommunikations-OSCE ($r = 0,60$, $p < 0,01$) sowie der im AC gemessenen kommunikativen Kompetenz und dem Peer-Feedback ($r = 0,50$, $p = 0,02$).

Die Ergebnisse der Selbsteinschätzung im Rahmen des Peer-Feedback korrelierten dagegen weder positiv mit dem Gesamtergebnis des AC noch mit der Fremdeinschätzung der Kommilitonen im Peer-Feedback. Auch die kommunikative Kompetenz im AC korrelierte nicht signifikant mit dem Kommunikations-OSCE ($r = 0,17$)

Tabelle 2

Untersuchung der Kriteriumsvalidität des Assessment Centers über die Korrelationen der im AC gemessenen Kompetenzen mit den Ergebnissen anderer kompetenzorientierter Assessments im weiteren Verlauf des Studiums.

Kriteriumsvalidität

AC	n	Mittelwert	Sozial-Eth. K.	Komm.- K.	Selbst- K.	Lehr- K.
Anamnese (im OSCE)	26	0,08	0,15	0,07	-0,14	0,27
OSCE (gesamt)		0,07	0,02	0,20	-0,06	-0,02
CoMeD-OSCE (Mittel)	21	0,41*	0,60**	0,17	-0,03	0,38*
Peer-FB (Selbst)	13	-0,14	-0,13	0,12	0,19	-0,28
Peer FB (Fremd)	18	0,38	0,19	0,50*	0,15	0,35

* p < 0,05; ** p < 0,01; r ≥ 0,50 = starker Effekt; r ≥ 0,30 = mittlerer Effekt; r ≥ 0,10 = geringer Effekt

3.2 Erstellung und Validierung von Messinstrumenten zur Erfassung des Lehr- und Lernklimas

Da die individuelle Wahrnehmung des Lehr-/Lernklimas einen wichtigen Einfluss auf das Verhalten und die in dieser Lernumgebung erzielten Ausbildungsergebnisse / Outcomes hat und ein positives Klima durchschnittlich zu besseren Leistungen führt als ein schlechtes Klima (s.o.) wurde aufgrund fehlender Verfügbarkeit validierter deutschsprachiger Inventare mit den drei nachfolgenden aufgeführten Arbeiten im deutschen Sprachraum erstmalig zwei Fragebögen zur Erfassung des Lehr- und Lernklimas psychometrisch validiert. In Deutschland wurde bisher keine systematisch valide Analyse des Lehr-/Lernklimas an einer Medizinischen Fakultät durchgeführt. Genn zufolge sind Fragebögen Mittel der Wahl zur Messung des Lehr- und Lernklimas an medizinischen Fakultäten (Genn 2001b). Der Vorteil von standardisierten quantitativen

Fragebögen besteht in ihrer einfachen Anwendung und Auswertbarkeit. Darüber hinaus ist ein Vergleich der Ergebnisse beim Einsatz in verschiedenen medizinischen Fakultäten möglich.

Zur Messung des Lehr-/Lernklimas kommen in erster Linie validierte Inventare wie beispielsweise das *Dundee Ready Education Environment Measure (DREEM)* (Roff et al. 1997) zur Anwendung. Aber auch Leitbilder werden als gut geeignet beschrieben, um das Lehr- und Lernklima sowie etwaigen Handlungsbedarf an einer medizinischen Fakultät zu erfassen (Genn 2001a), wiewohl solche Arbeiten bisher noch nicht publiziert wurden. International wurde bisher die Wahrnehmung des Lehr-Lernklimas auch vorwiegend bei Studierenden und kaum bei Lehrenden untersucht (Fraser 1986; Genn 2001b; Genn 2001a; Miles & Leinster 2009). Curriculare Veränderungen sind jedoch gerade auch für die Lehrenden eine besondere Herausforderung (Till 2005).

Nachfolgend werden eigenen Untersuchungen in Form einer Adaptation des DREEM (*Dundee Ready Education Environment Measure*) (Roff et al.1997) und eine Fragebogen-Neuentwicklung aus dem Leitbild „Lehre“ der Medizinischen Fakultät der HHUD (Leitbild Lehre Med. Fakultät der HHU 25.6.2009) integrativ und vergleichend dargestellt.

Publikation 5

Rothhoff T, Ostapczuk M, de Bruin J, Decking U, Schneider M, Ritz-Timme S. **Assessing the learning environment of a Faculty. Validation of the Dundee Ready Education Environment Measure (DREEM) with Students and Teachers.** *Med Teach.* 2011; 33(11):e624-36.

Publikation 6

Ostapczuk M, Hugger A, de Bruin J, Ritz-Timme S, Rothhoff T. **DREEM on, dentists! Students' perception of the educational environment in a German dental school as measured by the Dundee Ready Education Environment Measure.** *Eur J Dent Educ.* 2012; 16(2):67-77

Publikation 7

Rothhoff T, Ostapczuk M, de Bruin J, Kröncke K, Decking U, Schneider M, Ritz-Timme S. **Development and evaluation of a questionnaire to measure the perceived implementation of the mission statement of a competency based curriculum.** *BMC Med Educ* 2012; 12:109

Das *Dundee Ready Education Environment Measure (DREEM)* (Roff et al. 1997) wurde als ein nicht kulturell spezifisches Instrument zur Erfassung des Lehr- und Lernklimas in der Wahrnehmung von Studierenden in den Gesundheitsberufen entwickelt (Roff 2005). Der DREEM besteht in seiner heutigen Form aus 50 Items, die auf einer Skala von 0 („*Stimme überhaupt nicht zu*“) bis 4

(„*Stimme stark zu*“) beantwortet werden müssen. Maximal können 200 Punkte erzielt werden, die sich auf fünf Subskalen verteilen: Wahrnehmung der Lehre, Wahrnehmung der Lehrenden, Akademische Selbstwahrnehmung, Wahrnehmung der Atmosphäre und Soziale Selbstwahrnehmung. Lai et al. (2009) sowie McAleer und Roff (2001) haben ein Interpretationssystem der erzielbaren Punktzahlen vorgeschlagen, das in den meisten publizierten DREEM-Studien auch verwendet wird (Ostapczuk 2014).

Für den Gesamtwert gilt dabei:

- *0-50 Punkte*: Sehr schlechtes Lehr- und Lernklima
- *51-100 Punkte*: Viele Probleme
- *101-150 Punkte*: Mehr positive als negative Aspekte
- *151-200 Punkte*: Exzellentes Lehr- und Lernklima

Betrachtet man ausschließlich bei Studierenden der Medizin durchgeführte Untersuchungen mit dem DREEM, so liegt das nach Stichprobengröße gewichtete Mittel bei 121.0 Punkten (Ostapczuk 2014; S. 12-14).

Auch für Leitbilder wird postuliert, dass über sie das Lehr- und Lernklima gemessen werden kann (Genn 2001a). Das Leitbild Lehre der Medizinischen Fakultät der HHU wurde von einer von der Fakultät eingesetzten Arbeitsgruppe unter Beteiligung von Studierenden erarbeitet und von der Fakultät verabschiedet. Es bildete die Grundlage für eine Curriculumreform. Wie mehrere Autoren betonen (Vgl. Bouhaimed, Thalib & Doi 2009; Rothman & Ayoade 1970), ist die Untersuchung des Lehr- und Lernklimas ein erster Schritt bei der Implementierung eines neuen Curriculums, um nachfolgend mit wiederholten Messungen zu überprüfen, ob Maßnahmen zu angedachten Veränderungen bzw. Verbesserungen geführt haben. Um den aktuellen Umsetzungsgrad des Leitbildes in Bezug auf die Zielvorstellungen zu erfassen, wurden aus dem Leitbild Items für einen Fragebogen abgeleitet.

Methode

Die Befragungen mit dem kombinierten DREEM und Leitbild-Fragebogen fanden zwischen Juli 2010 und Februar 2011 in den Studiengängen Human- und Zahnmedizin statt. In beiden Studiengängen war die Teilnahme für die Studierenden und Lehrenden freiwillig und anonym. Sowohl bei Studierenden als auch bei den Lehrenden wurden demographische Informationen u.a. zu Geschlecht, Alter und Muttersprache erfragt. Die deutschsprachige Version des DREEM wurde nach der durch Jones et al. (2001) modifizierten Rückübersetzungsmethode von Brislin (1970) angefertigt (Ostapczuk 2014, S. 44). Der Fragebogen wurde in zwei Versionen - als Studierendenversion und als Version für die Lehrenden - erstellt.

Statistik

Für die Untersuchung der Gütekriterien des Fragebogens wurde für die Bestimmung der Homogenität der Items und Bestimmung der Konstruktvalidität eine explorative Faktorenanalyse durchgeführt, außerdem wurde die korrigierte Trennschärfe der Items und die Reliabilität des Tests über die interne Konsistenz mit der Berechnung von Cronbach's alpha bestimmt. Des Weiteren fand eine Prüfung hinsichtlich Unterschieden zwischen den Geschlechtern statt. Gruppenunterschiede wurden mittels univariater Varianzanalyse analysiert und Zusammenhänge zwischen intervallskalierten Daten durch Berechnung von Pearson-Produkt-Moment-Korrelationen überprüft (Ostapczuk 2014, S. 50).

Ausgewählte Ergebnisse

Im Studiengang Humanmedizin der HHUD konnten Fragebögen von 1.119 Studierenden (=55,0 %; davon 66 % Frauen) und 258 Lehrenden (19,9 %; davon 27,9 % Frauen) ausgewertet werden. Im Studiengang Zahnmedizin konnten Fragebögen von 205 Studierenden (65,9%; davon 74,6 % Frauen; Durchschnittsalter = 23,2 Jahre, $SD = 3,2$) ausgewertet werden. Da hier nur acht Lehrende an der Befragung teilnahmen werden diese Daten der Zahnmedizin nicht berücksichtigt. Für beide Fragebögen - DREEM und Leitbild-FB - kann eine hohe Inhaltsvalidität angenommen werden, da die Items des DREEM und das Leitbild der Medizinischen Fakultät der HHUD jeweils von Experten mit Hilfe des Delphi-Verfahrens entwickelt wurden (Linstone & Turoff, 1975; Wottawa & Hossiep 1987). Beide Fragebögen korrelierten bei Studierenden und Lehrenden im Sinne der konvergenten Validität hoch positiv und signifikant miteinander. Hieraus kann hypothesenkonform gemäß Genn (Genn 2001) abgeleitet werden, dass der Düsseldorfer Leitbildfragebogen ebenfalls das Lehr-Lernklima misst (Tabelle 3).

Tabelle 3: Konvergente Validität von DREEM und Leitbild Fragebogen bei Studierenden und Lehrenden der Humanmedizin.

	Studierende (n = 1119)			Lehrende (n = 258)		
	M (Punkte) (SD)	DREEM (α)	Leitbild- FB	M (Punkte) (SD)	DREEM (α)	Leitbild- FB
DREEM- Gesamtwert	109,75 (21,71)	0,92	1	117,63 (20,80)	0,94	1
Leitbild- Gesamtwert	77,12 (16,97)	0,92	0,79***	79,43 (17,24)	0,93	0,80***

Anmerkungen: α = Cronbachs α (interne Konsistenz); FB = Fragebogen; M (SD) = Mittelwert (Standardabweichung); *** $p < 0,001$ im entsprechenden t-Test für Korrelationskoeffizienten (gerichtet).

Herausgestellt werden nachfolgend die Vergleiche in der Wahrnehmung des Lehr-/Lernklimas zwischen Studierenden und Lehrenden der Humanmedizin sowie zwischen Studierenden der Human- und Zahnmedizin.

Studierende und Lehrende der Humanmedizin

Die Studierenden der Humanmedizin nahmen sowohl das Klima über den DREEM, als auch die Umsetzung des Leitbildes negativer wahr als die Lehrenden. Beide Gruppen beurteilten sich selbst besser als sie von der jeweils anderen Gruppe beurteilt wurden (sowohl im DREEM- als auch im Leitbild-Fragebogen).

Über den DREEM Fragebogen wurde das Klima von den Lehrenden signifikant besser wahrgenommen als von den Studierenden ($p < 0,001$), wenngleich der Effekt auch eher klein war $d = 0,37^4$. Auf den fünf DREEM-Subskalen beurteilten die Studierenden ($M = 22,92$, $SD = 6,91$) die Lehre „negativ“ (13-24 Punkte, vgl. Lai et al., 2009; McAleer & Roff, 2001), die Lehrenden ($M = 27,05$, $SD = 6,85$) dagegen „eher positiv“ (25-36 Punkte), $p < 0,001$, $d = 0,60$ (mittlerer Effekt). Gemessen mit dem Leitbildfragebogen empfanden die Studierenden im Vergleich zu den Lehrenden das Leitbild als weniger gut umgesetzt ($p < 0,05$), wobei der Effekt jedoch klein war ($d = 0,14$). Auf den Subskalen des Leitbildfragebogens beurteilten die Studierenden ($M = 15,11$, $SD = 4,67$) die Lehrenden schlechter als die Lehrenden sich selber ($M = 17,57$, $SD = 4,73$) ($p < 0,001$) $d = 0,52$ (mittlerer Effekt). Sich selbst beurteilten die Studierenden besser ($M = 33,65$, $SD = 7,35$) als sie von den Lehrenden eingeschätzt wurden ($M = 32,57$, $SD = 7,38$), $p < 0,05$; $d = 0,15$ (kleiner Effekt). Für die Beurteilung des Studiums ergaben sich keine signifikanten Unterschiede zwischen den Studierenden und Lehrenden (*Tabelle 4*).

⁴ (Für das Effektstärkemaß d gelten bei t-Tests gemäß Cohen (1988) folgende Richtlinien: $d = 0,20$ – kleiner Effekt; $d = 0,50$ – mittlerer Effekt; $d = 0,80$ – großer Effekt.).

Tabelle 4

Skalenwerte des DREEM und Leitbildfragebogens bei Studierenden und Lehrenden der Humanmedizin. Dargestellt sind die maximal erreichbaren Punktzahlen beider Fragebögen und der verschiedenen Fragebogenkategorien sowie die im Mittel von den Teilnehmern erzielten Punktzahlen (M). Die Interne Konsistenz (Cronbachs α) gibt die Reliabilität der Messung an.

Fragebogen	Studierende <i>n</i> = 1119			Lehrende <i>n</i> = 258			
	Maximal erreichbare Punktzahl	Interne Konsistenz α	M (SD)	Min-Max	Interne Konsistenz α	M (SD)	Min-Max
DREEM-Gesamtwert (50 Items)	200	0,92	109,75*** (21,71)	28-182	0,94	117,63*** (20,80)	52-182
Wahrnehmung der Lehre (12 Items)	48	0,84	22,92*** (6,91)	2-47	0,86	27,05*** (6,85)	5-47
Wahrnehmung der Lehrenden (11 Items)	44	0,75	24,33*** (5,08)	5-44	0,80	29,10*** (5,32)	12-44
Akadem. Selbstwahrnehmung (8 Items)	32	0,68	17,39*** (4,47)	0-31	0,67	17,50*** (3,78)	5-26
Wahrnehmung der Atmosphäre (12 Items)	48	0,75	27,39*** (6,04)	5-48	0,80	29,01*** (5,40)	13-45
Soziale Selbstwahrnehmung (7 Items)	28	0,57	17,71*** (3,68)	2-26	0,53	14,97*** (2,76)	6-23
Leitbild-Gesamtwert (37 Items)	148	0,92	77,12*** (16,97)	10-144	0,93	79,43*** (17,24)	20-133
Lehrende (8 Items)	32	0,81	15,11*** (4,67)	0-32	0,82	17,57*** (4,73)	4-31
Studierende (15 Items)	60	0,83	33,65*** (7,35)	6-60	0,87	32,57*** (7,38)	10-55
Curriculum (14 Items)	56	0,81	28,36*** (7,06)**	3-55	0,83	29,29*** (7,02)	4-49

Anmerkungen: α = Cronbachs α ; M (SD) = Mittelwert (Standardabweichung); Min-Max = minimal und maximal erreichter Wert; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ im entsprechenden t-Test für unabhängige Stichproben (ungerichtet).

Studierende der Zahnmedizin im Vergleich mit Studierenden der Humanmedizin

Das Klima insgesamt (beide Fragebögen) wurde von der Mehrheit der Zahnmedizinierenden (90.2%) mit „eher gut“ bewertet (129-297 Punkte)⁵. Im Vergleich zu den international publizierten Studien, in denen der DREEM bei Studierenden unterschiedlicher gesundheitswissenschaftlicher Fächer eingesetzt wurde (nach Stichprobengröße gewichteter Mittelwert $M = 120,3$), empfanden die Studierenden der Zahnmedizin ($M = 122,95$, $SD = 15,52$) – anders als ihre Kommilitonen der Humanmedizin – das Lehr- und Lernklima an der HHUD sogar als signifikant etwas besser, $t(204) = 2,45$, $p < 0,05$ $d = 0,17$ (kleiner Effekt) (Ostapczuk 2014; S.93) (Tabelle 5).

Tabelle 5

Skalenwerte des DREEM und Leitbildfragebogens bei Studierenden der Human- und Zahnmedizin

Fragebogen	Studierende $n = 1119$	Humanmedizin		Studierende Zahnmedizin			
		Interne Konsistenz α	M (SD)	Min- Max	Interne Konsistenz α	M (SD)	Min- Max
DREEM- Gesamtwert (50 Items)	200	0,92	109,75 (21,71)	28-182	0,87	122,95 (15,52)	81-171
Leitbild-FB Gesamtwert (37 Items)	148	0,92	77,12** (16,97)	10-144	0,88	88,32** (12,78)	48-126

Anmerkungen: α = Cronbachs α ; M (SD) = Mittelwert (Standardabweichung); *Min-Max* = minimal und maximal erreichter Wert; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ im entsprechenden t-Test für unabhängige Stichproben (ungerichtet).

Bei der Betrachtung ausgewählter Items aus beiden Fragebögen, die eine besondere Relevanz für das Thema Feedback und Feedbackkultur haben, fanden sich deutliche Unterschiede zwischen Studierenden und Lehrenden der Humanmedizin und auch zwischen Studierenden der Human- und Zahnmedizin (Tabelle 6).

⁵ Addition der jeweiligen Minimal- und Maximalwerte von DREEM und Leitbild FB bei Studierenden der Zahnmedizin

Tabelle 6

Mittelwerte, Standardabweichungen und part-whole-korrigierte Trennschärfen ausgewählter Items des DREEM und Leitbildfragebogens bei Studierenden und Lehrenden der Humanmedizin sowie Studierenden der Zahnmedizin; Skala der Items von 0 („Stimme überhaupt nicht zu“) bis 4 („stimme stark zu“)

Item	Studierende	Lehrende	Studierende
	Humanmedizin <i>n</i> = 1119 <i>M</i> (<i>SD</i>)	Humanmedizin <i>n</i> = 258 <i>M</i> (<i>SD</i>)	Zahnmedizin <i>n</i> =205 <i>M</i> (<i>SD</i>)
<i>Leitbild</i>			
1. Die Lehrenden sind Vorbild für die Studierenden.	1,91 (0,91)	2,34 (0,84)	2,31 (0,86)*
2. Die Lehrenden stehen in lebendigem Dialog mit den Studierenden und anderen Lehrenden.	1,80 (0,95)	2,28 (0,97)	2,52 (0,72)*
4. Die Lehrenden geben stimulierendes Feedback.	1,68 (0,90)	2,24 (0,84)	2,22 (0,83)*
5. Die Lehrenden sind Ansprechpartner und Vertrauenspersonen für die Studierenden.	1,34 (0,98)	2,18 (0,86)	1,98 (0,92)*
6. Die Lehrenden erhalten für ihre Tätigkeit Anerkennung von der Fakultät.	1,91 (0,82)	1,24 (0,98)	2,25 (0,75)*
<i>DREEM</i>			
9. Die Lehre ist studierendenzentriert.	1,71 (0,96)	2,10 (0,94)	2,20 (0,80)
13. Die Lehrenden sind gut darin, Studierenden Feedback zu geben.	1,45 (0,91)	2,04 (0,93)	2,20 (0,87)

Anmerkungen: *M* (*SD*) = Mittelwert (Standardabweichung); **p* < 0,05 im entsprechenden t-Test für unabhängige Stichproben (ungerichtet)

Für verschiedene demographische Variablen wurde ebenfalls der Einfluss bzw. Zusammenhang auf das Lehr- und Lernklima explorativ untersucht (Tabelle 7).

Tabelle 7

Einfluss von Geschlecht, Muttersprache und abgeleiteter Ausbildung vor dem Studium bei Studierenden der Humanmedizin.

	Frauen <i>n</i> =739	Männer <i>n</i> =380	Mutter- sprachler <i>n</i> =985	Nicht-Mutter- sprachler <i>n</i> =134	Studierende ohne Ausbildung <i>n</i> =972	Studierende mit Ausbildung <i>n</i> =241
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
DREEM	110,96* (20,72)	107,41* (23,35)	109,04** (21,68)	114,95** (21,24)	111,14§ (21,04)	104,67§ (22,48)
Leitbild- Fragebogen	78,06% (16,06)	75,28% (18,51)	85,12° (18,73)	76,03° (16,43)	78,42§ (16,46)	72,17§ (17,55)

M (SD) = Mittelwert (Standardabweichung); * $p < 0,01$, $d = 0,16$; ** $p < 0,01$, $d = 0,28$; § $p < 0,001$, $d = 0,30$; ° $p < 0,01$, $d = 0,16$; ° $p < 0,001$, $d = 0,52$; § $p < 0,001$, $d = 0,37$

Bei den Studierenden der Zahnmedizin fanden sich keine signifikanten Unterschiede für die Gesamtskalen beider Fragebögen bezüglich der aufgeführten Parameter. Mit fortschreitendem Studium bewerteten die Studierenden im DREEM speziell die Lehre zunehmend schlechter als die Studierenden in früheren Semestern (Vorklinik: $M = 23,91$, $SD = 6,52$; Klinik: $M = 22,08$, $SD = 7,14$; $p < 0,001$; $d = 0,27$), dafür jedoch die Atmosphäre signifikant besser (Vorklinik: $M = 26,56$, $SD = 5,89$; Klinik: $M = 28,12$, $SD = 6,07$, $p < 0,001$; $d = 0,26$). Studierende des klinischen Studienabschnitts bewerteten die Umsetzung des Leitbildes im Vergleich zu ihren Kommilitonen der Vorklinik als schlechter.

4 Diskussion

Die eigenen Arbeiten werden nun bezüglich ihres Beitrages zur Evidenz von Validität und Reliabilität für kompetenzbasierten Prüfungen diskutiert. Es folgt dann ein Ausblick, wie kompetenzorientiertes Prüfen künftig bestmöglich umgesetzt werden kann. Die Diskussion schließt mit einer Analyse der eigenen Arbeiten zur Messung des Lehr-/Lernklimas und seiner Bedeutung für das kompetenzorientierte Prüfen.

4.1 Geschaffene Evidenz zur Validität und Reliabilität von kompetenzorientierten Prüfungen

Prüfung von Teilaspekten einer Kompetenz (diagnostisch-therapeutische Kompetenz)

Wir konnten mit einer computer- und fallbasierten „Key-Feature“ Prüfung zeigen, dass klinische Entscheidungsfindung (Handlungswissen) als Teilaspekt der diagnostisch-therapeutischen Kompetenz summativ über Long Menu Fragen geprüft werden kann (Rotthoff et al. 2006a; Rotthoff et al. 2006b). Gemäß den „best practice“ Empfehlungen wurde hier die Verknüpfung des gemessenen Wissens mit einer definierten Kompetenz (McClarty & Gaertner 2015) umgesetzt.

Die Ergebnisse zeigten, dass Long Menu Fragen (und Antworten) gleichwertig einsetzbar sind zu Freitextfragen („short answer questions“) (Rotthoff et al. 2006b). Im Gegensatz zur manuellen Auswertung der Antworten von Freitextfragen können die Antworten von Long Menu Fragen automatisiert und korrekt ausgewertet werden (Rotthoff et al. 2006b). Als Herausforderung stellte sich dabei die präzise und fehlerfreie Vorformulierung der Long Menu Antworten sowie eine umfassende Berücksichtigung ebenfalls korrekter Synonyme als Antwortoptionen dar. In der reduzierten Variabilität der Antwortformulierung sehen manche Autoren allerdings eine Einschränkung für die Nutzung dieses Fragentyps in summativen computerbasierten Prüfungen (van Bruggen et al. 2012). Unsere Untersuchungsergebnisse zeigen jedoch, dass die Verwendung kurzer (begrifflicher) Long Menu Antworten die Variabilität in der Antwortformulierung reduziert, eine umfassende Definition synonyme Begriffe erlaubt und darüber eine automatisierte Auswertung ermöglicht. Für die Frage nach klar definierten Begriffen wie z.B. einer Diagnose, einem diagnostischen Verfahren oder einer Therapie, hat sich das Fragenformat in der vorliegenden Untersuchung als zuverlässig erwiesen. Prosaantworten sind dagegen für die computerbasierte Auswertung ungeeignet, da diese nicht in allen Formulierungsvariationen synonym verschlüsselt werden können und damit eindeutig messbare Ergebnisse erschweren. Limitierend für die Aussagekraft dieser Untersuchung, ist die geringe Anzahl ($n = 9$) getesteter

Long Menu Fragen aus dem Bereich der Inneren Medizin. Ob Fragen mit anderen Fachinhalten zu ähnlichen Messergebnissen gekommen wären, bleibt zu prüfen. Da sich jedoch weniger der Inhalt einer Long Menu Antwort, sondern vielmehr deren sprachliche Morphologie und Semantik als entscheidend für die Zuverlässigkeit des Messergebnisses zeigte, kann eine Kontextunabhängigkeit bei Beachtung der Formulierung von kurzen (begrifflichen) Antworten postuliert werden. Long Menu Fragen sollten dabei nicht alternativ zu MC Fragen, sondern ergänzend in einer computerisierten Prüfung für das Setzen eines anderen Prüfungsstimulus verwendet werden. Der Einbettung von Prüfungsfragen in einen klinischen Fall wird für das Prüfen von klinischer Entscheidungsfindung dabei die höchste Bedeutung beigemessen (Schuwirth & van der Vleuten 2004; van Bruggen et al. 2012). Auch wenn klinische Entscheidungsfindung als Bestandteil der diagnostisch-therapeutischen Kompetenz inhaltlich valide und reliabel mit einer computerbasierten Key-Feature Prüfung getestet werden konnte, bleibt mit der vorgelegten Untersuchung unbeantwortet, inwieweit das Ergebnis eines Prüflings dessen klinische Entscheidungsfindung im realen klinischen Kontext (= Performanz) vorhersagen kann.

Computerbasierte Prüfungen mit Key Feature Fällen und Long Menu Fragen werden heute in internationalen Prüfungsverbänden wie dem Umbrella Consortium for Assessment Networks (UCAN) mit mehr als 40 nationalen und internationalen Partnern verwendet (<https://www.ucan-assess.org/cms/de/Quelle> URL UCAN). Limitierend bei der Durchführung computerbasierter Prüfungen ist der Ressourcenaufwand für die Unterhaltung von (stationären) Computersystemen. Zunehmend werden daher heute Tablet-basierte Prüfungen durchgeführt, die im Hörsaal möglich sind und keine eigenen Computerräume benötigen.

Prüfung einer spezifischen Kompetenz (Kommunikative Kompetenz)

Im Unterschied zur computerbasierten Key-Feature Prüfung, welche Teilaspekte der diagnostisch, therapeutischen Kompetenz prüfte, berücksichtigte der Kommunikations-OSCE die Prüfung einer „gesamten“ Kompetenz. Wir konnten zeigen, dass anspruchsvolle ärztliche Gesprächssituationen in einem OSCE *prinzipiell* reliabel, valide und summativ prüfbar sind.

Limitierend für die Durchführung als summative (hier scheinrelevante) Prüfung, muss die mit 0,64 nicht ausreichende Gesamtreliabilität der Prüfung angemerkt werden. In der Literatur wird allerdings auf eine steigende Reliabilität eines OSCE mit Erhöhung der Anzahl an Prüfungsstationen hingewiesen: so wird für OSCE mit < 10 Stationen ein alpha von 0,56 und für OSCE mit >10 Stationen ein alpha von 0,74 beschrieben (Brannick, Erol-Korkmaz & Prewett 2011). Nach Berechnung mit der Spearman-Brown-Formel hätte die Reliabilität des vorgelegten OSCE mit zwei weiteren Stationen auf 0,7 und mit fünf weiteren Stationen auf 0,8 erhöht werden

können. In der Literatur wird eine Reliabilität von 0,8 als Zielgröße für summative Prüfungen genannt (Möltner et al. 2006).

Das im vorliegenden OSCE mit nur vier Stationen erzielte Cronbachs alpha von 0,64 kann damit als ein sehr gutes Ergebnis bewertet werden, zumal für OSCE Stationen mit kommunikativen Inhalten im Vergleich zu Prüfungsstationen mit anderen Fertigkeiten eher eine geringere Reliabilität beschrieben wird (Brannick, Erol-Korkmaz & Prewett 2011). Eine Erklärung für diese sehr gute Reliabilität des vorliegenden OSCE trotz geringer Zahl an Prüfungsstationen kann in der inhaltlich gleichen Ausrichtung sämtlicher OSCE-Stationen liegen. In allen Stationen wurde die kommunikative Kompetenz geprüft. In der Literatur werden für Testelemente, die ähnlich bezüglich des Inhaltes sind, höhere Korrelationen beschrieben als für Testelemente unterschiedlichen Inhalts, aber gleichen Formates (Schuwirth & Van der Vleuten 2012; S. 122). In weiteren eigenen Untersuchungen konnten wir diese Aussage bestätigen. So erreichten Studierende in OSCE Prüfungen für ärztliche Basisfertigkeiten mit jeweils nur einer Station zur kommunikativen Kompetenz (Anamnese) in dieser Station im Mittel weniger Punkte als in den anderen Prüfungsstationen mit Fokus auf manuelle Fertigkeiten (Vgl. Rotthoff et al. 2008; Rotthoff et al. 2014).

Im analysierten OSCE wurden vier komplexe kommunikative ärztliche Gesprächssituationen geprüft. Die inhaltliche Validität des OSCE kann in Bezug auf die Messung der allgemeinen ärztlichen kommunikativen Kompetenz als gut bewertet werden. Wie gut eignete sich das gewählte Messinstrument (BGR) für die Messung der Kommunikativen Kompetenz vor dem Hintergrund, dass komplexe Kompetenzen als kaum zuverlässig erfassbar beschrieben werden (Huddle & Heudebert 2007)?

Reliabilität der Messungen – Globales Rating versus Checkliste

Mit dem globalen Rating wurde jeweils die Gesamtdimension der kommunikativen Kompetenz in den einzelnen Stationen getestet. Obwohl keine differenzierte Prüfung spezifischer Stationsinhalte stattfand, weist die gute Unterscheidbarkeit der Prüfungsstationen in drei von vier Fällen auf eine erfolgreiche Messung verschiedener inhaltlicher Dimensionen mit der verwendeten globalen Rating Skala (BGR) hin. Die Ergebnisse unterstützen die Befunde, dass ein globales Rating in Bezug auf Reliabilität, interne Konsistenz und Validität gegenüber der Verwendung von Checklisten als ebenbürtig angesehen werden kann (Regehr et al. 1998). Auch weitere Kompetenzebenen wie Empathie und ethische Haltungen konnten ebenfalls gut abgebildet werden, was in der Literatur als Vorteil von globalen Ratings gegenüber Checklisten beschrieben wird (Nikendei & Jünger 2006). So kann abweichendes Verhalten, das auf einer Checkliste nicht

aufgeführt ist, vom Prüfer weder honoriert noch sanktioniert werden (Scheffer 2009, S. 20) und auch die Expertise von Prüfern wird beim Einsatz detaillierter Checklisten deutlich weniger genutzt (ebd.).

Interrater Reliabilität

Die Interrater Reliabilität im Kommunikations-OSCE war zufriedenstellend, für eine summative Prüfung jedoch verbesserungsbedürftig (Mortsiefer, Immecke & Rothhoff et al. 2014). Es stellt sich die Frage, mit welchen Maßnahmen die Interrater-Reliabilität beeinflusst und verbessert werden kann. In einer weiteren Untersuchung konnten wir zeigen, dass insbesondere dasselbe Geschlecht, eine große Rater-Erfahrung und ein klinischer beruflicher Hintergrund die Interrater Reliabilität signifikant positiv beeinflussen, eine Raterschulung dagegen keinen Effekt zeigte (Mortsiefer, Karger & Rothhoff et al. 2017). Diese für das Prüfungsergebnis relevanten Rater-Faktoren könnten künftig in einem simulierten Prüfungssetting wie dem Kommunikations-OSCE prinzipiell berücksichtigt werden, jedoch ginge dieses zu Lasten der Umsetzbarkeit durch eine geringere Verfügbarkeit von Prüferinnen und Prüfern.

Für die Verwendung des Messinstrumentes (Berliner Global Rating) zur Performanzmessung am ärztlichen Arbeitsplatz können solche Prüfermerkmale insbesondere im ambulanten, niedergelassenen Bereich nicht mehr berücksichtigt werden. In einer Untersuchung zur Prüfung der Kommunikativen Fähigkeiten von Studierenden mit Patientinnen und Patienten in Praxen der primärärztlichen Versorgung konnten wir bei Verwendung der BGR nur eine geringe Übereinstimmung zwischen den Bewertungen durch die Lehrärzte in den Praxen und externen Prüfern messen, welche ihre Bewertung auf Basis von Videoaufzeichnungen dieser Prüfungen vornahmen (Weyers et al. 2016). Hieraus lässt sich ableiten, dass im Falle der Verwendung des Messinstrumentes (BGR) am ärztlichen Arbeitsplatz die teststatistischen Anforderungen an eine summative Prüfung hinsichtlich Objektivität und Reliabilität nicht erfüllt werden können. Die Verwendung des Global Ratings für die Performanzmessung sollte daher aufgrund unzureichender Standardisierungsmöglichkeit nur als formatives Prüfungsinstrument für eine strukturierte Rückmeldung am Arbeitsplatz eingesetzt werden.

Die Untersuchung zum Kommunikations-OSCE zeigt, dass eine Prüfung der Kommunikativen Kompetenz in einer simulierten und standardisierten OSCE-Prüfung inhaltlich valide möglich ist. Mit Umsetzung der zuvor genannten Maßnahmen zur Verbesserung der Reliabilität, wie die Aufnahme zusätzlicher Prüfungsstation(en) und eine gezieltere Auswahl der Rater, ist die Prüfung auch summativ reliabel und objektiv durchführbar. Inwieweit die in einer simulierten Prüfungsumgebung gemessene kommunikative Kompetenz die spätere Performanz der

Studierenden im ärztlichen Arbeitsumfeld vorhersagen kann (prädiktive Validität), kann mit der vorgelegten Untersuchung jedoch nicht beantwortet werden.

Dass Prüfungsergebnisse aus einer simulierter Umgebung (OSCE) die spätere Performanz vorhersagen können, konnte Norcini et al. in einer Untersuchung zur United States Medical Licensing Examination (USMLE) belegen. In dieser Untersuchung fand sich eine signifikant geringere Mortalität von Patienten in einer Klinik, je besser die behandelnden Ärzte im Step 2 der USMLE abgeschnitten hatten (Norcini et al. 2014). Im Step 2 der USMLE Prüfung findet neben einer Wissensprüfung (Clinical Knowledge) auch eine klinisch praktische Prüfung (Clinical Skills) statt. In dem klinisch praktischen Teil durchlaufen die Prüflinge 12 klinische Situationen, welche von standardisierten Simulationspatienten dargestellt werden. Bewertet werden in dieser Prüfung neben den kommunikativen Fähigkeiten inklusive der Anamneseerhebung auch die Durchführung einer fokussierten körperlichen Untersuchung und die Erstellung eines Behandlungsplans (Vgl. USMLE - Federation of State Medical Boards of the United States and the National Board of Medical Examiners 2013-2016). Mit qualitativ hochwertig simulierten Patientenfällen kann in Kombination mit einer Wissensprüfung also kriteriumsvalide und zuverlässig geprüft werden. Die USMLE prüft dabei Kompetenzen nicht separat, sondern integrativ an realistisch simulierten klinischen Situationen.

Hieraus leitet sich die Frage ab, ob Kompetenzen überhaupt trennscharf einzeln geprüft werden können. Wie gut kommen unterschiedliche Messinstrumente bezüglich einer Kompetenz zum selben Messergebnis? Diese Fragen werden nachfolgend am Beispiel der Untersuchung zum Assessment Center diskutiert.

Beziehung der Messung einer Kompetenz in verschiedenen Messverfahren

Die Messung der Konstruktvalidität des Assessment Centers (AC) war in den Analysen nur schwach und es fanden sich keine signifikanten Korrelationen für die Messungen der einzelnen Kompetenzen mit den verschiedenen Formaten des AC ($r's = -0,29$ bis $0,21$) (Rotthoff et al. 2014). Hieraus kann abgeleitet werden, dass die Messung derselben Kompetenz mit unterschiedlichen Messinstrumenten im AC nicht sicher zum selben Ergebnis kommt, was auch in der Literatur für AC beschrieben und kritisch diskutiert wird (Woehr & Arthur 2001; Goodstone & Lopez 2001). Die Konstruktvalidität hätte wahrscheinlich über eine Reduktion der Anzahl zu messender Kompetenzen (Arthur et al. 2003; Gaugler et al. 1987) oder auch über den Einsatz eines zweiten Raters für jedes AC Element erhöht werden können (Gaugler et al. 1987). Gerade bezüglich des Einsatzes eines zweiten Raters für jedes AC-Element stellt sich die Frage nach der Machbarkeit im Fakultätsbetrieb mit einer als angespannt wahrgenommenen Personalsituation.

Trennschärfe bei der Messung einzelner Kompetenzen

Die Ergebnisse zeigen außerdem, dass eine trennscharfe Unterscheidung von einzelnen Kompetenzen wie beispielsweise der kommunikativen von der sozial-ethischen Kompetenz sehr schwierig ist. Für den vorgelegten AC wurde beispielsweise im Rollenspiel das Kriterium „Benennt die Problempunkte offen“ von der Expertengruppe der sozial-ethischen Kompetenz und das Kriterium „Monologisiert, lässt Gegenüber kaum zu Wort kommen“ der kommunikativen Kompetenz zugeordnet. Hier ist durchaus auch eine gegenteilige kompetenzbezogene Zuordnung der Kriterien denkbar. Durch diese Unschärfe in der Kompetenzzuordnung wird die Validität geschwächt, denn die Validität einer Prüfung wird umso stärker, je klarer umrissen das zu messende Konstrukt ist (Lienert & Ratz 1998). Die inhaltliche Validität wird einem Test u.a. durch ein Rating von Experten zugebilligt (ebd.), welche sich in diesem Fall für die genannte Zuordnung entschieden haben.

Ob über einen Konsensusprozess von Expertinnen und Experten eine allgemeingültige und trennscharfe Zuordnung von Kriterien zu einzelnen Kompetenzen überhaupt möglich ist, muss bezweifelt werden.

Trotz der Schwächen in der Konstruktvalidität, waren die im AC integrativ gemessenen Kompetenzen gute Prädiktoren für das spätere Verhalten der Studierenden im Follow-Up (Rotthoff et al. 2014). Diese Zusammenhänge von Kompetenzmessung und späteren Messergebnissen bezüglich derselben Kompetenz werden nun genauer analysiert.

Prädiktive Validität von Kompetenzmessungen

Gemäß der „best practice“ Empfehlung für einen Bezug kompetenzorientierter Prüfungen mit späteren Outcomes im Sinne der prädiktiven Validität (McClarty & Gaertner 2015) konnten wir für den Assessment Center nachweisen, dass die Messergebnisse für die Sozial-ethische und Kommunikative Kompetenz die Performanz der Studierenden über einen Zeitraum von vier Jahren recht gut vorhersagen konnten (Rotthoff et al. 2014). Diese Schlussfolgerung leitet sich aus den hoch positiven Korrelationen zwischen der Sozial-ethischen Kompetenz und dem Kommunikation-OSCE ($r = 0,60$) sowie der Kommunikativen Kompetenz und dem Peer-Feedback ($r = 0,50$) ab. Studierende, die im AC des ersten Studienjahres schlecht bewertet wurden, wurden auch im Umgang mit Schauspielpatienten oder Kommilitonen vier Jahre später von anderen „Prüfern“ nicht wesentlich besser bewertet. Die vorliegenden Ergebnisse unterstützen damit bisherige Forschungsergebnisse, welche AC insgesamt eine gute Kriteriums- bzw. prädiktive Validität für die spätere Berufsausübung zuschreiben (Bieri & Schuler 2011; Melancon et al. 2006; Arthur et al. 2003; Gaugler et al 1987).

Es scheint auf den ersten Blick überraschend, dass die kommunikative Kompetenz im AC nicht gut mit den Ergebnissen im Kommunikations-OSCE korrelierte. Erklärbar ist dieses mit der Kontextspezifität von kommunikativen Situationen. Fokussierten die Stationen des Kommunikations-OSCE wesentlich auf sehr spezifische Aspekte der Arzt-Patient Beziehung und die Umsetzung von Kommunikationsmodellen, so standen im AC mehr allgemeine Aspekte der Kommunikation wie Augenkontakt, Körpersprache etc. im Fokus. Diese Ergebnisse verstärken die Kritik an der CBE, dass gerade die Inhalt- und Kontextspezifität in verschiedenen Situationen zu unterschiedlicher Performanz führen kann (Lurie 2012; Ende & Davidoff 1992). Andererseits sprechen die Ergebnisse des AC auch gegen die Aussage, dass kompetenzorientierte Prüfungen nur für leicht messbare Inhalte wie Wissen oder einfache Fertigkeiten möglich sind (Vgl. Hodges 2010; Brightwell & Grand 2013; Malone & Supri 2012).

Unterscheidung zwischen kompetenten und inkompetenten Kandidatinnen und Kandidaten hinsichtlich einzelner Kompetenzen

Die vorgelegten Arbeiten zur fallbasierten, elektronischen Prüfung (Rotthoff et al. 2006a und 2006b) und zum Kommunikations-OSCE (Mortsiefer, Immecke & Rotthoff et. al 2014) zeigen, dass in der *Simulation* die Messung einer erreichten Kompetenzstufe u.a. über eine Standardsetzung (durch Experten) valide und prinzipiell auch reliabel möglich ist. Vorhersagen über die spätere Performanz lassen die Ergebnisse der elektronischen Key-Feature Prüfung und des Kommunikations-OSCE bisher jedoch nicht zu, da weitere Langzeitbeobachtungen dieser Studierenden fehlen. Folglich bedeutet ein gutes Abschneiden im Kommunikations-OSCE auch nicht zwangsläufig, dass ein Transfer der Kompetenz aus der Simulation auf eine Gesprächssituation mit einem realen Patienten im klinischen Kontext gelingt. Wir nehmen lediglich an, dass die Gesamtprüfung Kommunikations-OSCE ein stabiles Konstrukt „Ärztliche Gesprächsführung“ misst, obwohl der Kommunikations-OSCE nur eine Auswahl an spezifischen Kommunikationssituationen beinhaltet. In der klinischen Praxis ist die Varianz von Fall zu Fall von Bedeutung.

Demgegenüber ermöglichte der ebenfalls in *simulierter* Umgebung durchgeführte AC mit einer integrativen Messung von Kompetenzen zwar Aussagen zur späteren Performanz, jedoch erfüllte dieser die klassischen Anforderungen der Testgüte für eine summative Prüfung - wie für AC auch in der Literatur beschrieben - in Bezug auf die Objektivität und Reliabilität dagegen nicht (Lienert & Raatz 1998; American Educational Research Association 1999 u. 2014). Als Limitation für diese Aussage kann die kleine Studierenden-Stichprobe im präsentierten AC aufgeführt werden.

Eine umfassende, separate Prüfung einzelner Kompetenzen ist nicht möglich

Die drei vorgelegten Prüfungen (elektronische Key-Feature Prüfung, Kommunikations-OSCE, Assessment Center) lassen im Kontext der Literatur den Schluss zu, dass eine komplexere Integration von Kompetenzen in einer Prüfung zu Lasten der inhaltlichen Validität geht (vgl. Lienert & Raatz 1998). Die Trennschärfe zwischen den einzelnen Kompetenzen nimmt ab und eine differenzierte, reliable Bewertung einzelner Kompetenzen ist nicht mehr sicher möglich. Auch der Kommunikations-OSCE prüfte zwar wesentlich Kommunikative Kompetenz, jedoch wurden über die Berliner Global Rating Scala (BGR) ebenso affektive Komponenten wie z.B. Empathie erfasst, die auch der sozialen Kompetenz zugeordnet werden können (Mortsiefer, Immecke & Rothhoff et al. 2014).

Die trennscharfe Prüfung einzelner Kompetenzen ist folglich nur für *Teilaspekte* und nur in einer frühen Phase der Kompetenzentwicklung sinnvoll und möglich. In der frühen Kompetenzentwicklung können Wissensgrundlagen, Fertigkeiten und evtl. auch Haltungen einer spezifischen Kompetenz vermittelt und geprüft werden. Die in der Einleitung dieser Arbeit (S. 16f) beschriebenen Lern- und Prüfungsebenen des Cambridge Modells („Wissen“, „Wissen Wie“, „Zeigen Wie“ und „Tun“) (Ramani & Leinster 2008; Rethans et al. 2002) erleichtern die Operationalisierung von Kompetenzen und werden auch im NKLM verwendet. Mit steigender Komplexität von Aufgabenstellungen und steigender Lernebene werden zusätzliche Kompetenzen gefordert, die integrativ angewendet werden müssen. Der NKLM definiert beispielsweise für die Medizinisch-wissenschaftlichen Fertigkeiten insgesamt 48 Lernziele (NKLM 2015 - Kapitel 14a). Diese beinhalten sowohl grundlegende Ziele des wissenschaftlichen Denkens und Handelns, wie z.B. die „unterschiedlichen Vorgehensweisen bei der Hypothesengenerierung erläutern“ oder „unterschiedliche Hypothesenformen erklären“, als auch Handlungswissen, wie z.B. eine „Problemstellung in eine präzise, überprüfbare wissenschaftliche Fragestellung übersetzen können“ sowie auch sehr komplexe Ziele wie „wissenschaftliche Ergebnisse für ein Fachpublikum nach den Regeln wissenschaftlicher Publikationen aufbereiten und mündlich präsentieren“. Das grundlegende Wissen kann theoretisch vermittelt, gelernt und trennscharf über eine schriftliche Prüfung geprüft werden. Wenn zum Erreichen der Lernziele jedoch auch praktische Übungen in Gruppenarbeit durchgeführt werden und die Arbeitsergebnisse anschließend präsentiert werden sollen, sind von den Studierenden auch zusätzliche Kompetenzen wie die Kommunikative Kompetenz oder Lehrkompetenz gefordert. Bei komplexen und realitätsnahen Aufgabenstellungen, wie der eigenständigen Aufbereitung wissenschaftlicher Ergebnisse und deren Präsentation, sind neben einem fundierten Wissen auch Kompetenzen wie die Selbstkompetenz, Kommunikative- und Lehrkompetenz gefordert.

Aus den bisherigen Analysen lässt sich schlussfolgern, dass die umfassende, separate Prüfung einzelner Kompetenzen nicht möglich ist. Die Einführung von Leistungsnachweisen für einzelne Kompetenzen ist demnach nicht sinnvoll und abzulehnen!

Es besteht somit ein Spannungsfeld, Kompetenzen einerseits objektiv, trennscharf, valide und reliabel sowie andererseits realitätsnah und integrativ prüfen zu wollen.

Objektivität und Standardisierung versus Authentizität bei kompetenzbasierten Prüfungen

Eine Fokussierung auf die Objektivität und Standardisierung einer Prüfung steigert zwar deren Reliabilität, birgt aber gerade im klinischen Kontext die Gefahr, sich von der Wirklichkeit und einem authentischen Prüfungs-Szenario mit wechselnden Kontexten zu entfernen (Govaerts et al. 2007). Im realen Arbeitsumfeld gestaltet sich eine objektive und reliable Prüfung der Performanz entsprechend schwieriger (vgl. Weyers et al. 2016). Begründen lässt sich dieses damit, dass die Prüfung von Performanz eine Beurteilung und ein Entscheidungsfindungsprozess ist, welcher von den Interaktionen zwischen Personen und dem Kontext beeinflusst werden, in dem das Prüfen stattfindet (Govaerts et al. 2007). Die Inhalt- und Kontextspezifität des klinischen Handelns werden von einigen Autoren auch als Widerspruch zu einem isolierten Training verschiedener Kompetenzen angeführt, denn Studierende können in dem einen Fall eine gute und in einem anderen Fall mit abweichenden Kontext eine schlechte Performanz zeigen (Lurie 2012).

Bei einer Fokussierung auf die Operationalisierung von einzelnen Kompetenzen und der Entwicklung von passenden Prüfungsformaten tritt in der Einschätzung von Hodges und Lingard (2012) die Diskussion darüber, was Kompetenz eigentlich ist, zunehmend in den Hintergrund (Hodges & Lingard 2012, S. 8). Die Vorstellung davon, was Kompetenz bzw. kompetentes ärztliches Handeln ist, ändert sich mit der Zeit (Hodges 2012, S. 15) und ist damit kein stabiles Konstrukt (ebd. S. 20). Dieses gilt für alle Kompetenzen, ob Wissenschaftliche Kompetenz, Präventive, Diagnostische und Therapeutische Kompetenz oder auch Soziale-Ethische Kompetenz. Die eigenen Arbeiten haben gezeigt, dass die Definition einzelner Kompetenzen inhaltlich unscharf ist und auch die Entscheidungen über Kompetenz bzw. Inkompetenz auf Expertenmeinungen basieren. Auch den im NKLM definierten und auf Kompetenzen aufsetzenden Berufsrollen kann eine fehlende umfassende Prüfbarkeit attestiert werden, da bei jeder ärztlichen Tätigkeit die Rollen je nach Anforderung wechseln und ineinandergreifen. Diese operationalisierten Berufsrollen erschweren einerseits deren konkrete praktische Verwendung für die curriculare Entwicklung, andererseits machen sie jedoch die inhaltliche Dimension der verschiedenen Berufsrollen explizit und bieten einen nationalen Orientierungsrahmen für die medizinische Ausbildung in Deutschland.

Es gibt aktuell allenfalls eine Übereinstimmung darüber, wie ein guter Arzt bzw. eine gute Ärztin sein sollte und darauf basierend entscheiden wir, was in ein Curriculum integriert wird und wie unterrichtet werden sollte (Schuwirth & van der Vleuten 2012, S. 123). Dabei sind wir geprägt von einer Umgebung, in der Ideen kommen und gehen, beeinflusst von politischen, sozialen und ökonomischen Vorstellungen sowie Bedingungen der Zeit (Hodges 2012, S. 20). So verständlich und nachvollziehbar der Ansatz einer kompetenzorientierten Ausbildung bezogen auf die späteren beruflichen Anforderungen ist, so wird er bis heute eben doch auch wesentlich von politischen, sozialen und regulatorischen Motiven beeinflusst und weniger auf wissenschaftlicher Evidenz basierend angesehen (Brightwell & Grand 2013; Hawkins et al. 2015; Harden 2007). Bisweilen wird eine objektive Kompetenz- oder Performanzmessung sogar als gar nicht möglich angesehen; denn was heute als Attribut eines guten Arztes oder einer guten Ärztin angesehen wird, war es vor 50 Jahren noch nicht (Schuwirth & van der Vleuten 2012, S. 123). Wenn eine objektive und vor allen Dingen umfassende Kompetenzmessung nicht möglich ist, wie können Kompetenzen dann geprüft werden?

4.2 Lösungsansätze für die integrative Prüfung von Kompetenzen

Ein Lösungsansatz für die integrative Bewertung von Kompetenzen kann häufiges prozessorientiertes (= formatives) und *subjektives* Assessment (Feedback) sein, um die Entwicklung der Studierenden zu erfassen, diagnostische Entscheidungen über das erreichte Kompetenzniveau zu treffen und die Studierenden in ihrem Lernprozess zu begleiten (Holmboe et al. 2011). Formative Prüfungsansätze mit regelmäßigem Feedback im Sinne eines „*Assessment for Learning*“ anstelle eines ausschließlichen „*Assessment of Learning*“ sind wichtige Steuerungsinstrumente für die persönliche und berufliche Weiterentwicklung. Im Düsseldorfer Curriculum Medizin wird beispielsweise allen Studierenden, die den Kommunikations-OSCE nicht bestanden haben, ein persönliches Coaching unter Verwendung der individuellen Videoaufzeichnung der Prüfung angeboten.

Feedback ist besonders vor dem Hintergrund bedeutsam, als die klinische Expertise und das Beherrschen von klinischen Fertigkeiten nicht linear als eine Funktion der zunehmenden beruflichen Erfahrung wächst. Für den Bereich der ärztlichen Gesprächsführung konnte beispielsweise gezeigt werden, dass eine zunehmende Erfahrung mit Gesprächssituationen (z.B. das Überbringen schlechter Nachrichten) nicht automatisch die kommunikative Kompetenz verbessert, sondern diese in vielen Fällen auf einem Niveau stehen bleibt oder sich sogar trotz zunehmender Praxis verschlechtert (Yudokowsky, Downing & Ommer 2006). Trotz dieser

Erkenntnis werden Kommunikationstrainings auch in der ärztlichen Fortbildung kaum angeboten und nachgefragt (Rotthoff et al. 2011a).

Validität durch häufige subjektive Beobachtungen

Eine häufige und vor allen Dingen *subjektive* Bewertung und Erfassung von Kompetenz erscheint zunächst befremdlich, denn das typische Prüfungsprogramm besteht aus Einzelteilen, deren Leistungspunkte zu einem Ganzen zusammengeführt werden, um damit eine Entscheidung über Kompetenzen zu treffen. Diesem liegt auch der allgemeine Glaube zugrunde, dass Zahlen objektiver sind als individuelle Berichte (Tekian et al. 2015). Es besteht heute jedoch ein zunehmendes Übereinkommen darüber, dass gerade die persönliche Beurteilung durch einen Prüfer für die Beurteilung von Kompetenzen eine wichtige Rolle spielt, obwohl die menschliche Beurteilung als alleiniges Messinstrument in Bezug auf die Validität natürlich auch große Limitationen hat (Schuwirth & van der Vleuten 2012, S. 117). Dennoch: so wie eine Diagnose oder Behandlung nicht alleine und ausschließlich auf Basis von Laborwerten gestellt werden kann, sondern auch das eigene Urteilsvermögen einbezieht, so sollte auch das kompetenzbasierte Lehren und Prüfen das eigene – und hier nicht ausschließlich fachbezogene - Urteilsvermögen der Prüfer mit einbeziehen (Tekian et al. 2015). Ausgehend von dem von Kane (Kane 2006) vorgeschlagenen theoretischen Rahmenmodell für die Validität ist hierfür eine große Anzahl von Beobachtungen mit einer hohen Varianz der Beobachtungsbedingungen erforderlich.

Kompetenzorientiertes Prüfen muss künftig stärker von den Situationen des ärztlichen Handelns und weniger von der Kompetenz ausgehend gedacht werden.

Da Kompetenzen - wie zuvor postuliert - einzeln nicht umfassend geprüft werden können, ist eine integrative Prüfung von Kompetenzen notwendig. Gemäß der Definition von Kompetenz geht es um die erfolgreiche, verantwortungsvolle und *variable* Nutzung verschiedener Kompetenzen, um Probleme zu lösen (Vgl. Weinert 2002). Dabei ist die wissenschaftliche Kompetenz ebenso relevant wie die Präventive, Diagnostische und Therapeutische Kompetenz oder die Kommunikative, Wirtschaftliche, Sozial-ethische, Selbst- oder Lehrkompetenz. Das integrative Training von Kompetenzen setzt damit eine stärkere und frühzeitigere Integration der Medizinstudierenden in die klinischen und wissenschaftlichen Arbeitsprozesse voraus.

Das Düsseldorfer Curriculum Medizin trägt dem Ansatz von häufigem Feedback bereits Rechnung (Rotthoff et al. 2015). Neben dem Unterricht am Krankenbett müssen alle Studierenden im Verlauf des Studiums mindestens $n = 60$ reale Patienten kontaktieren und ausgehend von deren Behandlungsanlass eine Anamnese und körperliche Untersuchung durchführen. Sie erarbeiten sich dann, je nach Studienjahr mit unterschiedlichem Schwerpunkt, die pathophysiologischen und

pathobiochemischen Zusammenhänge bzw. das diagnostische und therapeutische individuelle Vorgehen und reflektieren das eigene Handeln (Rotthoff et al. 2015).

Ihre Ergebnisse stellen sie einem Arzt oder einer Ärztin vor, die ihnen anschließend Feedback geben. Zusätzlich stellen sie jeweils $n = 7$ Patientenfälle in Tutorien vor, die eine tiefgehende inhaltliche Auseinandersetzung erfordern und die Entwicklung verschiedener Kompetenzen einschließlich der Wissenschaftlichen Kompetenz (z.B. Kritisches Hinterfragen, Studieninterpretation oder Anwendung von evidenzbasierter Medizin), Sozial-ethischen Kompetenz, Lehrkompetenz, Selbstkompetenz oder auch der Wirtschaftlichen Kompetenz über ärztliches Feedback vertiefen sollen und können. Darüber hinaus sind mindestens $n = 8$ klinische praktische Prüfungen als Mini Clinical Examination (Mini-CEX) abzulegen, bei denen klinisch praktische Fertigkeiten vorwiegend am Patienten geprüft werden und die verschiedene Kompetenzen integrativ erfordern.

So wird bei dem Mini-CEX „Neuropathie-Screening“ beispielsweise nicht nur die manuelle Fertigkeit bei der Untersuchungstechnik, sondern auch die kritische Interpretation der erhobenen Untersuchungsbefunde als diagnostische Kompetenz und die Kommunikation mit dem Patienten beobachtet und bewertet. Die Kliniken und Institute haben bisher insgesamt $n = 86$ solcher Mini-CEX Prüfungssituationen entwickelt (Stand 2/2017).

Diese beschriebenen Formate sehen Beobachtungen mit einer hohen Varianz der Beobachtungsbedingungen und Prüfer vor und steigern damit die Validität (Rotthoff et al. 2015; Kane 2006). Die an Tätigkeiten des Arztberufes ausgerichteten Prüfungsszenarien stellen einen Transfer von Kompetenzen in die klinische oder wissenschaftliche Praxis dar und können als Anvertraubare Professionelle Tätigkeiten (APT) bezeichnet werden. Kompetenzen gehen von der einzelnen Person aus, APT von konkreten Anforderungen im Arbeitsumfeld, für deren Umsetzung verschiedene Kompetenzen erforderlich sind (Ten Cate, 2013).

Das Prinzip des Vertrauens und Anvertrauens

Dieses international zunehmend eingesetzte Prinzip der Anvertraubaren Professionellen Tätigkeiten (APT) (im Englischen: „Entrustable Professional Activities“, EPAs) folgt einer Praxis, wie sie in den meisten Arbeitsbereichen weltweit angewandt wird. Am Ende steht eine Entscheidung darüber, wer welche Tätigkeit, unter welchem Anleitungs- und Aufsichtsgrad durchführen darf (= Anvertrauens-Entscheidungen) (AG PJ Logbuch 2.0. des Medizinischen Fakultätentages). Der Akt des Anvertrauens basiert auf einem expliziten und transparenten Entscheid, der überwiegend auf kontinuierlicher Supervision durch verschiedene Personen und punktuell und ergänzend auf explizit dokumentierten Beobachtungen wie z.B. der Bewertung von Arbeitsergebnissen oder Praxisbeobachtungen (z.B. Patientenaufnahme, Wundbehandlung) beruht

(ebd.). Dem Akt des Anvertrauens liegt damit immer auch ein gewisser Grad an Unsicherheit inne und bleibt ein *subjektiver* Prozess zwischen Ausbilder und Auszubildenden, was bereits durch das Wort „Anvertrauen“ impliziert wird (ebd.). Vertrauen und Anvertrauen bildet eine der Grundsäulen im individuellen, institutionellen und gesellschaftlichen Umgang mit Gesundheit und Krankheit. Die Gesellschaft vertraut der Ärzteschaft, den Gesundheitsberufen und ihren Institutionen. Patienten vertrauen sich individuellen Ärzten an. Ärzte wiederum vertrauen schrittweise ihren Aus- und Fortzubildenden die Ausführung von ärztlichen Tätigkeiten an (ebd.). Es braucht vielleicht künftig mehr Bewusstsein und auch Mut, solche individuellen Entscheidungen bei Medizinstudierenden im Verlauf des Medizinstudiums zu treffen. Im ärztlichen Alltag wird dieses Vorgehen bereits heute regelhaft praktiziert: Oberärzte und Oberärztinnen vertrauen Assistenzärztinnen und -ärzten Handlungen an. Diese wiederum vertrauen Famulantinnen und Famulanten oder Studierenden im Praktischen Jahr Handlungen an. Limitierend für solch ein Vorgehen sind neben rechtlichen Aspekten auch die großen Studierendenzahlen anzuführen, die eine individuelle und kontinuierliche Beobachtung erschweren. Hier können künftig eine Stärkung der ambulanten Medizin wie im Masterplan 2020 für die Weiterentwicklung des Medizinstudiums angedacht, oder auch eine Neuausrichtung der insgesamt viermonatigen Famulatur neue Möglichkeiten aufzeigen.

Entsprechend der Definition von Frank et al. (2010) und der Analyse von Brightwell & Grand (2013) wird kompetenzbasierte Ausbildung auch am Fortschritt des individuellen Kompetenzerwerbs festgemacht und weniger an der rein zeitlichen Dauer einzelner Trainingseinheiten. Für das Fortschreiten im Medizinstudium an einer Universität ist diese Vorgehen prinzipiell denkbar, wenngleich auch prüfungslogistisch herausfordernd. Eine Verkürzung des Studiums insgesamt ist aufgrund der gesetzlichen Rahmenbedingungen bisher nicht möglich (ÄAppO, Richtlinie 2005/36/EG über die Anerkennung von Berufsqualifikationen).

Kompetenzorientiertes Prüfen erfordert Austausch und Diskussion zwischen den Prüfern

Neben den gesellschaftlichen Erwartungen an die Ärztinnen und Ärzte sind auch die Erwartungshaltung und Einstellung der Menschen einer Fakultät gegenüber einer kompetenzorientierten Ausbildung und die damit verbundene tatsächliche Umsetzung solcher Prüfungen bedeutsam; insbesondere dann, wenn die Bedeutung des individuellen Prüfers und der Prüferin an Bedeutung gewinnt.

Für eine gelingende Implementierung kompetenzorientierter Prüfungen kommt der Fakultätsentwicklung daher eine besondere Bedeutung zu. Trainings in der Fakultät sollten sich nicht auf einzelne Workshops für einzelne Assessment Instrumente beschränken, bei dem es

beispielsweise darum geht, wie eine Checkliste für einen Mini-Clinical Examination auszufüllen ist. Vielmehr sollen sich die Prüfer über ihre Bewertungen und die Prüfungssituationen austauschen und diese diskutieren (Schuwirth & van der Vleuten 2012, S. 126-127). Realitätsnahe Prüfungssituationen können dabei im klinischen wie im wissenschaftlichen Umfeld geschaffen werden. Dieses setzt eine Fakultätsentwicklung, entsprechende Kapazitäten und eine Lehr-Lernkultur voraus, die solchen Diskussionen Raum für ernsthaften Austausch ermöglicht. Wenn Feedback nicht konstruktiv gegeben wird, laufen die Ziele für den Kompetenzerwerb ins Leere. Für ein funktionierendes Feedback ist sowohl ein vertrauenswürdiges Verhältnis zwischen Feedbackgeber und -nehmer (Crommelinck & Anseel 2013) also auch die Lernumgebung (Lehr-/Lernkultur) entscheidend, in der Feedback stattfindet (Kinicki et al. 2004). Um eine Bewertung des lokalen Lehr- und Lernklimas vornehmen zu können, muss dieses zunächst messbar gemacht werden.

4.3 Bedeutung und Messung des Lehr-/Lernklimas

In den vorliegenden Untersuchungen (Rotthoff et al. 2011; Rotthoff et al. 2012, Ostapczuk et al. 2012) wurden Messinstrumente zur Erfassung des Lehr-/Lernklimas für den deutschsprachigen Raum validiert. Die Fragebögen beziehen erstmalig sowohl die Studierenden als auch die Lehrenden als Zielgruppe mit ein. Gerade die Perspektive der Lehrenden auf das Lehr-/Lernklima wird im Verlauf curricularer Veränderungen als besonders wichtig angesehen (Till, 2005).

Der Kenntnis nach zum ersten Mal wurde aus dem Leitbild einer medizinischen Fakultät ein Fragebogen entwickelt, um den Umsetzungsgrad aus Sicht der Studierenden und Lehrenden zu untersuchen (Rotthoff et al 2012). Der Leitbild-Fragebogen zeigte sehr gute Item- und Testkennwerte. Die Ergebnisse der Testanalyse bescheinigen beiden Fragebögen sowohl für die Human- als auch für die Zahnmedizin eine vergleichbare und gute Reliabilität. Die sehr gute interne Konsistenz auf Gesamtwert-Ebene und die hohe positive Korrelation zwischen Leitbild-Fragebogen und DREEM belegen damit die Eignung des Leitbild-Fragebogens zur Untersuchung des Lehr- und Lernklimas (Rotthoff et al 2012).

Die Lehrenden in der Humanmedizin schätzen in der vorliegenden Arbeit die Atmosphäre und das gesamte Klima mittels DREEM signifikant positiver ein als die Studierenden (Rotthoff et al. 2011). Diese unterschiedlichen Wahrnehmungen können über eine bisher eher geringe Interaktion zwischen Studierenden und Lehrenden erklärt werden, wodurch die Lehrenden nur wenig Einblick in die Lebenswirklichkeit der Studierenden bekommen. Als mitursächlich können hier die großen Studierendenzahlen und vorwiegend kurze Kontaktzeiten zwischen Studierenden und Lehrenden angeführt werden. Dieses legen auch die deutlich diskrepanten Einschätzungen von Studierenden

und Lehrenden für die Aspekte „Vorbildfunktion“ und „Ausmaß einer lebendigen Interaktion“ nahe. Die bessere Bewertung des Lehr-/Lernklimas bei den Studierenden der Zahnmedizin kann demzufolge auch auf der kleineren Studierendenzahl und dem intensiveren Kontakt mit den Lehrenden basieren.

Ein weiterer Grund kann in einer gefühlt stärkeren Verantwortung für die Lehre bei Lehrenden im Vergleich zu den Studierenden liegen, welche definitionsgemäß eher als die passiveren „Konsumenten“ der Lehre wahrgenommen werden (Vgl. Ostapczuk 2014, S. 116). Wie aus der sozialpsychologischen Forschung zum *fundamentalen Attributionsfehler* gut bekannt ist, „(...) könnten die Studierenden die Verantwortung für die vermeintlich schlechte Lehre externalisieren, d.h. den Lehrenden zuschreiben, und diese daher negativer beurteilen“ (ebd.). In der Wahrnehmung der Studierenden geben die Lehrenden im Mittel kein gutes (stimulierendes) Feedback (DREEM und Leitbild-Fragebogen). Gerade Feedback setzt - wie zuvor vorbeschrieben - eine Kultur des wechselseitigen Vertrauens und der Anerkennung voraus (Crommelinck & Anseel 2013). Neben dem Ausmaß der bestehenden Interaktion zwischen beiden Gruppen existieren weitere Faktoren, die einen Einfluss auf das Lehr-/Lernklima haben. Mit dem Fortschreiten des Studiums wird das wahrgenommene Lehr-/Lernklima im Querschnitt schlechter. Studierende des klinischen Studienabschnitts beurteilten vor allem die Lehre schlechter, die Atmosphäre jedoch besser als ihre Kommilitonen aus dem vorklinischen Abschnitt. Die Unterschiede blieben auch signifikant, wenn das Alter aus der Korrelation zwischen der Semesterzahl und der Wahrnehmung des Klimas herausgerechnet wurde, was hier erstmalig untersucht wurde. Die hohe Motivation zu Beginn des Medizinstudiums scheint bei vielen Studierenden im Laufe des Studiums abzunehmen (vgl. Miles & Leinster 2007), und zwar unabhängig von konkreten Erfahrungen. Ähnlich wie bei Miles und Leinster (2009) nahmen die Studierenden sich selbst sozial positiver wahr als sie von den Lehrenden wahrgenommen wurden, wohingegen sich die Lehrenden selbst und ihre Lehre positiver beurteilten als dies die Studierenden taten. Ursächlich für die bessere Einschätzung der jeweils eigenen Gruppe könnte ein Innergruppenbias gewesen sein, da Menschen gemäß der Theorie der sozialen Identität häufiger die Gruppe bevorzugen, der sie sich zugehörig fühlen. Auch überschätzen Menschen gerne ihre eigenen Fähigkeiten (Lewicki 1983; Svenson 1981; Ostapczuk 2014, S. 119). In Übereinstimmung mit anderen Arbeiten (Bassaw et al. 2003; Bouhaimed et al. 2009; de Oliveira Filho, Vieira & Schonhorst 2005; Roff et al. 1997) bestätigt sich bei den Studierenden der Humanmedizin, dass das Lehr-/Lernklima sowohl über den DREEM als auch über den Leitbildfragebogen von Frauen im Vergleich zu Männern als besser wahrgenommen wird. Gemäß Roff (2005) hängen Geschlechtsunterschiede bei der Wahrnehmung des Lehr- und Lernklimas von einer Vielzahl auch kultureller Faktoren ab und fallen insgesamt geringer aus, wenn die Gesamtzufriedenheit hoch ist (Roff 2005). Dieses bestätigt sich anhand der vorliegenden

Ergebnisse, da bei den Studierenden der Zahnmedizin die Gesamtzufriedenheit über das Lehr-/Lernklima größer war als bei den Studierenden der Humanmedizin und signifikante Unterschiede zwischen Frauen und Männern nicht nachweisbar waren.

Nutzung von Messergebnissen zum Lehr-/Lernklima für die Fakultätsentwicklung

Die im Leitbild Lehre der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf (Leitbild Lehre 2009) definierten Ziele waren zum Zeitpunkt der Befragung in der Einschätzung von Studierenden und Lehrenden noch ein Stück weit von ihrer Umsetzung entfernt. Da ein Leitbild eher langfristige Ziele definiert und das Düsseldorfer Leitbild Lehre zum Zeitpunkt der Befragung erst ein Jahr veröffentlicht war, überrascht dieses Ergebnis nicht. Es gibt viele empirische Hinweise darauf, dass ein Leitbild auch nicht automatisch akzeptiert wird, sondern ganz im Gegenteil gezielte Maßnahmen getroffen werden müssen, um die erwünschte Umsetzung zu fördern: Voraussetzung sind seine Kenntnis (Davis, Ruhe, Lee & Rajadhyasksha, 2006) und ein Vorleben seiner Umsetzung durch die Verantwortlichen (Rough, 2009) (Vgl. auch Ostapczuk 2014, S. 8) Insbesondere eine weitere Verbreitung und Diskussion über die Leitbildinhalte erscheinen dabei wichtig. Durch die Transformation des Leitbildes in einen Fragebogen und dessen Einsatz bei Studierenden und Lehrenden konnten konkrete Bereiche für die weitere Fakultätsentwicklung aufgezeigt werden. Für den Standort Düsseldorf waren es die von den Lehrenden wahrgenommene geringe Wertschätzung ihrer Lehrtätigkeit durch die Fakultät sowie die von Studierenden und Lehrenden als diskrepant wahrgenommene Beziehung zwischen beiden Gruppen. Zur Steigerung des Dialogs zwischen Studierenden und Lehrenden trägt heute z.B. das Mentoring Projekt A² der Fakultät bei. Die Vergabe von Lehrpreisen an besonders engagierte Lehrende oder auch die individuelle Rückmeldung besonders positiver studentischer Evaluation an einzelne Lehrpersonen durch das Studiendekanat sind Maßnahmen für eine gesteigerte Übermittlung und Wahrnehmung von Wertschätzung.

Entwicklung einer Standortkultur

Zu berücksichtigen ist auch, dass die Lehrenden als Mitglieder der Fakultät vielfach auch gleichzeitig Mitarbeiterinnen und Mitarbeiter eines Universitätsklinikums sind. Sie vereinen damit die Tätigkeiten von Krankenversorgung, Forschung und Lehre in sich als Person. Die formal vorhandene und in der Wahrnehmung auch gelebte Trennung von „Lehre = Fakultät“ und „Krankenversorgung = Klinikum“ kann zu intrapersonalen Konflikten in Bezug auf die unterschiedlichen Anforderungen führen. Die Zufriedenheit mit dem Arbeitsplatz „Klinik“ bei einer Person hat sicherlich Einfluss auf die Bewertung des Curriculums und die Wahrnehmung der Studierenden. Wenn also Forschung, Lehre und Krankenversorgung in einer Person integrativ zu

leisten sind, sollte die Unternehmenskommunikation von Klinikum und Medizinischer Fakultät (Dekanat, Studiendekanat) soweit wie möglich vernetzt werden, um gemeinsam an einer einzigen Unternehmenskultur zu arbeiten. Die Lehr-/Lernkultur und die Unternehmenskultur eines Universitätsklinikums/einer Fakultät sind zwei Seiten derselben Medaille, die sich wechselseitig bedingen. Mitarbeiterinnen und Mitarbeiter leben die Einheit und Konflikte von Lehre und Krankenversorgung, werden jedoch von zwei mitunter konkurrierenden Systemen angesprochen. Eine in Teilen besser abgestimmte und wertschätzende Kommunikationsstrategie von Fakultät und Klinikum könnte zur stärkeren Entwicklung einer Standortkultur und Standortprofilierung beitragen, auch wenn äußere - z.B. wirtschaftliche - Zwänge als Störgrößen bestehen bleiben. Die Zufriedenheit und Identifikation mit dem Standort haben Einfluss auf die Wahrnehmung des *Arbeits-, Lehr und Lernklimas* und im positiven Sinne können sie zu verbesserten Ergebnissen führen. Psychometrisch sind in zukünftigen deutschsprachigen, aber auch internationalen Untersuchungen eine weitergehende konvergente, aber auch diskriminante Validierung der Fragebögen sowie die Untersuchung der faktoriellen Struktur des DREEM und seiner Retest-Reliabilität unter Verwendung eines adäquaten Retest-Intervalls von Interesse.

5 Zusammenfassung

Die Vermittlung und der Erwerb medizinischen Wissens sind wesentlich für den Arztberuf! 'Verbunden mit der Kompetenzorientierung ist eine Akzentuierung von Fertigkeiten und Haltungen in Kombination zum Wissen – also eine Stärkung des berufsspezifischen, praxisorientierten Spektrums des Medizinstudiums' (WR 2014).

Die *theoretischen Wissensgrundlagen und Basisfertigkeiten* einzelner Kompetenzen können (summativ) valide, objektiv und reliabel geprüft werden. Hier haben Prüfungsformate in der Simulation wie z.B. die vorgelegte elektronische Key-Feature Prüfung oder der Kommunikations-OSCE ihren Stellenwert. Das *integrative* Prüfen verschiedener Kompetenzen an realitätsnahen Aufgaben hat über den AC jedoch gezeigt, dass die testtheoretischen Kriterien von Validität und Reliabilität für die Einzel-Kompetenzen nicht mehr erfüllt werden, obwohl diese Prüfung insgesamt eine gute Vorhersage über die spätere Performanz ermöglichte.

Die vorgelegten Arbeiten lassen den Schluss zu, dass je komplexer die Prüfungssituation ist und je mehr Kompetenzen integriert geprüft werden sollen, es umso schwieriger wird, eine objektive und reliable kompetenzspezifische Prüfung abzubilden. Die Integration verschiedener Kompetenzen in einer Prüfung geht demnach zu Lasten der inhaltlichen Validität für die einzelne Kompetenz, deren Trennschärfe schwächer wird.

Auf Basis dieser Überlegungen und testtheoretischer Erkenntnisse muss postuliert werden, dass eine *umfassende*, isolierte Prüfung einzelner Kompetenzen (und auch Berufsrollen) nicht möglich ist! Eine alleinige Ausrichtung von Prüfungen an den operationalisierten Zielen einzelner Kompetenzen ist folglich nicht sinnvoll und Leistungsnachweise für einzelne Kompetenzen sind abzulehnen.

Es ist daher ein Perspektivwechsel auf das kompetenzorientierte Prüfen notwendig! Kompetenzorientiertes Prüfen muss künftig weniger von der einzelnen Kompetenz (oder Rolle) ausgehend gedacht werden, sondern stattdessen von ärztlichen Tätigkeiten aus dem klinischen und wissenschaftlichen Arbeitsumfeld, die verschiedene Kompetenzen integrativ und in unterschiedlicher Gewichtung erfordern.

Für das Medizinstudium bedeutet das, Situationen für die Studierenden im Curriculum zu schaffen, in denen häufiges prozessorientiertes (= formatives) und *subjektives* Assessment (Feedback) in realitätsnahen oder realen Situationen erfolgt. Für eine hinreichende Validität solch eines formativen und subjektiven Assessments ist auf eine hohe Varianz der Beobachtungsbedingungen und Prüfer zu achten. Das Düsseldorfer Curriculum Medizin zeigt mit seinen verschiedenen Lehr- und Lehrformaten innerhalb von 40 Praxiswochen entsprechende Möglichkeiten auf, den Wissenserwerb in die Praxis zu integrieren und dabei Kompetenzen formativ und integrativ zu prüfen und rückzumelden.

Die Entwicklung von „Entrustable Professional Activities“ (EPA) - deutsch: „Anvertraubare Professionellen Tätigkeiten“ (APT) - zeigen einen weiteren Weg auf. Solche APT können in unterschiedlicher Granularität vorgehalten werden. In früheren Phasen des Studiums können einfache diagnostische und therapeutische Tätigkeiten unter Supervision bereits zusammen mit anderen Kompetenzen verknüpft gelehrt und geprüft werden. Im Praktischen Jahr können die Tätigkeiten dann umfassender sein und z.B. die Betreuung eines Patienten mit chronischer Erkrankung oder das komplette Entlassungsmanagement eines Patienten beinhalten. Integratives Prüfen von Kompetenzen ist – wie mit dem AC gezeigt - je nach Aufgabenstellung von Beginn des Studiums an möglich, da Studierende bereits Kompetenzen in unterschiedlicher Ausprägung mitbringen. Die Notwendigkeit der Vermittlung wissenstheoretischer Grundlagen, Fertigkeiten und Haltungen für einzelne Kompetenzen ist demnach abhängig von der bereits bestehenden individuellen Kompetenz der Studierenden und der konkreten Aufgabenstellung.

Fakultätsinterne Prüfungen sollten die Studierenden auch auf das 3. Staatsexamen vorbereiten. Dieses bekommt gerade vor dem Hintergrund der Planungen zur Weiterentwicklung des Medizinstudiums mit dem Entwurf des Masterplan 2020 neue Aktualität. Hiernach sind klinisch

praktische Prüfungsszenarien in einer OSCE Prüfung zu Beginn und am Ende des praktischen Jahres (3. Staatsexamen) vorgesehen. Diese Prüfungsszenarien sollen sich an klinischen Situationen orientieren (Quelle IMPP, persönliche Mitteilung) und werden darüber verschiedene Kompetenzen integrieren. Das solche Prüfungen – sofern qualitativ hochwertig erstellt - in Kombination mit Wissensprüfungen die spätere Performanz im klinischen Umfeld vorhersagen können, konnte in den Vereinigten Staaten an der USMLE gezeigt werden.

Die Ausrichtung der Ausbildung an professionellen Tätigkeiten in unterschiedlicher Komplexität ist ein gangbarer Weg, ohne dass bisher ausreichende Evidenz für die Machbarkeit und Kriteriumsvalidität vorliegt. Die nominelle Einführung eines solchen Konzeptes bedeutet auch noch nicht automatisch dessen gelingende Umsetzung an einer Fakultät. Da die einzelnen Prüferinnen und Prüfer an Bedeutung gewinnen, Objektivität (und Reliabilität) abnehmen, ist die Akzeptanz kompetenzbasierter Lehre und Prüfungen bei den Beteiligten mitentscheidend für das Gelingen. Für ein funktionierendes Feedback ist neben einem vertrauenswürdigen Verhältnis zwischen Feedbackgeber und –nehmer auch das Lehr- und Lernklima entscheidend, in dem Feedback stattfindet. Für eine Bewertung des lokalen Lehr- und Lernklimas, muss dieses zunächst messbar gemacht werden. Mit überzeugenden Gütekriterien auf Item- wie Testebene steht das DREEM und ein Leitbildfragebogen nun in deutscher Sprache sowohl für Studierende und international erstmalig auch für Lehrende zur Erfassung des Lehr- und Lernklimas für die Fakultätsentwicklung zur Verfügung. Damit solch ein Vorgehen gelingt, sind Maßnahmen der Fakultätsentwicklung erforderlich, die wesentlich auch die Lehr- und Lernkultur und eine Diskussion über die Ausbildungsziele beinhalten. Neben Initiativen, die den Austausch von Studierenden und Lehrenden fördern oder gute Lehre auszeichnen, kann eine intensivere Vernetzung der Kommunikationsstrategie von Fakultät und Universitätsklinikum zur stärkeren Entwicklung einer einzigartigen Standortkultur und Profilierung beitragen.

Die in dieser Arbeit gezogenen Schlussfolgerungen basieren auf den Ergebnissen der vorgelegten Arbeiten unter Einbeziehung der Erkenntnisse aus der Literatur. Limitierend ist anzuführen, dass alle Untersuchungen an nur einer Fakultät durchgeführt wurden und die Teilnehmerzahlen wie z.B. beim AC teilweise gering waren. Inwieweit die Ergebnisse also mit anderen Studierenden, an anderen Standorten oder in anderen Ländern vergleichbare Befunde ergeben hätten, kann nicht beantwortet werden. Im Kontext der bisher vorhandenen Evidenz sind ähnliche Befunde jedoch wahrscheinlich.

Die Arbeit zeigt einen Weg auf, wie kompetenzorientiertes Prüfen künftig gedacht und modellhaft am Beispiel des Düsseldorfer Curriculum Medizin auch praktisch durchgeführt werden kann. Dabei können komplexere kognitive Ausbildungsziele ausreichend berücksichtigt werden. Für eine

Fakultät stellt die Kompetenzbasierte Ausbildung eine Herausforderung dar und erfordert das Verständnis und den Willen zur Umsetzung. Welche strukturellen Anpassungen oder Veränderungen der Lehr-/Lernkultur notwendig sind, um an einem Standort kompetenzorientierte Ausbildungskonzepte zielführend umzusetzen oder besser gesagt „zu leben“, bedarf weiterer Forschung.

Die politisch gewollte Akzentuierung von Fertigkeiten und Haltungen in Kombination mit Wissen sowie die Stärkung der berufsspezifischen und praxisorientierten medizinischen Ausbildung erfordern eine weitere Suche nach Evidenz über eine Stärkung der evaluativen Ausbildungsforschung.

6 Literatur

A

AG PJ Logbuch des Medizinischen Fakultätentages – unveröffentlichter Text in eigener Kenntnis als Mitautor des Entwurfs „PJ Logbuch 2.0“ 2016; Medizinischer Fakultätentag der Bundesrepublik Deutschland, 2016.

Albanese MA, Mejjano G, Mullan P, Kokotailo P, Gruppen L. 2008: Defining characteristics of educational competencies. *Med Educ* 2008; 42(3):248-55.

American Educational Research Association (Publisher). *Testing Standards for Educational and Psychological Testing*, Washington, DC, 2014, Kapitel 1, S.23-31; Kapitel 2, S.33-40.

Anthony DE, McNelly TL, Edens PS. A meta-analysis of the criterion-related validity of assessment center dimensions. *Pers Psychol* 2003; 56:125–54.

Approbationsordnung für Ärzte vom 27. Juni 2002 (BGBl. I S. 2405), zuletzt geändert durch Artikel 2 der Verordnung vom 2. August 2013 (BGBl. I S. 3005).

Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010 Jan;44(1):101–8.

Arthur W, Jr, Anthony DE, McNelly TL, Edens PS. A meta-analysis of the criterion-related validity of assessment center dimensions. *Pers Psychol*. 2003; 56:125–54
Ashworth PD, Saxton J. On 'Competence'. *Journal of Further and Higher Education* 1990; 14(2):1-25.

B

Bachmann C, Abramovitch H, Barbu CG, Cavaco AM, Elorza RD, Haak R, et al. A European consensus on learning objectives for a core communication curriculum in health care professions. *Pat Educ Couns* 2013; 93:18–26.

Back AL, Arnold RM, Baile WF, Tulsky JA, Fryer-Edwards K. Approaching difficult communication tasks in oncology. *CA Cancer J Clin* 2005; 55:164–77.

Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES—a six-step protocol for delivering bad news: application to the patient with cancer. *Oncologist* 2000; 5:302–11.

Bassaw B, Roff S, McAleer S, Roopnarinesingh S, de Lisle J, Teelucksingh S, Gopaul S. Students' perspectives on the educational environment, Faculty of Medical Sciences, Trinidad. *Med Teach* 2003; 25:522-6.

Bieri C, Schuler P. Cross-curricular competencies of student teachers: a selection model based on assessment centre admission tests and study selection after the first year of teacher training. *Assess Eval High Educ* 2011; 36:399–415.

Bloom B.S. 1956. Taxonomy of educational objectives. *Handbook 1: Cognitive domain*. New York: David McKay Company.

-
- Bouhaimed M, Thalib L, Doi SAR. Perception of the educational environment by medical students undergoing a curricular transition in Kuwait. *Med Princip Pract* 2009; 18:204-8.
- Bordage G, Page G. An alternate approach to PMPs, the key feature concept. In: Harden R (Hrsg.). *Further Developments in Assessing Clinical Competence*. Montreal: Can-Heal Publications 1987.
- Brannick, M. T., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Med Educ* 2011; 45(12):1181-89.
- Brightwell A. Grand J – competency-based Training: who benefits? *Postgrad Med J* 2013; 89:107-10.
- Brislin RW. Back-translation for cross-cultural research. *J Cross Cult Psychol* 1970; 1:185-216.
- Bürgi H, Bader C, Bloch R, Bosman F, Gasser C, Gerke W, Humair JP, Im Hof V, Kaiser H, Lefebvre D, Schläppi P, Sottas B, Spinass GA, Stuck AE. *Swiss Catalogue of Learning Objectives for Undergraduate Medical Training* – June 2008. Working Group under a Mandate of the Joint Commission of the Swiss Medical Schools.
- C**
- Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multi method matrix. *Psychol bull* 1959; 56(2), 81.
- Carraccio, C., Wolfsthal, SD., Englander, R., Ferentz, K., Martin, C. Shifting paradigms: from Flexner to competencies. *Acad Med* 2002; 77(5):361-67.
- Carraccio, C.L.; Englander, R. From Flexner to competencies: Reflections on a decade and the journey ahead, in: *Acad Med* 2013; 88:1067-73.
- Chen HC, Naquin SS. An integrative model of competency development, training, design, assessment center, and multi-rater assessment. *Adv Dev Hum Resour* 2006; 8:265–82.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (2nd edition). Hove: Lawrence Erlbaum 1988.
- Combes JR Physician competences for a 21st century health care system. *J Grad Med Educ* 2012; 4:401-5.
- Cooke M, Irby DM, O'Brien BC. *Educating Physicians – a call for reform of medical school and residency*, 2010; Vol 16. John Wiley & Sons.
- Crommelinck M, Anseel F. Understanding and encouraging feedback-seeking behaviour: a literature review. *Med Educ* 2013; 47(3):232–41.
- Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Med Educ* 2012; 46(1):28-37.
- Cumming A, Ross M. The Tuning Project for Medicine-learning outcomes for undergraduate medical education in Europe. *Med Teach* 2007; 29(7):636-41.

D

Davis JH, Ruhe JA, Lee M, Rajadhyaksha U. Mission possible: Do school mission statements work? *J Bus Ethics* 2006; 70:99-110.

de Oliveira Filho GR, Vieira JE, Schonhorst L. Psychometric properties of the Dundee Ready Educational Environment Measure (DREEM) applied to medical residents. *Med Teach* 2005; 27:343-7.

Dijkstra IS, Pols J, Remelts P, Rietzschel EF, Schotanus JC, Brand PLP, *Perspect Med Educ* 2015; 4:300–307.

Dornic S, Ekehammar B Extraversion, neuroticism, and noise sensitivity. *Pers Indiv Dif* 1990; 11(9):989-92.

Düsseldorfer Curriculum Medizin: übergeordnete Lernziele des Modellstudiengangs der medizinischen Fakultät, Düsseldorf, www.medizin.hhu.de/medizin/lernziele.

Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, Buffone N, et al. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. *Acad Med* 2004; 79:495–507.

E

Elwyn G, Edwards A, Wensing M, Atwell C, Hood K, Grol R. Fleeting glimpses: measuring shared decision making in primary care using the OPTION instrument. *Qual saf health care* 2003; 12: 93–99.

Ende J, Davidoff F. What is a curriculum? *Ann Intern Med* 1992; 116(12 Pt 2):1055-7.

Ericsson KA. Deliberate practice and maintenance of expert performance in medicine and related domains. *Acad Med* 2004; 79(10 Suppl): S70-81.

Ericsson KA, Chanress N. Expert performance: it's structure and acquisition. *Am Psychol* 1994; 49(8):725-47.

Erridge, A., & Perry, S. The validity and value of national vocational qualifications: The case of purchasing. *The Vocational Aspects of Education*, 1994; 46(2):139-154.

F

Fallowfield, L., & Jenkins, V. Communicating sad, bad, and difficult news in medicine. *The Lancet* 2004; 363(9405): 312-319.

Fischer MR, Kopp V, Holzer M, Ruderich F, Jünger J: A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach* 2005; 27:450-55.

Fisseni HJ. *Lehrbuch der psychologischen Diagnostik*. Mit Hinweisen zur Intervention (3., überarbeitete und erweiterte Auflage). Göttingen; Hogrefe 2004 S.47.

Frank, JR., Snell, LS., Cate, OT., Holmboe, ES., Carraccio, C., Swing, SR., Harden, RM. Competency-based medical education: theory to practice. *Med Teach* 2010; 32(8):638-645.

Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach* 2010; 32:631-637.

Fraser BJ. Differences between students' and instructors' perceptions of actual and preferred classroom environment in higher education. *High Educ Res Dev* 1986; 5:191-9.

Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet* 2010; 376: 1923-58.

G

Gaugler BB, Rosenthal DB, Thornton GC, Benson C. Meta-analysis of assessment center validity. *J Appl Psychol* 1987; 72:493–511.

General Medical Council. Education Committee. Tomorrows' doctors 2009.

Genn JM (a). AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education – a unifying perspective. *Med Teach* 2001; 23:445-54.

Genn JM (b). AMEE Medical Education Guide No. 23 (Part 1): curriculum, environment, climate, quality and change in medical education - a unifying perspective. *Med Teach* 2001; 23:337–44.

Goodstone M, Lopez F. The frame of reference approach as a solution to an assessment center dilemma. *Consult Psychol J Pract Res* 2001; 53:96–107.

Govaerts MJB, Van der Vleuten DPM, Schuwirth LWT, Muijtjens AMM. *Adv Health Sci Educ* 2007; 12:239–60.

H

Haertel GD, Walberg HJ, Haertel EH. Socio-psychological environments and learning: A quantitative synthesis. *Br Educ Res J* 1981; 7:27–36.

Harden RM, AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education, *Med Teach* 1999; 21(1):7-14.

Harden RM. Developments in outcome-based education. *Med Teach* 2002; 24(2):117-20.

Harden RM 2007; Outcome-based education – The ostrich, the peacock and the beaver. *Med Teach* 2007; 29(7):666-71.

-
- Hargie, O., Dickson, D., Boohan, M., & Hughes, K. (1998). A survey of communication skills training in UK schools of medicine: present practices and prospective proposals. *Med Educ* 1998; 32(1): 25-34.
- Harris J, Keller S. Assessment Measures Needed for Competency-Based Higher Education. *Peabody Journal of Education* 1976; 53 (4):241–47.
- Hawkins, RE., Welcher, CM., Holmboe, ES., Kirk, LM., Norcini, JJ., Simons, KB., Skochelak, SE. Implementation of competency-based medical education: are we addressing the concerns and challenges? *Med Educ* 2015; 49(11):1086-1102.
- Heron, S. L., Ander, D. S., Houry, D., Hassani, D. M., & Quest, T. (2010). Standardized Patients to Teaching Medical Students about Intimate Partner Violence. *West J of Emerg Med* 2010;11(5). 500-505.
- Hodges BD, Lingard L. Introduction in: Hodges BD, Lingard L Ed.; *The Question of Competence - Reconsidering Medical Education in the Twenty-First Century*. Cornell University Press, New York 2012, S.1-13.
- Hodges BD. The Shifting Discourses of Competence in: Lingard L. Ed.; *The Question of Competence - Reconsidering Medical Education in the Twenty-First Century*. Cornell University Press, New York 2012, Kapitel 1, S. 15.
- Hodges BD. A Tea-Steeping or i-Doc Model for Medical Education. *Acad Med* 2010; 85(9): S34-44.
- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach* 2010; C32(8):676-82.
- Howard BJ. Calm: listening skills for diffusing anger. In: Jellinek M, editor. Bright futures in practice: mental health—Volume II. Tool kit. Arlington, VA: National Center for Education in Maternal and Child Health; 2002.
- Huddle TS, Heudebert GR. Taking apart the art: the risk of anatomizing clinical competence. *Acad Med* 2007; 82(6):536–41.
- Hyland T. Competence, Knowledge and Education. *Journal of Philosophy of Education* 1993; 27(1), 57-68.
- Hyland T. Teaching, learning and NVQs: challenging behaviourism and competence in adult education theory and practice. Reproduced from 1994 Conference Proceedings, pp. 56-59 SCUTREA 1997.
- I**
- International Task Force on Assessment Center Guidelines Guidelines and ethical considerations for assessment center operations. *Int J Sel Assess.* 2009; 17:243–53.
- Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005; 27(1):10–28.

J

Janson K. Die Sicht der Nachwuchsmediziner auf das Medizinstudium - Ergebnisse einer Absolventenbefragung der Abschlussjahrgänge (2007) -2008. Eine Sonderauswertung des Projekts "Studienbedingungen und Berufserfolg." Internationales Zentrum für Hochschulforschung, Universität Kassel (INCHER-Kassel), 2010; S.9).

Joiner DA. Guidelines and ethical considerations for assessment center operations International Task Force on Assessment Center Guidelines. *Publ Pers Manag* 2000; 29:315–32.

Jones EA, Voorhees R, Paulson K. Defining and Assessing Learning: Exploring Competency-Based Initiative, NCES 2002-159U.S. Department of Education, National Center for Education Statistics. Council of the National Postsecondary Education Cooperative Working Group on Competency-Based Initiatives. Washington, DC: 2002.

Jones PS, Lee JW, Phillips LR, Zhang XE; Jaceldo KB. An Adaption of Brislin's translation model for cross-cultural research. *Nurs Res* 2001; 50:300-304.

K

Kane M. Content-Related Validity Evidence in Test Development in: *Handbook of Test Development*, Downing SM, Haladyna TM Ed. Lawrence Erlbaum Associates, Inc., Publishers Mahwah, New Jersey 2006.

Kinicki AJ, Prussia GE, Wu B, McKee-Ryan FM. A Covariance Structure Analysis of Employees' Response to Performance Feedback. *Journal of Applied Psychology*. 2004; 89(6):1057–69.

Koens F, Ten Cate OJT, Custers EJFM. Context-dependent memory in a meaningful environment for medical education: in the classroom and at the bedside. *Adv Health Sci Educ* 2003; 8:155-65.

Koerfer, A., Köhle, K., Obliers, R., Sonntag, B., Thomas, W., & Albus, C. Training und Prüfung kommunikativer Kompetenz. Aus und Fortbildungskonzepte zur ärztlichen Gesprächsführung. *Gesprächsforsch Z verbal Interaktion* 2008; 9:34-78.

Kopp V, Möltner A, Fischer MR. Key-Feature-Probleme zum Prüfern von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild* 2006; 23(3): Doc50.

Kurtz SM, Silverman JD, Draper J. *Teaching and learning communication skills in medicine*. Oxford: Radcliffe Medical Press; 1998.

Kusurkar RA, Ten Cate TJ, Vos CMP, Westers P, Croiset G. How motivation affects academic performance: a structural equation modelling analysis. *Adv Health Sci Educ* 2013; 18(1):57–6.

L

Lai N, Nalliah S, Jutti RC, Hla Y, Lim VKE. The educational environment and self-perceived clinical competence of senior medical students in a Malaysian medical school. *Educ Health* 2009; 22:148.

Leitbild Lehre der Medizinischen Fakultät der Heinrich-Heine-Universität beschlossen vom
Fachbereichsrat der Medizinischen Fakultät am 25.6.2009;
<http://www.medizin.hhu.de/studium-und-lehre/leitbild-lehre.html> (Zugriff 3.3.2017)

Leong SL, Baldwin CD, Adelman AM. Integrating Web-based Computer Cases into a Required Clerkship: Development and Evaluation. *Acad Med* 2003; 78:295-301.

Leung WC. Competency based medical training: Review. *BMJ* 2002; 325:693-96.

Lewicki P. Self-image bias in person perception. *J Pers Soc Psychol* 1983; 45:384-93.

Liénard, A., Merckaert, I., Libert, Y., Bragard, I., Delvaux, N., Etienne, A. M., ... & Razavi, D. (2010). Is it possible to improve residents breaking bad news skills & quest; A randomized study assessing the efficacy of a communication skills training program. *Br J Canc* 2010; 103(2):171-77.

Lienert GA, Raatz U. *Testaufbau und Testanalyse* (6.Auflage). Weinheim: Beltz 1998 S.9.

Linstone, H. A., & Turoff, M. (Eds.). *The Delphi method: Techniques and applications* (Vol. 29). Reading, MA: Addison-Wesley 1975.

Lurie SJ. History and practice of competency-based assessment. *Med Educ* 2012; 46:49–57.

M

Mager RF. *Lernziele und Unterricht*, 1977, Beltz 1994

Malone K, Supri S. A critical time for medical education: the perils of competence-based reform of the curriculum. *Adv Health Sci Educ* 2012; 17(2):241-46.

McAleer S, Roff S. A practical guide to using the Dundee Ready Education Environment Measure (DREEM). *AMEE medical education guide* 2001; 23:29-33.

McClarty KL, Gaertner MN. Measuring Mastery – Best Practices for assessment in competency-based education, AEI Series on CompetencyBased Higher Education, Center on higher education reform, American Enterprise Institute, April 2015.

Melancon S, Williams M. Competency-based assessment center design: a case study. *Adv Dev Hum Resour* 2006; 8:283–314.

Meng X, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychol Bull* 1992; 111:172–5.

Miles S, Leinster SJ. Medical students' perceptions of their educational environment: Expected versus actual perceptions. *Med Educ* 2007; 41:265-72.

Miles S, Leinster SJ. Comparing staff and student perceptions of the student experience at a new medical school. *Med Teach* 2009; 31:539-46.

Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65(9 Suppl):63–7.

Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild* 2006; 23(3): Doc53.

Morcke AM, Dornan T, Eika B. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv in Health Sci Educ* 2013; 18(4):851-63.

Morrison GR, Ross SM, Kemp JE. Developing evaluation instruments. *Designing effective instruction. 4th edition.* Joh Wiley & Sons Inc; 2004.

Mortsiefer A, Rotthoff T, Schmelzer R, Immecke J, Ortmanns B, in der Schmitt J, Karger A. Implementation of the interdisciplinary curriculum Teaching and Assessing Communicative Competence in the fourth academic year of medical studies (CoMeD). *GMS Z Med Ausbild* 2012; 29(1), Doc06.

Mortsiefer A, Karger A, Rotthoff T, Raski B, Pentzack M. Examiner characteristics and interrater reliability in a communication OSCE. *Pat Educ Couns* 2017; (in press)
<http://dx.doi.org/10.1016/j.pec.2017.01.013>

N

Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM). MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e.V.; Juli 2015.

Neufeld, VR, Maudsley, RF., Pickering, RJ., Turnbull, JM., Weston, WW., Brown, MG., Simpson, JC. Educating future physicians for Ontario. *Acad Med* 1998; 73(11):1133-48.

Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 2004; 38:199–203.

Nikendei C, Jünger J. OSCE - praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. *GMS Z Med Ausbild* 2006; 23(3): Doc47.

Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The Relationship Between Licensing Examination Performance and the Outcomes of Care by International Medical School Graduates. *Acad Med* 2014; 89(8):1157-1162

Norman J. Editorial - outcomes, objectives, and the seductive appeal of simple solutions. *Adv Health Sci Educ* 2006; 11:217-220.

O

Ostapczuk MS. Entwicklung, Adaptation und Validierung deutschsprachiger Inventare zur Erfassung des Lehr- und Lernklimas an der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf, Dissertation, Dezember 2014.

Ostapczuk M, Hugger A, de Bruin J, Ritz-Timme S, Rotthoff T. DREEM on, dentists! Students' perception of the educational environment in a German dental school as measured by the Dundee Ready Education Environment Measure. *Eur J Dent Educ.* 2012; 16(2):67-77

P

Page G, Bordage G: The medical council of Canada's key feature project: a more valid written examination of clinical decision making skills; *Acad Med* 1995, 70:104-110.

Page G, Bordage G, Allen T: Developing key-feature problems and examinations to assess clinical decision making skills *Acad Med* 1995, 70:194-201.

Patricio M, Juliao M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of OSCEs. *Med Teach* 2009; 31:112–124.

Pimparyon P, Caleer SM, Pemba S, Roff S. Educational environment, student approaches to learning and academic achievement in a Thai nursing school, *Med Teach* 2000; 22(4):359-64.

R

Ramani S, Leinster S. AMEE Guide No. 34: Teaching in the clinical environment. *Med Teach* 2008; 30(4): 347-64.

Regehr G, MacRae H, Reznick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; 73:993-97.

Rethans J-J, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, et al. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002; 36(10):901–9.

Rider EA, Hinrichs MM, Lown BA. A model for communication skills assessment across the undergraduate curriculum. *Med Teach* 2006; 28:127–34.

Roff, S., McAleer, S., Harden, R. M., Al-Qahtani, M., Ahmed, A. U., Deza, H., Pimparyon, P. (1997). Development and validation of the Dundee ready education environment measure (DREEM). *Med Teach* 1997; 19(4): 295-99.

Roff S. The Dundee Ready Educational Environment Measure (DREEM) – a generic instrument for measuring students' perceptions of undergraduate health professions curricula. *Med Teach* 2005; 27:322-5.

Rothman AI, Ayoade F. The development of a learning environment: A questionnaire for use in curriculum evaluation. *J Med Educ* 1970; 45:754-9.

Roter DL, Hall JA, Aoki Y. Physician gender effects in medical communication: a meta-analytic review. *JAMA* 2002; 288:756-64.

Rotthoff T, Baehring T, Dicken HD, Fahron U, Fischer M.R., Adler M, Hege I., Scherbaum W.A. Fallbasierte Online-Prüfungen im Medizinstudium – Zielsetzung, Implementierung und praktische Erfahrungen. *GMS Med Inform Biom Epidemiol* 2006a; 2(3):Doc 09

Rotthoff T, Baehring T, Dicken HD, Fahron U, Fischer M.R., Scherbaum W.A. Comparison of Long Menu and Open-Ended Questions in computerized medical assessments – A randomized controlled trial. *BMC-Med Educ* 2006b; 6:50

-
- Rotthoff T, Willers R, Siebler M, Lindner S, Scherbaum W, Soboll S. Partielle OSCE Prüfung zur Einsparung von Ressourcen. *GMS Z Med Ausbild.* 2008; 25(1):Doc09.
- Rotthoff T, Bachring T, David D.M., Bartnick C., Linde F., Willers R., Schäfer R. D., Scherbaum W.A., The Value of Training in Communication Skills for Continuing Medical Education, *Patient Educ Couns* 2011a; 84(2):170-5
- Rotthoff T, Ostapczuk M, de Bruin J, Decking U, Schneider M, Ritz-Timme S. Assessing the learning environment of a Faculty. Validation of the Dundee Ready Education Environment Measure (DREEM) with Students and Teachers. *Med Teach* 2011b; 33(11):e624-36.
- Rotthoff T, Ostapczuk M, de Bruin J, Kröncke K, Decking U, Schneider M, Ritz-Timme S. Development and evaluation of a questionnaire to measure the perceived implementation of the mission statement of a competency based curriculum. *BMC Med Educ* 2012; 12:109
- Rotthoff T, Ostapczuk M, Zimmerhofer A, Decking U, Schneider M, Ritz-Timme S. Criterion Validity of a Competency Based Assessment Center in Medical Education. A 4 – year follow-up study. *Med Educ Online* 2014; 19: 25254
- Rotthoff T, Schneider M, Ritz-Timme S, Windolf J. Theory in practice instead of theory versus practice – curricular design for task-based learning within a competency oriented curriculum. *GMS Z Med Ausbild* 2015; 1(32): Doc4.
- Rough D. An analysis of the mission statements of K-12 Christian schools: Their components and the value they hold for the organizational leader. *Diss Abstr Int A: Hum Soc Sci* 2009;70(2-A):618.
- Ruschin S. Prüfungen kompetenzorientiert gestalten. In: *Wissen, was zählt. Ideen für die Lehre Ruhr-Universität-Bochum*, Abt. IFB; Auflage: 1. Auflage 2010); Zentrum für Hochschul- und Qualitätsentwicklung Universität Duisburg-Essen.

S

- Scheffer S. “Validierung des Berliner Global Rating” (BGR)-ein Instrument zur Prüfung kommunikativer Kompetenzen Medizinstudierender im Rahmen klinisch-praktischer Prüfungen (OSCE).” Dissertation, Medizinische Fakultät, Charité –Universitätsmedizin, Berlin, Berlin 2009.
- Scheffer, S., Muehlinghaus, I., Froehmel, A., & Ortwein, H. (2008). Assessing students’ communication skills: validation of a global rating. *Adv health sci educ* 2008; 13(5):583-92.
- Schirmer, J. M., Mauksch, L., Lang, F., Marvel, M. K., Zoppi, K., Epstein, R. M., ... & Pryzbylski, M. Assessing communication competence: a review of current tools. *Fam Med* 2005; 37(3), 184-92.
- Svenson O. Are we all less risky and more skillful than our fellow drivers? *Acta Psychol* 1981; 47: 143-8.

-
- Schuwirth LWT, van der Vleuten CPM, Stoffers HEJH, Peperkamp AGW: Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Med Educ* 1996; 30:50-55.
- Schuwirth LWT, Verheggen MM, van der Vleuten CPM, Boshuizen HPA, Dinant GJ: Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ* 2001; 35:348-56.
- Schuwirth LWT, van der Vleuten CPM: Different written assessment methods: what can be said about their strength and weaknesses? *Med Educ* 2004; 38:974-79.
- Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011; 33:478–85.
- Schuwirth LWT, Van der Vleuten CPM. Assessing Competence, Extending the Approaches to Reliability in: Hodges BD. The Shifting Discourses of Competence in: Lingard L. Ed.; *The Question of Competence - Reconsidering Medical Education in the Twenty-First Century*. Cornell University Press, New York 2012, Kapitel 5. S. 113 – 130.
- Simonsohn AB, Fischer MR. Evaluation of a case-based computerized learning program (CA SUS) for medical students during their clinical years. *Dtsch Med Wochenschr* 2004; 129(11):552-6.
- Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, et al. Doctor–patient communication: the Toronto consensus statement. *BMJ* 1991; 303:1385–7.
- Simpson JG, Furnace J, Crosby J, Cumming AD, Evans PA, David MFB, McPhate GF. The Scottish doctor – learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. *Med. Teach* 2002; 24(2):136-43.
- Spady WG. Organising for results: The basis of authentic restructuring and reform. *Educ Leadership*, 1988; 46(2):4-8.
- Spady WG. Outcome-based education: Critical issues and answers. American Association of School Administrators, 1801 North Moore Street, Arlington, USA. 1994; p 55-60.
- Spencer J. ABC of learning and teaching medicine. *BMJ* 2003; 326:591-4.
- Stacy RD. *Complex Responsive Processes in Organizations*, Routledge, New York 2003.
- Stenhouse, L. *An introduction to curriculum research and development*. Heinemann Educational Publishers, London Guilford, 1975.
- Stewart MA. Effective physician–patient communication and health outcomes: a review. *Can Med Assoc J* 1995; 152:1423–33.
- Street Jr RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Pat Educ Couns* 2009; 74:295–301.

T

- Talbot M. Monkey see, monkey do: a critique of the competency model in graduate medical education *Med Educ* 2004; 38(6):587-92.
- Tekian A, Hodges BD, Roberts TE, Schuwirth L, Norcinin J. Assessing competencies using milestones along the way. *Med. Teach* 2015; 37(4):399-402.
- Thistlethwaite DL. College Environments and the Development of Talent: Characteristics of colleges are related to the percentage of graduates who attain the Ph.D. *Science* 1959; 130(3367):71–6.
- Till H. Climate studies: Can students' perceptions of the ideal educational environment be of use for institutional planning and resource utilization? *Med Teach* 2005; 27:332-7.
- Tyler RW. Principles of Curriculum and Instruction. The University of Chicago Press, Chicago, 1949, 1969, 2013 by The University of Chicago. Published 2013.
- Ten Cate O. Nuts and Bolts of Entrustable Professional Activities. *J Grad Med Educ* 2013; 5(1):157-8
- Ten Cate O, Snell L, Carracio C. Medical competence: the interplay between individual ability and the health care environment *Med Teach* 2010; 32(8):669-75.

U

- Umbrella consortium for assessment networks (UCAN) – Item Management System (IMS)
Zugang nur mit Berechtigung möglich. <https://www.ucan-assess.org/cms/de>.
- USMLE -Federation of State Medical Boards of the United States, Inc., and the National Board of Medical Examiners® (NBME®) 2013-2016 (booklet USMLE Step 2 Clinical Skills (CS) updated March 2016.

V

- Van Bruggen L, Manrique-van Woudenberg M, Spierenburg E, Vos J. Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspect Med Educ* 2012; 1:162-71.
- Van Herwaarden CLA, Laan RFJM, Leunissen RRM. The 2009 *Framework for undergraduate Medical Education in the Netherlands*. 2009, Utrecht: Dutch Federation of University Medical Centres.
- Van Nuland M, Van Den Noortgate W, Degryse J, Goedhuys J. Comparison of two instruments for assessing communication skills in a general practice objective structured clinical examination. *Med Educ* 2007; 41:676–83.
- Van der Vleuten CPM, Schuwirth LWT: Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39:309-17.
- Van der Vleuten CPM, Dannefer EF. Towards a systems approach to assessment *Med Teach* 2012; 34(3):185-86.

Veloski JJ, Rabinowith HK, Robeson MR: A solution to the cueing effects of multiple choice questions: The UN-Q format. *Med Educ* 1993, 27:371-5.

von Fragstein M, Silverman J, Cushing A, Quilligan S, Salisbury H, Wiskin C. UK consensus statement on the content of communication curricula in under-graduate medical education. *Med Educ* 2008; 42:1100–7.

W

Weinert FE. Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: Weinert FE (Hrsg.). *Leistungsmessung in Schulen*. Weinheim: Beltz; 2002. S. 27f.

Weyers S, Jemi I, Karger A, Raski B, Rotthoff T, Pentzek M, Mortsiefer A. Workplace-based assessment of communication skills: A pilot project addressing feasibility, acceptance and reliability. *GMS J Med Educ* 2016; 33(5): Doc70.

Whittle S, Whelan B, Murdoch-Eaton DG. DREEM and beyond; studies of the educational environment as a means for its enhancement. *Educ Health* 2007; 20(1):7.

Winter R. The Assessment of Professional Competences: The Importance of General Criteria in: Edwards A & Knight P.(Eds.) *Assessing competence in higher education*. London Kogan Page Ltd. 1995.

Wissenschaftsrat. Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland auf Grundlage einer Bestandsaufnahme der humanmedizinischen Modellstudiengänge; Wissenschaftsrat Drs. 4017-14, Dresden, 11.07.2014.

Woehr DJ, Arthur W., Jr. The construct-related validity of assessment center ratings: a review and meta-analysis of the role of methodological factors. *J Manag* 2003; 29:231–58.

Wottawa H, Hossiep R. *Grundlagen psychologischer Diagnostik*. Göttingen: Hogrefe 1987.

Y

Yudokowsky R, Downing SM, Ommer D. Prior experiences associated with residents' scores on a communication and interpersonal skill OSCE. *Pat Educ Couns* 2006; 62(3):368-73.

7 Anhang

7.1 Danksagung

Prof. Dr. Stefanie Ritz-Timme und Prof. Dr. Matthias Schneider danke ich ganz besonders herzlich für die uneingeschränkt sehr gute, vertrauensvolle und partnerschaftliche Betreuung in jeder Phase dieser Arbeit!

Den ehemaligen Doktorandinnen und Doktoranden – Dr. Cora Bartnick, Dr. Aurèle Comparot, Dr. Fabian Linde und ganz besonders Dr. Dr. Martin Ostapczuk – danke ich für die sehr gute Zusammenarbeit, inspirierende Diskussionen und Unterstützung bei der Durchführung der Projekte.

Mein besonderer Dank gilt den Kolleginnen und Kollegen - Dr. Thomas Baehring, André Karger, Prof. Dr. Klaus Kröncke, Dr. Achim Mortsiefer, Dr. Michael Pentzek, Dr. Bianca Raski, Dr. Simone Weyers und Prof. Dr. Stefan Wilm - für die sehr gute wissenschaftliche Kooperation.

Den Mitarbeiterinnen und Mitarbeitern des Studiendekanates danke ich für die Unterstützung bei der Datenerhebung und Evaluation in verschiedenen Projekten. Insbesondere bedanke ich mich hier bei Frau Dr. Judith de Bruin.

Zu guter Letzt bedanke ich mich ganz herzlich bei den Studierenden der Medizinischen Fakultät. Ohne sie wären die in dieser Arbeit vorgelegten Untersuchungen nicht möglich gewesen.

7.2 Verzeichnis der wissenschaftlichen Veröffentlichungen und Vorträge

Gesamtübersicht Originalarbeiten

Summe IF: 13,244

Summe gewichtete IF: 15,86

Mortsiefer A, Karger A, **Rotthoff T**, Raski B, Pentzek M. Examiner characteristics and interrater reliability in a communication OSCE. *Pat Educ Couns* 2017 doi:org/10.1016/j.pec.2017.01.013

IF 2.232; 5-Year-IF 2.892

Weyers S, Jemi I, Karger A, Raski B, **Rotthoff T**, Pentzek M, Mortsiefer A. Workplace-based assessment of communication skills: A pilot project addressing feasibility, acceptance and reliability *GMS J Med Educ*. 2016; 33(5):Doc70.

IF not assigned

Rotthoff T, Ostapczuk M, Zimmerhofer A, Decking U, Schneider M, Ritz-Timme S. Criterion Validity of a Competency Based Assessment Center in Medical Education. A 4 – year follow-up study 2014 *Med Educ Online* 2014, 19: 25254

IF 0.833; 5-Year-IF established in 2013

Pentzek M, Abholz HH, Mortsiefer A, Weyers S, Wilm S, **Rotthoff T** Unterrichtskonzepte für die Berufsfelderkundung: Bewertung durch Studierende des ersten Semesters. *Z Allg Med* 2014; 90(6):271-76

IF not assigned

Mortsiefer A, Immecke J, **Rotthoff T**, Karger A, Schmelzer R, Raski B, Schmitten JI, Altiner A, Pentzek M. Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Düsseldorf CoMeD-OSCE. *Pat Educ Couns*. 2014 Jun;95(3):348-55.

IF 2.199; 5-Year-IF 2.975

Rotthoff T, Ostapczuk M, de Bruin J, Kröncke K, Decking U, Schneider M, Ritz-Timme S. Development and evaluation of a questionnaire to measure the perceived implementation of the mission statement of a competency based curriculum. *BMC Med Educ* 2012, 12:109

IF 1.409; 5-Year-IF established in 2013

Mortsiefer A, **Rotthoff T**, Schmelzer R, Immecke J, Ortmanns B, in der Schmitt J, Karger A. Implementierung eines interdisziplinären Unterrichtscurriculums „Kommunikative Kompetenz lehren und prüfen“ im vierten Studienjahr Humanmedizin (CoMeD). *GMS Z Med Ausbild.* 2012; 29(1):Doc06

IF not assigned

Ostapczuk M, Hugger A, de Bruin J, Ritz-Timme S, **Rotthoff T**. DREEM on, dentists! Students' perception of the educational environment in a German dental school as measured by the Dundee Ready Education Environment Measure. *Eur J Dent Educ.* 2012; 16(2):67-77

IF 1.012; 5-Year-IF 1.325

Rotthoff T, Ostapczuk M, de Bruin J, Decking U, Schneider M, Ritz-Timme S. Assessing the learning environment of a Faculty. Validation of the Dundee Ready Education Environment Measure (DREEM) with Students and Teachers. *Med Teach.* 2011; 33(11):e624-36.

IF 1.217; 5-Year-IF 1.653

Büchtemann D, Wollny A, Mortsiefer A, in der Schmitt J, **Rotthoff T**, Karger A, Altiner A. Einstellungen von Studierenden im 4. Studienjahr zur Erlangung kommunikativer Kompetenzen. *Z Allg Med*, 2011. 87 (3): p. 135-142

IF not assigned

Rotthoff T, Baehring T, David D.M., Bartnick C., Linde F., Willers R., Schäfer R. D., Scherbaum W.A., The Value of Training in Communication Skills for Continuing Medical Education, *Pat Educ Couns* 2011; 84(2):170-5

IF 2.237; 5-Year-IF 2.733

Rotthoff T, Baehring T, David D.M., Bartnick C., Willers R., Linde F., Schäfer R.D., Scherbaum W.A. Angebot und Fortbildungsverhalten bei der Continuing Medical Education (CME) im Bereich der Ärztekammer. *Z. Evid. Fortbild. Qual. Gesundh. wesen (ZEFQ)* 2010; 104(10):754-60

IF not assigned

Rotthoff T, Fahrion U, Baehring T, Scherbaum W.A. Die Qualität von CME-Fragen in der Ärztlichen Fortbildung. Eine empirische Studie. *ZaeFQ* 101 (2008); 667-678

IF not assigned

Rotthoff T, Baehring T, Dicken HD, Fahrion U, Fischer M.R., Adler M, Hege I., Scherbaum W.A. Fallbasierte Online-Prüfungen im Medizinstudium – Zielsetzung, Implementierung und praktische Erfahrungen. *GMS Med Inform Biom Epidemiol.* 2006; 2(3):Doc 09

IF not assigned

Rotthoff T, Baehring T, Dicken HD, Fahrion U, Fischer M.R., Scherbaum W.A. Comparison of Long Menu and Open-Ended Questions in computerized medical assessments – A randomized controlled trial. *BMC Med Educ* 2006; 6:50

IF not assigned (established in 2010)

Rotthoff T, Soboll S. Qualitätsverbesserung von MC Fragen: ein exemplarischer Weg für eine medizinische Fakultät. *GMS Z Med. Ausbild.* 2006; 23(3):Doc45

IF not assigned

Lamounier-Zepter, **Rotthoff T**, Ansurudeen, Kopprasch et al. Increased Aldosterone/renin Quotient in Obese Hypertensive Women: A novel Role for Low-density Lipoproteins? *Horm Metab Res* 2006; 38(7):471-8

IF 1.997; 5-Year-IF (established in 2007)

Editorials / Übersichtsarbeiten

Rotthoff T, Schneider M, Ritz-Timme S, Windolf J – Theorie in Praxis statt Theorie versus Praxis - Curriculares Design für ein Lernen an Behandlungsanlässen in einem kompetenzorientierten Curriculum *GMS Z Med Ausbild.* 2015; 32(1):Doc4.

Rotthoff T, Baehring T, David D.M., Scherbaum W.A. Die Effektivität von CME- Qualität durch differenzierte Lehr-/Lernforschung. *Z. Evid. Fortbild. Qual. Gesundh.wesen (ZEFQ)* 103 (2009); 165-168

Rotthoff T, Scherbaum WA. Pharmakotherapie der Adipositas – aktueller Stand und Perspektiven. *Arzneimitteltherapie* 2006; 24:314-9 und in *Med. Monatszeitschrift für Pharmazeuten*

Rotthoff T, Berger M. Die Behandlung der Adipositas (1): Definition, Diagnostik, nicht medikamentöse Therapiemaßnahmen. *Z. Allg. Med.* 2002; 78: 372-378

Rotthoff T, Berger M. Die Behandlung der Adipositas (2): Medikamentöse und chirurgische Therapie – Indikation zur Therapie. *Z. Allg. Med.* 2002; 78:408-414

Buchbeiträge

Psychyrembel – Klinisches Wörterbuch 2011; Thema Diabetologie; Co-Autor

Baehring T, **Rotthoff T**, David D.M. Die ärztliche Online-Fortbildung – 5 Jahre Erfahrung aus Nordrhein. e-Health 2010. Informationstechnologien und Telematik im Gesundheitswesen / Frank Duesberg (Hrsg.) Medical Future Verlag Solingen.

Baehring T, **Rotthoff T**, Scherbaum W.A. Online-Lernen und Online-Prüfen: Wie findet E-Learning Eingang ins medizinische Curriculum in Deutschland? In: Jäckel, A. (Hrsg.) Telemedizinführer Deutschland, Ober-Mörlen, Ausgabe 2006, S. 243-247.

Rotthoff T, Schritt für Schritt. Ein Weg zur erfolgreichen Planung von Lehrveranstaltungen in Neues Handbuch Hochschullehre, Raabe Verlag 2007

Scherbaum WA, **Rotthoff T**: Über- und Unterernährung in Wolf Weihrauch Internistische Therapie 2010/2011, Elsevier, München Auflagen - 2006/2007 bis 2016/2017

Rotthoff T, Scherbaum WA CASUS: Fallbasiertes E-Learning zur: Adipositas therapie veröffentlicht in Wolf Weihrauch Internistische Therapie 2006/2007, Elsevier, München 2006-05-26

Verzeichnis der wissenschaftlichen Vorträge und Poster

Eingeladene Vorträge

Rotthoff T. Die Innere Medizin in Modellstudiengängen: Chance der Integration oder Risiko der inhaltlichen Zerstückelung? Jahrestagung der Deutschen Gesellschaft für Innere Medizin, Mannheim am 9.4.2016

Rotthoff T. „Von der Theorie zur Praxis“ - Initiative Lehre – Integriertes Lernen - Medizinische Universität Wien 27.11. 2015

Rotthoff T. **“Kompetenzbasierte Curriculumplanung in der Praxis”** - Eröffnungsvortrag der Jahrestagung Gesellschaft für Medizinische Ausbildung 1.10. 2015 in Leipzig.

Rotthoff T. „Lehr- und Lernforschung - Evaluation von Aus-, Fort-, Weiterbildung“ 10. Kongress der Deutschen Gesellschaft für Palliativmedizin und der 13. Deutsche Kongress für Versorgungsforschung finden 24. Juni 2014, Düsseldorf

Rotthoff T. „Theorie versus Praxis - Was zählt am Patienten(bett)?“ Tag der Lehre NRW in Münster, 16.11.2013

Wissenschaftliche Vorträge und Poster - Indexed abstracts

- Raski B, Böhm M, Schneider M, **Rotthoff T**. Influence of the personality factors rigidity and uncertainty tolerance on peer-feedback. In: 5th International Conference for Research in Medical Education (RIME 2017). Düsseldorf, 15.-17.03.2017. Düsseldorf: German Medical Science GMS Publishing House; 2017. DocP15
- Eissner A, Raski B, **Rotthoff T**. Influences of personality factors on the acceptance and perception of feedback by medical students. In: 5th International Conference for Research in Medical Education (RIME 2017). Düsseldorf, 15.-17.03.2017. Düsseldorf: German Medical Science GMS Publishing House; 2017. DocP18
- Rotthoff T**, Kroll B, Barz H, In der Schmitt J, Schneider M, Ritz-Timme S. Integration von Theorie in die Praxis – Eigenstudium von Medizinstudierenden im Rahmen von Praxisblöcken: Potential nicht ausgeschöpft. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bern, 14.-17.09.2016. Düsseldorf: German Medical Science GMS Publishing House; 2016. DocFL-10- 479. DOI: 10.3205/16gma034,
- Raski B, Eißner A, Berger R, Schneider M, **Rotthoff T**. Veränderung in Wahrnehmung und Akzeptanz von Peer-Feedback – Wege in Richtung Feedbackkultur. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ). Leipzig, 30.09.-03.10.2015. Düsseldorf: German Medical Science GMS Publishing House; 2015. DocP7-105.
- Eißner A, Raski B, Schneider M, **Rotthoff T**. Geschlechtsspezifische Unterschiede in der Selbst- und Fremdbeurteilung bei Medizinstudierenden. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ). Leipzig, 30.09.-03.10.2015. Düsseldorf: German Medical Science GMS Publishing House; 2015. DocP7-110.
- Rotthoff T**, Schneider M, Ritz-Timme S, Windolf J. Theorie in Praxis statt Theorie versus Praxis - Lernen an Behandlungsanlässen im klinischen Kontext. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg. 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. Doc P255
- Rotthoff T**, Ostapczuk MS, Kröncke KD, Zimmerhofer A, Decking U, Schneider M, Ritz-Timme S. Criterion Validity of a competency based Assessment Center in Undergraduate Medical Education. A 4-year follow-up study. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg. 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. Doc V312
- Eissner A, Raski B, **Rotthoff T**. Unterschiede in der Wahrnehmung von Peer-Feedback zwischen Medizinstudierenden des ersten und vierten Studienjahres unter Berücksichtigung von verschiedenen Persönlichkeitsmerkmalen. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg. 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. DocP245

-
- Raski B, Eissner A, Böhm M, Schneider M, **Rotthoff T**. 360° Peer-Feedback zur Kompetenzentwicklung bei Medizinstudierenden – Akzeptanz bei Studierenden im 1. und 4. Studienjahr des Düsseldorfer Modellstudiengangs Humanmedizin. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg. 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. DocP241
- Galonska L, Zey S, Bosse HM, **Rotthoff T**. Anleitung des Selbststudiums durch Study-Guides im Praxisteil des Düsseldorfer Modellstudiengangs. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg. 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. Doc P221
- Rotthoff T**, Raski B, Böhm M, Schneider M. 360° Peer-Feedback zur Kompetenzentwicklung bei Medizinstudierenden – Erfahrungen aus einer zweijährigen Implementierungsphase im 4. Studienjahr Humanmedizin. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Graz, 26.-28.9.2013. Düsseldorf: German Medical Science GMS Publishing House; 2013. DocP07_07
- Raski B, Böhm M, Schneider M, **Rotthoff T**. Effekte von rigiden Persönlichkeitskonstrukten bei Studierenden im 4. Studienjahr Humanmedizin im Rahmen eines 360° Peer-Feedback Lehrprojektes. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Graz, 26.-28.9.2013. Düsseldorf: German Medical Science GMS Publishing House; 2013. DocP08_06
- Rotthoff T**, Kröncke K, Ostapczuk M, Zimmerhofer A, Decking U, Schneider M. Assessment Center – Ein Prüfungsformat zur Beurteilung und Rückmeldung übergeordneter Kompetenzen bei Medizinstudierenden in kompetenzorientierten Curricula. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Aachen, 27.-29.09.2012. Düsseldorf: German Medical Science GMS Publishing House; 2012. DocV588.
- Rotthoff T**, Ostapczuk M, de Bruin J, Kröncke K, Decking U, Schneider M, Ritz-Timme S. How far away from the goal? Measuring the Implementation of a Mission Statement in Medical Education for Faculty Development. 2nd International Symposium on Research in Medical Education (RIME), Tübingen Mai 2011.
- Rotthoff T**, Kosubek T, Wiegand N, Nilges A, Schneider M. Ein Modell zur Implementierung einer Feedbackkultur in der medizinischen Ausbildung. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). München 05.-08.10.2011. Doc11gma005.
- Herrmann D, Twehues S, Raible M, Henden U, Meinert B, **Rotthoff T**. Technische Aspekte der Erstellung von Lehrvideos. VI. Skills Lab Symposium, Würzburg 8.-9. April 2011
- Rotthoff T**, Kosubek T, Nilges A, Wiegand N, Barz H, Wiegand N, Schneider M. Clinical Competence and the Hidden Curriculum – an Area of Tension 14th Ottawa Conference on The Assessment of Competence in Medicine and The Healthcare, Miami, USA, 2010
- Rotthoff T**, Ostapczuk M, Kröncke K, Schneider M, Decking U, Ritz-Timme S. Erfassung des Lehr-/Lernklimas in der Medizinischen Ausbildung - Validierung des DREEM (Dundee Ready Education Environment Measure) in deutscher Sprache und ergänzender Messinstrumente. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bochum, 23.-25.09.2010. Düsseldorf: German Medical Science GMS Publishing House; 2010. Doc10gma79.

-
- Rotthoff T**, Baehring T, David D.M, Bartnick C, Linde F, Schäfer R., Scherbaum W.A. Communications Training in Continuing Medical Education – A Current Analysis of Supply and Demand in the Area Served by the North Rhine Medical Society in Germany. Research in Medical Education – Changes and Challenges – International Conference, Heidelberg Mai. 2009
- Rotthoff T**, Willers R, Siebler M, Lindner S, Scherbaum WA, Soboll S. Partielle OSCE Prüfung zur Einsparung von Ressourcen. *GMS Z Med Ausb.* 2008; 25 (1): Doc 09
- Rotthoff T**, Richter B, Scherbaum W.A. Prüfungsformat für eine anwendungsorientierte Evidenzbasierte Medizin in der medizinischen Ausbildung Jahrestagung der Gesellschaft für Medizinische Ausbildung, Nov. 2006, Köln, Doc 06gma132. (Poster)
- Rotthoff T**, Baehring T, Dicken HD, Fahrion U. Fallbasierte Online-Prüfung – nur Trend oder auch Nutzen? *GMS Z Med Ausb* 2005; 22(4): Doc123, Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) Münster, November/2005.
- Rotthoff T**, Baehring T, Dicken HD, Fahrion U, Scherbaum WA. Long Menu questions in a computerized case-based assessment- a randomized controlled trial. Conference of the Association of Medical Education in Europe (AMEE), Amsterdam 2005; abstract 9H5 p 171.
- Rotthoff T**, Raski B, Böhm M, Schneider M. 360° Peer-Feedback zur Kompetenzentwicklung bei Medizinstudierenden. Erfahrungen aus einer zweijährigen Implementierungsphase im 4. Studienjahr Humanmedizin GMA Jahrestagung Graz, Österreich 2013)
- Herrmann D, Fox F, Iskender Ö, **Rotthoff T**. Video-Tutorien zur Qualitätssicherung in der medizinischen Ausbildung Ärztliche Untersuchungstechniken im Hein-e-Comp Symposium 2009, Heinrich-Heine-Universität Düsseldorf
- Lautwein F, **Rotthoff T**, Ritz-Timme S, Schneider M. Development, diversity and impact of Mentor-Mentee relationships of the A2 Mentoring Program in the context of the educational culture of the Medical Faculty of Düsseldorf. In: 5th International Conference for Research in Medical Education (RIME 2017). Düsseldorf, 15.-17.03.2017. Düsseldorf: German Medical Science GMS Publishing House; 2017. DocP17
- Barz H, Kroll B, **Rotthoff T**, Schneider M, In der Schmitt J, Ritz-Timme S. ZEITLast- Verlauf des realen Workloads bei Medizinstudierenden im Vergleich zu anderen Studiengängen. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bern, 14.-17.09.2016. Düsseldorf: German Medical Science GMS Publishing House; 2016. DocV12-449. DOI: 10.3205/16gma296
- Lautwein F, **Rotthoff T**, Ritz-Timme S, Schneider M. Entwicklung und Diversität von Mentor-Mentee-Beziehungen im „A²-Mentoring-Programm“ der Medizinischen Fakultät Düsseldorf. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ). Leipzig, 30.09.-03.10.2015. Düsseldorf: German Medical Science GMS Publishing House; 2015. DocP5-066.

-
- Mortsiefer A, Karger A, **Rotthoff T**, Raski B, Büchtemann D. Welche Bedürfnisse äußern Medizinstudierende zum Training kommunikativer Kompetenzen? Evaluation des Düsseldorfer CoMeD-Projekts. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg, 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. DocV334
- Mortsiefer A, Karger A, **Rotthoff T**, Raski B, Schmelzer R, Pentzek M. Wie viele Minuten brauchen Studierende, um kommunikative Kompetenz zu zeigen? Prüfungsqualität und Akzeptanz des Düsseldorfer CoMeD-OSCE nach Verkürzung der Prüfungszeit. 47.Kongress für Allgemeinmedizin und Familienmedizin, 12.-14.9.2013 in München
- Karger A, Raski B, Schaefer R, **Rotthoff T**, Mortsiefer A, Schmelzer R. Training communication skills to deal with difficult emotions - desirable but hard to achieve. In: International Conference of Communication in Healthcare (ICCH). Montreal, 29.09.-02.10.2013. Medical Encounter. 2014 Volume 27 Number 2, Winter 2014.
- Karger A, Raski B, Miekley J, Schmelzer R, **Rotthoff T**, Schäfer R, Körner U. Communicational Behaviour and Attitudes of 4th Year Medical Students. In: International Conference of Communication in Healthcare (ICCH). Montreal, 29.09.-02.10.2013. Medical Encounter. 2014 Volume 27 Number 2, Winter 2014.
- Mortsiefer A, **Rotthoff T**, Karger A, Raski B, Schmelzer R, in der Schmitt J, Pentzek M. Einfluss von Prüfermerkmalen auf die Interrater-Reliabilität im Düsseldorfer CoMeD – OSCE zur Prüfung kommunikativer Kompetenzen. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Aachen, 27.-29.09.2012. Düsseldorf: German Medical Science GMS Publishing House; 2012. DocV593.
- Raski B, Schmelzer R, Mortsiefer A, **Rotthoff T**, Karger A. Lässt sich Interaktionskompetenz objektiv prüfen? – Ergebnisse des Düsseldorfer CoMeD-OSCE. Psychologische Medizin – Österreichische Fachzeitschrift für Medizinische Psychologie, Psychotherapie und Psychosomatik. 2012
- Nilges A, **Rotthoff T**, Wiegand N, Kosubek T, Knoefel W.T, Schneider M. The challenge to implement a feedback culture to improve medical education. International Symposium on Research in Medical Education, Tübingen Mai 2011.
- Mortsiefer A, Pentzek M, Immecke J, Schmelzer R, **Rotthoff T**, Karger A. Prüfung der kommunikativen Kompetenz für die Bewältigung schwieriger Arzt-Patient-Gespräche. Ergebnisse der Düsseldorfer CoMeD-OSCE im 4. Studienjahr. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). München, 05.-08.10.2011. Düsseldorf: German Medical Science GMS Publishing House; 2011. Doc11gma107
- Nilges A, **Rotthoff T**, Meyer F, Wiegand N, Knoefel WT, Schneider M. Formative Prüfungen im Praktischen Jahr zur Entwicklung und Förderung klinischer Expertise. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bochum, 23.- 25.09.2010. Düsseldorf: German Medical Science GMS Publishing House; 2010. Doc10gma15(9)
- Tegeler H, Schmelzer R, Karger A, Mortsiefer A, **Rotthoff T**, Franz M. Einsatz von E-Learning und Schauspielpatienten in der psychosomatischen Lehre Jahrestagung der Deutschen Gesellschaft für Psychosomatische Medizin und ärztliche Psychotherapie, Berlin 2010

Schmelzer R, Mortsiefer A, **Rotthoff T**, Karger A. "Hilfe, mein Patient zeigt Gefühle!" Wie angehende Mediziner die Bedeutung von sowie den Umgang mit Emotionen in der Arzt-Patienten-Interaktion erlernen. Gemeinsamer Kongress der Deutschen Gesellschaft für Medizinische Psychologie (DGMP) und der Deutschen Gesellschaft für Medizinische Soziologie (DGMS) 15. - 18. September 2010 in Gießen

Werner C, **Rotthoff T**, Richter B, Scherbaum W.A. Verwendung des psychometrischen Testverfahrens SCL-90-R bei Patienten einer Adipositasambulanz. DGIM 2006

7.3 Dieser Habilitationsschrift wesentlich zu Grunde liegende Originalarbeiten

Rotthoff T, Baehring T, Dicken HD, Fahron U, Fischer M.R., Adler M, Hege I., Scherbaum W.A. Fallbasierte Online-Prüfungen im Medizinstudium – Zielsetzung, Implementierung und praktische Erfahrungen. *GMS Med Inform Biom Epidemiol.* 2006; 2(3):Doc 09

Rotthoff T, Baehring T, Dicken HD, Fahron U, Fischer M.R., Scherbaum W.A. Comparison of Long Menu and Open-Ended Questions in computerized medical assessments – A randomized controlled trial. *BMC Med Educ* 2006; 6:50

Mortsiefer A, Immecke J, **Rotthoff T**, Karger A, Schmelzer R, Raski B, Schmitten JI, Altiner A, Pentzek M. Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Düsseldorf CoMed-OSCE. *Pat Educ Couns.* 2014 Jun;95(3):348-55.

Rotthoff T, Ostapczuk M, Zimmerhofer A, Decking U, Schneider M, Ritz-Timme S. Criterion Validity of a Competency Based Assessment Center in Medical Education. A 4 – year follow-up study 2014 *Med Educ Online* 2014, 19: 25254

Rotthoff T, Ostapczuk M, de Bruin J, Decking U, Schneider M, Ritz-Timme S. Assessing the learning environment of a Faculty. Validation of the Dundee Ready Education Environment Measure (DREEM) with Students and Teachers. *Med Teach.* 2011; 33(11):e624-36.

Ostapczuk M, Hugger A, de Bruin J, Ritz-Timme S, **Rotthoff T**. DREEM on, dentists! Students' perception of the educational environment in a German dental school as measured by the Dundee Ready Education Environment Measure. *Eur J Dent Educ.* 2012; 16(2):67-77

Rotthoff T, Ostapczuk M, de Bruin J, Kröncke K, Decking U, Schneider M, Ritz-Timme S. Development and evaluation of a questionnaire to measure the perceived implementation of the mission statement of a competency based curriculum. *BMC Med Educ* 2012, 12:109