

1 Einleitung

Der Reformstudiengang Medizin (RSM) der Humboldt-Universität zu Berlin wird parallel zum Regelstudiengang angeboten. Er gilt mit seinem integrierten, problemorientierten und studentenzentrierten Ansatz als Pilotprojekt zur Erprobung einer grundsätzlichen Reform des Medizinstudiums in Deutschland. Im Wintersemester 1999/2000 startete unter dem Dach der Charité der erste Jahrgang des Reformstudiengangs Medizin mit 63 Studierenden parallel zum Regelstudiengang mit 350 Studierenden pro Jahr.^[1]

Die Studierenden des RSM nehmen nicht am ersten Abschnitt der ärztlichen Prüfung (neue AO) teil. Stattdessen finden am Ende jeden Semesters Semesterabschlussprüfungen statt. Diese werden von einem eigenen Prüfungsbereich der Arbeitsgruppe Reformstudiengang konzipiert und durchgeführt, um eine weitgehende Homogenität zwischen Lehre und Prüfung sicherzustellen. Der Schwerpunkt der Prüfungen liegt im schriftlichen und praktischen Bereich. Die praktischen Prüfungen werden vor allem in Form der Objective Structured Clinical Examination (OSCE) (s. Kap. 1.2.3) durchgeführt. Der OSCE wurde erstmals 1975 von Harden et al. beschrieben^[2] und hat sich schon länger international in der medizinischen Ausbildung als Prüfungsmethode für praktische klinische Fähigkeiten etabliert.

Die Erlangung kommunikativer Fähigkeiten ist ein wichtiger Bestandteil des RSM. In der Übung Interaktion, die teilnahmeverpflichtend und im Stundenplan aller Semester fest integriert ist, sollen die Grundlagen von Kommunikation, Interaktion und sozialer Kompetenz vermittelt und von den Studierenden im Hinblick auf ihr eigenes Verhalten reflektiert werden. (s. Kap. 1.1.1) Diese Fähigkeiten wurden von den Prüfungen des RSM jedoch bisher nicht explizit erfasst. Mehrere Studien zur Überprüfung kommunikativer Fähigkeiten wenden dafür unter anderem den OSCE als Prüfungsmethode an.^[3] Im Wintersemester 2001/2002 wurden in allen Jahrgängen des RSM (1., 3. und 5. Semester) OSCE als Teil der Semesterabschlussprüfungen durchgeführt. Diese Gelegenheit wurde genutzt, um bei diesen OSCE-Prüfungen ein spezielles Augenmerk auf die Erfassung kommunikativer Fähigkeiten zu richten.

1.1 Der Reformstudiengang Medizin an der Charité (RSM)

Im Rahmen bundesweiter Protestmaßnahmen entstand aus dem Streik der Studierenden der Freien Universität Berlin (UNIMUT) im Wintersemesters 1988/89 eine studentische Initiative, die sich die Reformierung des Medizinstudiums zum Ziel setzte. Ausgehend von der Kritik an inhaltlichen und strukturellen Defiziten der bisherigen universitären Ausbildung des Medizinstudiums bildete sich eine Inhalts-AG, die ein neues Curriculum erarbeitete und letztendlich in die Arbeitsgruppe Reformstudiengang der Charité übergang. Zentrale

Kritikpunkte der bisherigen Ausbildung betrafen unter anderem die stetige Zunahme spezialisierten Fachwissens, eine mangelhafte Verknüpfung theoretischer und klinischer Inhalte, zu wenig praxisbezogener Unterricht und die Dominanz frontaler Lehrveranstaltungen.^[4]

Das in der Studienordnung festgeschriebene Ziel der Ausbildung „ist die Befähigung zur wissenschaftlichen und praktischen Tätigkeit als Ärztin oder Arzt zur eigenverantwortlichen und selbständigen Berufsausübung und zur Weiterbildung oder sonstigen ärztlichen Qualifizierung“.^[5] Die klassische Unterteilung in Vorklinik mit der Dominanz theoretischer Grundlagen und in Klinik (H-Modell; s. Abb. 1) ist im Curriculum des RSM aufgehoben. Von Anfang an soll entsprechend des N-Modells eine gemeinsame Vermittlung theoretischer Grundlagen und klinischer wie auch sozialwissenschaftlicher Inhalten erfolgen.

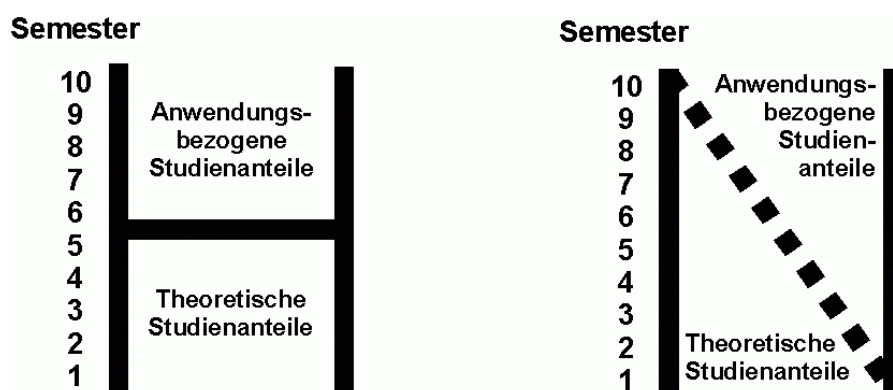


Abb. 1: H- und N- Modell ^[1]

Die Lerninhalte des Reformstudienganges sind in Themenblöcke aufgeteilt und derart gestaltet, dass sie im Sinne einer Lehr- und Lernspirale ^[6] aufeinander aufbauen und immer komplexere Wissensbereiche erschließen. Während sich im ersten Studienabschnitt die Themenblöcke primär an Organen und Organsystemen orientieren (s. Abb. 2), gliedern sich die Themenblöcke des zweiten Studienabschnitts nach Lebensabschnitten mit integrierten mehrwöchigen Blockpraktika in den unterschiedlichen klinischen Fachgebieten (s. Abb. 3). Der dritte Studienabschnitt besteht wie im Regelstudiengang aus dem Praktischen Jahr mit anschließendem zweiten Abschnitt der ärztlichen Prüfung (neue AO).

Wochen:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Semester		Orientierung	Bewegung		NFM 1		Flüssigkeitshaushalt/Herz-Kreislauf				Prüfungen							
2. Semester		Atmung		Pflegekurs		Ernährung/Stoffwechsel/Verdauung				Prüfungen								
3. Semester		Blut		Entzündung/Abwehr				Sexualität/Geschlechtsorgane/Hormone				Prüfungen						
4. Semester		Nervensystem/Koordination		BP Wahlpflicht		Elektrolyte/Niere				Prüfungen								
5. Semester		Haut		NFM 2		Sinnesorgane				Psyche und Erleben				Prüfungen				

Abb. 2: Blockgliederung des ersten Studienabschnitts ^[7]

Wochen:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
6. Semester	PP	BP Gynäkologie/Geburtshilfe			Schwangersch./Geburt/Neugeb.			Saugling/Kleinkind			Prüfungen							
7. Semester	PP	BP Pädiatrie			Schulkind			Adoleszenz			BZK	Lebensmitte I			Prüfungen			
8. Semester	PP	Lebensmitte II			BZK			BP Neurologie			BP Psychiatrie			Prüfungen				
9. Semester	PP	Lebensmitte III / Alter			BZK			BP Innere Medizin			BP Geriatrie			Prüfungen				
10. Semester	NFM 3		BP Chirurgie			BP Wahlpflicht			Vorbereitung für das Staatsexamen									

PP: Propädeutik BP: Blockpraktikum BZK: Berufs- und Zivilisationskrankheiten NFM: Notfallmedizin

Abb. 3: Blockgliederung des zweiten Studienabschnitts ^[7]

Die zentrale Lehr- und Lernform des RSM ist das Problem-Orientierte-Lernen (POL). Anhand eines konkreten Falles werden in POL-Kleingruppen, die sich aus sieben Studierenden und einem Dozenten zusammensetzen, Fragen entwickelt. Die Bearbeitung dieser Fragen erfolgt im Selbststudium. In der folgenden POL-Sitzung werden die Ergebnisse zusammengetragen und gemeinsam diskutiert. Die Arbeit in Kleingruppen soll die Verantwortung der Studierenden stärken und sie für Interaktion und Zusammenarbeit sensibilisieren. Das gelernte Wissen wird diskutiert und hinterfragt. Gleichzeitig soll die Motivation zum kontinuierlichen Wissenserwerb geweckt werden. Das Selbststudium ist fester Bestandteil des Stundenplans. Mit dem „Trainingszentrum für ärztliche Fertigkeiten“ (TÄF) sind dafür durch fachgerechte Betreuung und Bereitstellung von Räumen und Materialien gute Bedingungen geschaffen worden. Zur Vertiefung des Wissens und der Vermittlung interdisziplinärer Zusammenhänge werden fächerübergreifende Seminare in den Bereichen Grundlagen ärztlichen Denkens und Handelns, klinisch theoretische Grundlagen, Methoden wissenschaftlichen Denkens und Handelns, Gesundheitswissenschaften, sowie Übungen zu Diagnostik und Untersuchung angeboten.

Ab dem zweiten Semester findet einmal wöchentlich ein Praxistag statt, an dem die Studierenden bei einer/m niedergelassenen Kollegin/en die Möglichkeit haben, beobachtend und assistierend an der Sprechstunde teilzunehmen und möglichst in Untersuchung, Beratung und Behandlung einbezogen werden.

In den Ausbildungszielen der Studienordnung des RSM wird explizit darauf hingewiesen, dass „vor allem Fähigkeiten zu Kommunikation und Interaktion, das interdisziplinäre Denken, sowie die Bereitschaft zu lebenslangem Lernen gefördert werden“ sollen.^[5] Neben Praxistag, POL-Sitzungen, wöchentlichen Seminaren in Grundlagen ärztlichen Denkens und Handelns und Gesundheitswissenschaften, wie auch dem integrierten Studium Generale soll vor allem mit der Übung Interaktion in diesem Bereich ein Schwerpunkt gesetzt werden.

1.1.1 Übung Interaktion

Ausgangspunkt für den RSM war unter anderem der Wunsch, die kommunikativen Fähigkeiten der Studierenden im Studium zu fördern. Kommunikative Fähigkeiten sind Voraussetzung für erfolgreiche Arzt-Patientenkontakte.^[8] Sie können gelehrt und gelernt werden,^[9] gehen jedoch verloren, wenn sie nicht regelmäßig aufgefrischt werden.^[10, 11] Eine Studie aus den Niederlanden, die Fähigkeiten im medizinischen Interview vom ersten bis zum sechsten Studienjahr prüfte, konnte zeigen, dass mit kontinuierlichem intensivem Unterricht in Kleingruppenformat mit Einsatz von Simulationspatientinnen/-en (s. Kap. 1.1.2) und regelmäßigem Experten- bzw. Kommilitonen-Feedback während des gesamten Curriculums ein stetiger Zuwachs der Fähigkeiten erzielt wurde.^[12]

Um eine kontinuierliche Vermittlung kommunikativer Fähigkeiten sicherzustellen, ist die Übung Interaktion fester Bestandteil des Stundenplans des RSM und wird teilnahmeverpflichtend, 14-tägig über das gesamte Studium hinweg bis zum Beginn des Praktischen Jahres angeboten. Der Unterricht findet in den POL-Kleingruppen von sieben Studierenden und mindestens einer/m Dozentin/en statt. Zentraler Inhalt der Übung „Interaktion“ ist die Vermittlung von Grundlagen der Kommunikation und Interaktion, sozialer Kompetenz sowie die Unterstützung der Gruppenarbeit.

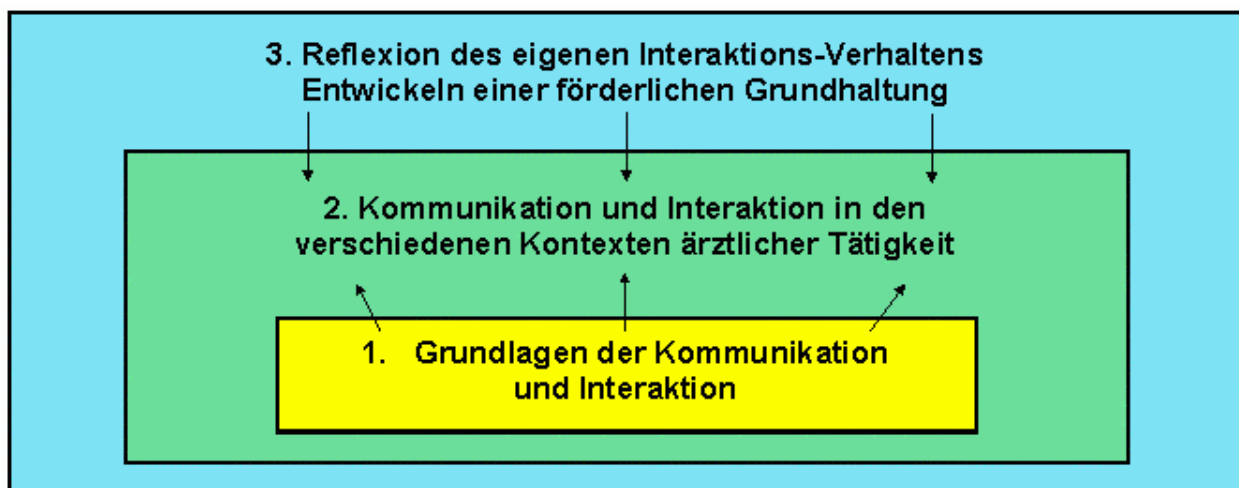


Abb. 4: Konzept der Übung Interaktion ^[13]

Ausgehend von der Vermittlung allgemeiner Grundlagen von Kommunikation und Gesprächsführung werden in Rollenspielen und Simulationspatientengesprächen die Fertigkeiten im ärztlichen Kontext praktisch geübt. Ziel der Übung ist schließlich die Vermittlung von Kompetenzen, die Studierende befähigen, eine Grundhaltung zu entwickeln, die einen partnerschaftliche Kontakt und die Reflexion des eigenen Verhaltens fördern (s. Abb. 4). Zum Zeitpunkt des Wintersemesters 2001/2002 wurden schwerpunktmäßig folgende Themen

behandelt.^[14]

1. Semester: Grundlagen der Kommunikation, Teamarbeit, Feedback
3. Semester: Einführung in die Gesprächsführung, Einführung in die Sexualanamnese
5. Semester: Psychosoziale Aspekte im ärztlichen Gespräch, Konzepte der Psychosomatik

Zur Überprüfung der Qualität bestehender Semesterkonzepte und der Fertigstellung und Implementierung des übergeordneten Gesamtkonzeptes der Übung Interaktion wurde 2001 die Arbeitsgruppe „Interaktionskonzepte“ (AGIK) gegründet. Ihr gehören Mitarbeiter/-innen des RSM, Studierende, Dozentinnen/-en der Übung Interaktion und Mitarbeiter/-innen des Institutes für Allgemeinmedizin an. Die folgende Zusammenstellung zeigt die Inhalte der jeweiligen Semester des von der AGIK entwickelten aktuell gültigen Gesamtkonzeptes:^[13]

Semester	Inhalte
1	In Kontakt kommen; soziale Kompetenzen; Selbst-/ Fremdwahrnehmung; verbale und nonverbale Kommunikation, Unterstützung der Zusammenarbeit der POL-Gruppen
2	Unterstützung der Studierenden im Praxistag; grundlegende Kommunikationsmodelle; Fragen stellen; Gesprächseinstieg und -abschluss; Unterstützung der Zusammenarbeit der POL-Gruppen
3	<i>Anamnese I:</i> Informationen bekommen: Anamnesen erheben; „Let’s talk about sex“: Wie man sexuelle Themen mit Patientinnen/-en bespricht; Erotik in der Beziehung zwischen Arzt/Ärztin und Patient/-in
4	<i>Anamnese II:</i> Anamnese erheben mit Fokus auf psychosoziale Aspekte, unterstützend dazu Video-Analyse des eigenen Gesprächsverhaltens
5	<i>Anamnese III:</i> Anamnesegespräche mit dem Schwerpunkt psychosomatische Patientinnen/en
6	Informationen geben und gemeinsame Entscheidungen ermöglichen: Aufklärungs- und Beratungsgespräche; Unterstützung der Studierenden im Blockpraktikum
7	Unterstützung der Studierenden im Blockpraktikum; Kommunikation mit Kindern, Jugendlichen und deren Eltern
8	Unterstützung der Studierenden in den Blockpraktika; Kommunikation mit neurologischen und psychiatrischen Patientinnen und Patienten
9	Unterstützung der Studierenden in den Blockpraktika; Überbringen schlechter Nachrichten; Umgang mit sterbenden /schwerkranken Patientinnen/-en und deren Angehörigen
10	Patientenpräsentation

Wichtiger Bestandteil der Übung Interaktion ist der Einsatz von Simulationspatientinnen/-en, mit deren Hilfe gelernte Fertigkeiten praktisch geübt werden können. Der Aufbau eines Simulationspatientenprogramms im Rahmen des RSM machte dies erst möglich.

1.1.2 Simulationspatientenprogramm

1964 wurde der Einsatz von Simulationspatientinnen/-en (SP) erstmalig beschrieben. Barrows setzte die von ihm damals „programmierte Patienten“ genannten SP zur Prüfung klinischer Kompetenzen im Bereich der Nervenheilkunde ein.^[15] In den USA sind Simulationspatientenprogramme mittlerweile weitgehend etabliert, während sie in Deutschland

noch im Entstehen sind. Zu Beginn des RSM wurde bereits im Wintersemester 1999/2000 an der Charité ein Simulationspatientenprogramm initiiert, das im Trainingszentrum für ärztliche Fertigkeiten (TÄF) integriert ist.^[16]

In der Literatur tauchen sowohl die Bezeichnung „simulated patient“ als auch „standardized patient“ auf. *Simulationspatientinnen/-en* sind normale Personen, die sorgfältig trainiert wurden, Symptome und Zeichen einer/s Patientin/-en zu präsentieren. Die Betonung liegt dabei auf der Simulation der Realität. *Standardisierte Patientinnen/-en* werden trainiert, einen medizinischen Fall in konsistenter Art und Weise zu präsentieren, wobei der Schwerpunkt auf der Konsistenz der Präsentation liegt. Standardisierte Patientinnen/-en können simulierte oder auch reale Patientinnen/-en sein.^[17] Der Trend, die Bezeichnung Simulationspatient durch standardisierten Patient zu ersetzen, spiegelt die Notwendigkeit wider, ein standardisiertes Patientenproblem zu präsentieren, das nicht von Studierenden zu Studierenden variiert.^[18] Die Konstanz der Präsentation ist eine wichtige Voraussetzung für den Einsatz von SP in Prüfungen. Die Abkürzung SP wird in der Literatur für beide Bezeichnungen benutzt. In dieser Arbeit steht SP für die Bezeichnung Simulationspatient/-in.

SP sind vielseitig einsetzbar. Sie werden ausgebildet, die verschiedensten Patientenrollen zu übernehmen. Sie können bei der Anamneseerhebung, bei Aufklärungsgesprächen oder beim Erlernen von Untersuchungstechniken eingesetzt werden. Zur Simulation gehören unter anderem die Wiedergabe der Patientengeschichte, die Darstellung der Erkrankungssymptome und die Persönlichkeitsmerkmale der Rolle. SP geben den Studierenden die Möglichkeit, frühzeitig in *sicherer Umgebung* praktische Fähigkeiten zu erlernen und zu erproben. Der Vorteil der SP liegt darin, verschiedene Krankheitsbilder zuverlässig mehrfach hintereinander vorzutragen, ohne reale Patienten zu belasten. SP können in für reale Patientinnen/-en unzumutbaren Situationen, beispielsweise bei der „Überbringung schlechter Nachrichten“ eingesetzt werden. Weitere wichtige Vorteile gegenüber echten Patientinnen/-en sind die ständige Verfügbarkeit unabhängig von Ort und Zeit und die Möglichkeit, nach einem Interview als betroffene Person den Studierenden gezieltes Feedback zu geben.^[18] Barrows weist jedoch darauf hin, dass SP echte Patientinnen/-en nicht ersetzen sollen, sondern mit ihren Vorteilen eine sinnvolle Ergänzung in der medizinischen Ausbildung sein können. Einschränkend ist zu sagen, dass nicht alle Symptome simuliert werden können wie beispielsweise pathologische Lungen- oder Herzgeräusche, Ödeme etc. Außerdem bedarf es für Training, Betreuung und Bezahlung der SP eines erhöhten Zeit- und Kostenaufwandes. Collins und Harden haben einen ausführlichen Artikel veröffentlicht, der Vor- und Nachteile der SP bzw. realen Patientinnen/-en beschreibt und Kriterien für deren Einsatz festlegt.^[17] Neben der Darstellung von Patientenrollen und der

Feedback-Gabe werden SP zunehmend auch eingesetzt, um die Leistungen von Studierenden zu bewerten. In einem Review von Vu und Barrows über die Anwendung von SP in klinischen Prüfungen wird beschrieben, dass SP die Leistung von Prüflingen, gerade auch in Bezug auf Interaktion, zuverlässig beurteilen können.^[19] Ein Beispiel ist das Clinical Skills Assessment (CSA), das ausländische Ärzte durchlaufen müssen, bevor sie in den USA arbeiten dürfen. Die Prüflinge treffen dabei mehrfach auf SP und werden von diesen bezüglich Anamnese und Kommunikationsfähigkeit bewertet.^[20]

Das Simulationspatientenprogramm des RSM wurde initiiert, um Studierenden die Möglichkeit zu geben, in einer sicheren quasi-realen Atmosphäre praktische Kompetenzen zu erproben und zu verbessern. SP werden im RSM regelmäßig in der Übung Interaktion und in den praktischen Prüfungen eingesetzt. Zunehmend kommen SP aber auch in Lehrveranstaltungen des Regelstudienganges bzw. in Fortbildungsveranstaltungen einzelner Abteilungen der Charité zum Einsatz.

1.2 Prüfungen in der medizinischen Ausbildung

Da Prüfungen das Lernverhalten von Studierenden wesentlich mitbestimmen, wird das Curriculum eines Studienganges entscheidend vom Prüfungssystem beeinflusst. Gelernt wird das, was geprüft wird. Studenten konzentrieren sich auf Themen, von denen sie wissen, dass diese in Prüfungen abgefragt werden. Themen, die nicht in Prüfungen behandelt werden, drohen vernachlässigt zu werden.^[21]

Prüfungen werden oft lediglich als summative, das heißt bestehensrelevante Erfolgskontrollen am Ende eines Kurses angewandt. Prüfungen erfüllen neben der Erfolgskontrolle jedoch noch weitere Aufgaben. Ein Aufsatz von Lowry fasst diese treffend zusammen: „Das Prüfen von Studierenden der Medizin soll ihnen während eines Kurses helfen, ihr Lernen zu fokussieren, ihre individuellen Stärken und Schwächen zu erkennen, eine Möglichkeit zur Verbesserung liefern, Unzulänglichkeiten im Inhalt oder der Durchführung eines medizinischen Kurses aufzeigen und letztendlich die Öffentlichkeit vor inkompetenten Absolventen schützen. Um dies alles zu leisten, muss das Prüfungssystem ein gut organisiertes formatives Element beinhalten.“^[22] Formative Prüfungen sind nicht bestehensrelevante, regelmäßige Tests der Lerninhalte, die angeboten werden, ohne eine Sanktion (z. B. Wiederholung des Kurses) befürchten zu müssen. Sie können als integraler Teil des Lehr- und Lernprozesses während eines Kurses durchgeführt werden. Prüfungen sind letztendlich eine Form des Lernens und sollen dabei Orientierung und Unterstützung geben.^[23] Im Prüfungssystem des RSM sind solche Leistungstests auch integriert, um den Lehrenden und Lernenden kontinuierlich Rückmeldung

über den Lehr- und Lernerfolg zu geben.^[24]

Miller stellte 1990 in einem Vortrag zur Prüfung klinischer Kompetenzen ein Pyramiden-Modell vor (s. Abb. 5), das die Entwicklung der Kompetenzen von Studierenden in Studium und Weiterbildung veranschaulicht: Die Basis (Knows) wird vom theoretischen Wissen gebildet, das die Studierenden benötigen. Die nächste Stufe (Knows how) bildet das Wissen über die praktische Umsetzung. Auf der darüber liegenden Stufe (Shows how) können die Studierenden zeigen, wie das Gelernte ausgeführt wird. Die vierte Stufe (Does) des Modells steht für das Tun in der alltäglichen Praxis unabhängig von einer künstlichen Prüfungssituation. Ziel einer jeden Prüfung sollte sein, auf der höchst möglichen Stufe zu prüfen.^[25]

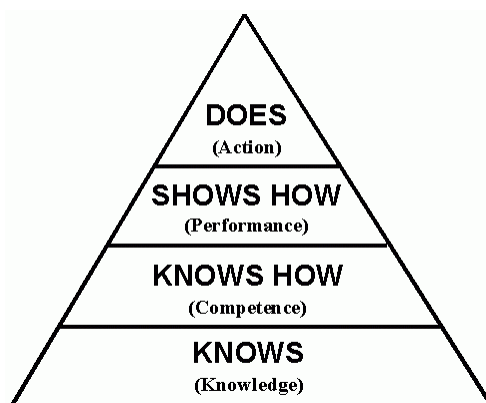


Abb. 5: Pyramide von Miller^[25]

Anhand des Modells wird deutlich, dass unterschiedliche Kompetenzen nicht von einer einzigen Prüfungsmethode ausreichend erfasst werden können. Es werden verschiedene Methoden benötigt, um unterschiedliche Kompetenzen auf der jeweiligen Stufe sinnvoll zu prüfen. In den Jahren der Forschung zur medizinischen Ausbildung entstand eine Vielzahl unterschiedlichster Prüfungsmethoden, wie beispielsweise die Multiple Choice Questions (MCQ) (s. Kap. 1.2.2) oder die Objective Structured Clinical Examination (OSCE) (s. Kap. 1.2.3.).

Für die Auswahl einer geeigneten Prüfungsmethode zur Überprüfung bestimmter Kompetenzen stehen verschiedene Kriterien zur Verfügung. Van der Vleuten entwickelte ein Modell, mit dem der Nutzen einzelner Prüfungsmethoden bestimmt wird. Die wesentlichen Kriterien dieses Modells sind Reliabilität, Validität, lernsteuernde Wirkung, Kosten/Nutzen-Verhältnis und die Akzeptanz bei Lehrenden und Lernenden.^[26] Die Gewichtung einzelner Kriterien kann jedoch unterschiedlich sein, je nachdem, ob es sich um „high-stakes“ Prüfungen (z. B. Staatsexamen) oder um Feedback an Studierende während eines Kurses handelt.^[27] Die *Reliabilität* ist ein Maß für die Reproduzierbarkeit des Prüfungsergebnisses. Sie wird als ein Koeffizient mit dem Wertebereich 0 (keine Reliabilität) bis 1 (absolute Reliabilität) dargestellt. Oft ist ein Wert von 0,8 als minimal akzeptabler Wert genannt, obwohl er in Abhängigkeit vom Zweck der Prüfung

höher oder niedriger angegeben wird.^[27] Die *Validität* besagt, inwieweit die Prüfungsmethode tatsächlich das misst, was sie zu messen beabsichtigt. Die *lernsteuernde Wirkung* beeinflusst die Lerninhalte, die Prüflinge beherrschen. Bestimmte Kompetenzen werden mit einzelnen Prüfungsformaten besser erfasst als andere, die vernachlässigt werden. Neben einem angemessenen *Kosten/Nutzen-Verhältnis* ist die *Akzeptanz* bei Studierenden und Lehrenden ein entscheidender Faktor für die Umsetzung einer neuen Prüfungsmethode.

Qualitativ hochwertige Prüfungen, die den wesentlichen Testgütekriterien wie Validität, Reliabilität und Objektivität genügen, erfordern einen aufwändigen Entwicklungs- und einen ebenso aufwändigen Qualitätssicherungsprozess. Da dies im Aufbauprozess des RSM nur begrenzt leistbar war, konzentrierte sich der Prüfungsbereich des RSM zunächst auf die Prüfungsformate „Multiple Choice Questions“ (MCQ) und „Objective Structured Clinical Examination“ (OSCE).^[24]

1.2.1 Prüfungen im RSM

Die Reform des Curriculums des RSM bedeutet eine grundlegende Umstrukturierung des Prüfungssystems gegenüber dem des Regelstudiengangs. Der in der neuen Approbationsordnung vorgeschriebene erste Abschnitt der ärztlichen Prüfung entfällt. Stattdessen finden am Ende jedes Semesters summative, das heißt teilnahmeverpflichtende und bestehensrelevante, benotete Semesterabschlussprüfungen statt. Nach dem zehnten Semester absolvieren die Studierenden des RSM, wie ihre Kommilitonen aus dem Regelstudiengang, den dritten Studienabschnitt in Form des Praktischen Jahres mit dem sich anschließenden zweiten Abschnitt der ärztlichen Prüfung.

Die Semesterabschlussprüfungen werden unter der Aufsicht eines entsprechend der Prüfungsordnung des RSM^[28] eingesetzten Prüfungsausschusses von den Mitarbeiterinnen/-n des Prüfungsbereichs der Arbeitsgruppe Reformstudiengang koordiniert. In Zusammenarbeit mit den an den Themenblöcken beteiligten Fachabteilungen erfolgt die Konzeption, Erarbeitung und Durchführung der Prüfungen. Inhaltliche Grundlage der Semesterabschlussprüfungen sind die in den Blockbüchern der Themenblöcke des Semesters aufgeführten Lernziele. Diese sollen Lernenden und Lehrenden bei der Bearbeitung der jeweiligen Themenblöcke Orientierung geben. Die Festlegung der Lernziele erfolgt von Dozentinnen/-en bzw. Vertreterinnen/-n der beteiligten Fachrichtungen in den Blockplanungssitzungen der einzelnen Themenblöcke. Die Lernziele beschreiben jeweils unterschiedliche Kompetenzen bzw. Ebenen der „Miller-Pyramide“. Dementsprechend werden die Lernziele jedes Semesters auf drei unterschiedliche Kategorien aufgeteilt: „kognitive“, „anwendungsbezogene“ und „emotionale“ Lernziele. In Abhängigkeit der Lernziele des aktuellen Semesters legt der Prüfungsausschuss des RSM zu

Beginn des Semesters die Prüfungsmethoden der Semesterabschlussprüfungen fest, um die Lernziele der verschiedenen Kategorien optimal zu prüfen. Bezüglich der summativen Semesterabschlussprüfungen legte sich der Prüfungsausschuss des RSM zunächst vornehmlich auf die Prüfungsformate „Multiple Choice Questions“ (MCQ) und „Objective Structured Clinical Examination“ (OSCE) fest.^[24] Weitere Formate wie beispielsweise fallbasierte schriftliche Prüfungen, computerbasierte Prüfungen oder strukturierte mündliche Prüfungen werden vereinzelt in formativen, nicht bestehensrelevanten Prüfungen angewandt. Aufgabe der Mitarbeiter/-innen des Prüfungsbereichs des RSM ist es unter anderem, neue Prüfungsformate zu erproben und weiterzuentwickeln.

In den folgenden Kapiteln werden die beiden im RSM regelmäßig zur Anwendung kommenden Prüfungsformate OSCE und MCQ genauer beschrieben. Daran anschließend wird nochmals auf die Testgütekriterien Reliabilität und Validität eingegangen.

1.2.2 Multiple-Choice-Question-Verfahren (MCQ)

Zu Beginn des RSM beschloss der Prüfungsausschuss, nach jedem Semester die kognitiven Lernziele mit dem Multiple-Choice-Question-Verfahren (MCQ) zu überprüfen. Das MCQ-Format wurde in die staatlichen Prüfungen der medizinischen Ausbildung mit der Absicht eingeführt, mehr Objektivität bei der Bewertung der Leistungen der Studierenden zu erlangen. Bis zu diesem Zeitpunkt wurden alle Prüfungen als mündliche, teils auch praktische Prüfungen durchgeführt, in denen in der Regel vier Kandidaten von einem Dozenten oder Hochschullehrer ohne Beisitzer bzw. Protokollant geprüft wurden.^[29] Das MCQ-Format ermöglichte es, eine große Menge kognitiven Wissens in angemessener Zeit ressourcenschonend zu überprüfen und auszuwerten. Nach jahrelangem Einsatz des MCQ-Formates mehrte sich jedoch auch negative Kritik, wie beispielsweise das Prüfen von isoliertem Detailwissen oder der Einfluss auf das Lernverhalten der Studierenden, die sich in der Prüfungsvorbereitung zunehmend auf das Lernen von „Altfragen“ konzentrierten. Ein Großteil der Kritikpunkte an MCQ-Prüfungen sind jedoch nicht durch das Format bedingt, sondern können durch sorgfältige Gestaltung der Fragen vermieden werden.^[30] Die Vorteile von MCQ-Prüfungen gegenüber anderen schriftlichen Formaten überwiegen. Sie sind objektiv, reliabel, bei großen Studierendenzahlen kostengünstig und bei Studierenden und Lehrenden bekannt.^[24] Um zu verhindern, dass Studierende wie bei den Staatsexamensprüfungen des IMPP lediglich „Altfragen“ lernen, werden die MC-Fragen des RSM nicht veröffentlicht. Die Fragenautorinnen/-en sind Angehörige der Fakultät und meist Dozentinnen/-en der zu prüfenden Themenblöcke. Vor Anwendung der MC-Fragen in den Semesterabschlussprüfungen durchlaufen die Fragen zur Qualitätssicherung einen aufwendigen

Review-Prozess. Alle Fragen werden von Reviewkomitees, denen Mitarbeiter/-innen mehrerer Fachbereiche angehören, begutachtet. Bei Fehlern oder auftretenden Kritikpunkten gehen die Fragen zurück an die Autorinnen/-en. Nach Korrektur kommt die Frage wieder in das Reviewkomitee.

1.2.3 Objective Structured Clinical Examination (OSCE)

Zur Überprüfung „anwendungsbezogener“ Lernziele wird bei den Semesterabschlussprüfungen des RSM überwiegend die Objective Structured Clinical Examination (OSCE) als Prüfungsmethode eingesetzt. Der OSCE ist mittlerweile ein weit verbreitetes Prüfungsformat für die Erfassung praktischer klinischer Fertigkeiten, der Anamneseerhebung wie auch interpersoneller Fähigkeiten. Das OSCE-Format wurde in den 70er Jahren erstmals von Harden beschrieben und wurde aus dem Wunsch heraus entwickelt, eine objektivere Herangehensweise für die Prüfung klinischer Kompetenzen zu finden.^[2] Bis zu diesem Zeitpunkt wurden klinische Prüfungen überwiegend unstrukturiert am Bett von Patientinnen/-en durchgeführt. Bei der Prüfung an nur einer/m Patientin/-en konnte die Schwerpunktsetzung und der Schwierigkeitsgrad der Prüfungen je nach Fall äußerst unterschiedlich sein, ebenso der Beurteilungsmaßstab der Prüfer/-in. Durch „Sampling“, dem Prüfen an mehreren unterschiedlichen Stationen und durch unterschiedliche Prüfer/-innen, sollte das OSCE-Format in strukturierter und standardisierter Weise der Prüfung klinischer Kompetenzen einen objektiveren Rahmen geben.

Ein OSCE besteht aus mehreren voneinander unabhängigen Stationen. Die Studierenden absolvieren einen Prüfungsparcours, in dem sie von einer Station zur nächsten rotieren, bis alle Studierenden jede Station absolviert haben. Die Stationen haben jeweils einen festen Zeitrahmen (5 bis 20 Minuten), in dem eine Aufgabenstellung erfüllt werden muss. Die Aufgabenstellungen sind vielfältig und beinhalten beispielsweise eine körperliche Untersuchung, eine Anamneseerhebung oder Basic Life Support an einer Reanimationspuppe. Die Leistung der Studierenden wird von Prüferinnen/-n anhand von Checklisten oder Global Ratings bewertet, wodurch eine weitgehende Objektivierung erzielt werden soll. Eine *Checkliste* versucht, die Leistung der Studierenden zu operationalisieren. Sie besteht aus einer Liste detaillierter Items (Punkte), die von den Studierenden in der Station gezeigt werden sollen. Die Items werden mit Ja oder Nein „abgecheckt“. *Global Ratings* beinhalten im Gegensatz zu Checklisten weniger oder nur ein Item, die allgemeiner formuliert sind und auf einer Likert-Skala mit mehreren Abstufungen bewertet werden (z. B. von sehr schlecht bis sehr gut auf einer Skala von 1 bis 10). Im RSM besteht ein OSCE im Regelfall aus 8 bis 10 Stationen mit einer Zeitvorgabe von 5 bis 10 Minuten pro Station. Die Zeitdauer richtet sich nach dem Wissensstand. In höheren

Semestern werden komplexerer Stationen konzipiert, für die längere Stationszeiten nötig sind. Im dritten Semester mit den Themenblöcken „Blut“, „Entzündung/Abwehr“ und „Sexualität/Geschlechtsorgane/Hormone“ bestand der OSCE im Wintersemester 2001/2002 beispielsweise aus folgenden Stationen: Bed-Side-Test, infektiös-entzündliche Anamnese, Lymphknotenuntersuchung, mikrobiologische Untersuchung, Sexualanamnese, gynäkologische Untersuchung und Brustuntersuchung. Die Checklisten bestehen aus jeweils 10 bis 20 Items, mit den Optionen „vollständig erfüllt“, „teilweise erfüllt“ und „nicht erfüllt“. Die Option „teilweise erfüllt“ wurde nach den Erfahrungen der ersten OSCE auf Anregung der Prüfer/-innen eingefügt, da in der Praxis eine klare Entscheidung zwischen Ja – „Item wurde erfüllt“ und Nein – „Item wurde nicht erfüllt“ in einigen Fällen schwer fiel. An jeder Station bekommen die Studierenden einen Prozentwert (erreichte Punkte / maximal mögliche Punkte). Das Gesamtergebnis jedes Prüflings ergibt sich aus dem arithmetischen Mittel der Prozentwerte aller Stationen. Die Bestehensgrenze liegt im RSM zur Zeit bei 60%.

1.3 Testgütekriterien

Die klassische Testtheorie definiert drei zentrale Testgütekriterien: Objektivität, Reliabilität und Validität. Die Objektivität eines Testes gibt an, inwieweit Testergebnisse vom Testanwender unabhängig sind. Ein Test ist objektiv, wenn verschiedene Testanwender (Prüfer/-in) bei denselben Personen zu den gleichen Resultaten gelangen. In Publikationen zur medizinischen Ausbildung wird häufig auf die beiden weiteren Testgütekriterien Reliabilität und Validität verwiesen. Wegen der Komplexität dieser beiden Begriffe wird in den folgenden beiden Abschnitten genauer darauf eingegangen.

1.3.1 Reliabilität

Die Reliabilität (Zuverlässigkeit) bezeichnet den Grad der Messgenauigkeit eines Messinstrumentes. Nach der klassischen Testtheorie setzt sich ein Testergebnis aus dem „wahren Wert“ und dem Messfehler zusammen. Die Reliabilität gibt an, wie stark Messwerte durch Störeinflüsse und Fehler belastet sind.^[31] Ein äußerst reliabler Test misst so genau, dass er für ein Individuum bei jeder erneuten Messung das gleiche Ergebnis liefert. Es gibt mehrere Methoden, mit denen Reliabilitäten berechnet werden. Eine davon ist die Bestimmung der inneren Konsistenz eines Instrumentes, die eine Aussage über die Homogenität der Items einer Checkliste trifft. Ein Maß für die innere Konsistenz ist beispielsweise „Cronbach's Alpha“. Eine andere Methode ist die Retest-Reliabilität. Sie ist definiert als Korrelation zwischen zwei Messwerten derselben Stichprobe. Die Retest-Reliabilität gibt an, wie viel Prozent der Gesamtunterschiedlichkeit (Varianz) der Testergebnisse auf wahre Merkmalsunterschiede

zurückzuführen sind. Der Rest geht auf Fehlereinflüsse zurück. In der Literatur zu OSCE-Prüfungen wird beispielsweise häufig die Interrater-Reliabilität erfasst. Sie beschreibt die Reproduzierbarkeit des Testergebnisses eines Studierenden bei der Bewertung durch unterschiedliche Prüfer/-innen.

Um die Reliabilität einer Methode zu erhöhen, müssen möglichst viele Fehler- und Störungsquellen vermieden werden. Für das Beispiel der Interrater-Reliabilität gilt als wesentliche Voraussetzung für stabile Bewertungen eine gute Schulung der Prüfer/-innen, was unter anderem eine genaue Vorbereitung auf die Prüfungssituation und detaillierte Kenntnis der qualitativen Abstufung der Bewertungsskalen beinhalten.^[32] Das Streben nach Objektivierungen von Prüfungsmethoden (z. B. MCQ, OSCE) zielt unter anderem auf die Erhöhung der Reliabilität. Van der Vleuten weist darauf hin, dass eine Objektivierung nicht automatisch zu einer dramatischen Verbesserung der Reliabilität führt. Ergebnisse subjektiver Methoden können ebenso reliabel sein.^[33] Es gibt per se keine unrelia- blen oder reliablen Prüfungsmethoden. Die entscheidende Voraussetzung für reliable Ergebnisse ist das optimale Sampling über Bedingungen und Messungen (Stationen, Prüfer/-in, SP).^[27]

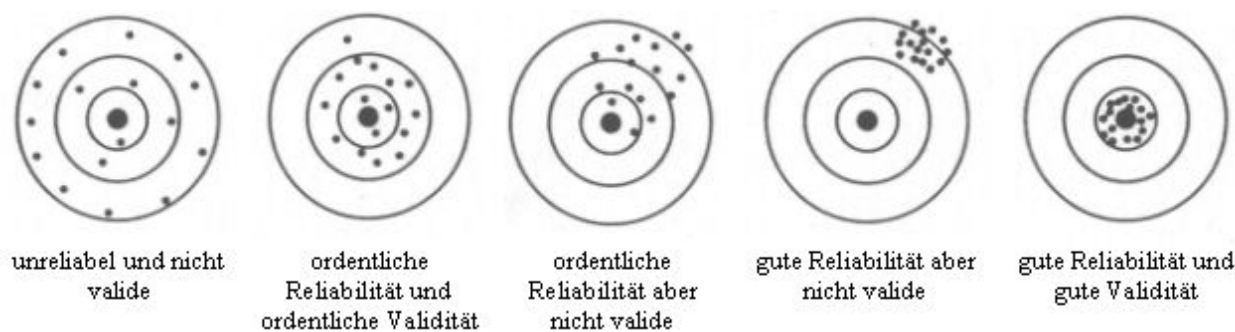


Abb. 6: Reliabilität und Validität ^[34]

Abb. 6 veranschaulicht den Zusammenhang zwischen Reliabilität und Validität. Das Zentrum der Zielscheibe repräsentiert den „wahren Wert“. Jeder Punkt steht für ein einzelnes Ergebnis des Testinstrumentes. Je reliabler ein Testinstrument, desto weniger streuen die Punkte. Je valider ein Testinstrument, desto genauer liegen die Punkte im Zentrum. Ein Punkt im Zentrum bedeutet, dass das erzielte Ergebnis mit dem „wahren Wert“ übereinstimmt.

1.3.2 Validität

Die Validität (Gültigkeit) gibt an, wie gut ein Test in der Lage ist, genau das zu messen, was er zu messen vorgibt. Je genauer eine Prüfung das misst, was sie zu messen beabsichtigt, desto valider ist die Prüfung. Es werden drei Hauptarten von Validität unterschieden:^[31] Die *Inhaltsvalidität* bezieht sich auf Inhalt und Format des Testinstrumentes. Sie ist gegeben, wenn der Inhalt der Test-Items das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend

erfasst. Zu beantwortende Fragen sind dabei: Repräsentiert die Itemauswahl adäquat den Inhalt, der geprüft werden soll? Ist das Format geeignet? *Kriteriumsvalidität* liegt vor, wenn das Ergebnis eines Tests zur Messung eines Konstruktes mit Messungen eines korrespondierenden Merkmals bzw. Kriteriums übereinstimmt. Sie ist definiert als Korrelation zwischen den Testwerten und den Kriteriumswerten einer Stichprobe. Bei Instrumenten, die das gleiche messen sollen, werden hohe Korrelationen erwartet und umgekehrt. Die *Konstruktvalidität* bezieht sich auf die Eigenschaften des psychologischen Konstruktes bzw. dessen Charakteristik, das vom Instrument gemessen wird. Wie gut erklärt das Konstrukt des Instrumentes Leistungsunterschiede? Als Voraussetzung gilt eine Beweiskette, die die Interpretation von Ergebnissen und Daten eindeutig mit einem Netzwerk aus Theorie, Hypothesen und logischen Zusammenhängen verbindet. Dieses Netzwerk unterstützt dann die gewünschten Interpretation angemessen oder lehnt sie ab. Die Konstruktvalidität wird in der medizinischen Ausbildung als wesentliche Validität angesehen, da fast alle Prüfungen in der medizinischen Ausbildung Konstrukten unterliegen, die von psychologischen und Bildungstheorien gebildet werden.^[35]

Die American Educational Research Association führt in ihren Standards für den Gültigkeitsbeweis ihrer Ergebnisinterpretation fünf unterschiedliche Quellen auf:^[36] 1. Inhalt, was wurde geprüft? – 2. Antwortprozess, wie gestaltet sich das Einlesen, die Auswertung und Qualitätskontrolle der Daten? – 3. Innerer Struktur, wie sind die statistischen Eigenheiten, z. B. die Reliabilität? – 4. Beziehung zu anderen Variablen, z. B. „Goldstandards“ – 5. Konsequenz des Ergebnisses, z. B. bestehen/nicht bestehen.

Boon und Stewart bemängeln in einem Review über Prüfungsinstrumente bezüglich der Kommunikationsfähigkeit, dass die Validität oft vernachlässigt wird und ein Großteil der beschriebenen Methoden lediglich bezüglich ihrer Reliabilität getestet wurde.^[37]

Auf die Besonderheiten der Überprüfung kommunikativer Fähigkeiten wird im folgenden Kapitel genauer eingegangen.

1.4 Prüfung kommunikativer Fähigkeiten

Warum sollen kommunikative Fähigkeiten von Studierenden in der medizinischen Ausbildung überprüft werden? Der große Stellenwert der kommunikativen Fähigkeiten der Ärztin/ des Arztes in der Arzt-Patienten-Interaktion ist bekannt und weithin akzeptiert. Kommunikative und soziale Kompetenzen gewinnen eine immer größer werdende Bedeutung in der derzeitigen Arbeitssituation der Ärztinnen/-e.^[38] Sie sind Voraussetzung für erfolgreiche Arzt-Patientenkontakte und führen zu verbesserten Behandlungsergebnissen und höherer Zufriedenheit sowohl bei Ärztinnen/-en als auch bei Patientinnen/-en.^[8] Dennoch wird der

Vermittlung kommunikativer Fähigkeiten im Rahmen der medizinischen Ausbildung oft nicht ausreichend Rechnung getragen. Bei der Implementierung neuer bzw. der Optimierung und Weiterentwicklung vorhandener Kommunikations-Kurse und Übungen ist die Überprüfung der Leistung der Studierenden ein wichtiger Bestandteil. Neben den in Kapitel 1.2 genannten Wirkungen von Prüfungen auf die Studierenden und deren Lernverhalten verdeutlicht die Überprüfung kommunikativer Fähigkeiten den Effekt der „Kommunikationskurse“ bzw. des auf Kommunikation und Interaktion ausgerichteten Curriculums. Die Evaluierung bzw. Prüfung von Kommunikationskursen steigert unter anderem deren Akzeptanz in der Fakultät. Sobald ein Bereich legitimiert wird, in dem er Teil des Evaluierungssystems wird, wird dieser eher als wichtiges Element der medizinischen Ausbildung anerkannt.^[8]

In der Literatur sind unterschiedlichste Instrumente zur Erfassung kommunikativer Fähigkeiten beschrieben. Sie unterscheiden sich vor allem bezüglich der Fragen, was geprüft wird, wie geprüft wird und wer die Leistung der Studierenden bewertet.

Was wird geprüft? Je nach Zielsetzung und Fragestellung einzelner Studien fokussieren unterschiedliche Methoden auf jeweils andere Teilbereiche der kommunikativen Kompetenzen der Studierenden. In einem Review von Anne Cushing zur Überprüfung nicht-kognitiver Faktoren professioneller Kompetenz werden diese Faktoren in folgende Bereiche aufgeteilt: Die *Anamneseerhebung* (history taking skills), wobei der Fokus vor allem auf den Inhalt der Fragen gerichtet ist. Die *Kommunikationsfähigkeit* (communication skills), die auf das Verhalten und die Technik hinzielt, um im Gespräch einen angemessenen Austausch und gegenseitiges Verstehen zu ermöglichen. Die *interpersonellen Fähigkeiten* (interpersonal skills), als Verhaltensweisen in Bezug auf Beziehung, Vertrauen und Akzeptanz im Interview. Und schließlich die *Einstellung* (attitudes) als emotionale und intellektuelle Voraussetzung für ein Gespräch.^[39]

Eine andere Einteilung favorisiert Suzanne Kurtz, die kommunikative Fähigkeiten (communication skills) einteilt in: 1. *Content skills* (Inhalt), die die Substanz der Frage und der erhaltenen Antwort betreffen: Wie angemessen und vollständig sind die gegebenen und erhaltenen Informationen? 2. *Process skills* (Prozess) beinhalten die Art und Weise, wie Fragen gestellt werden, die Interaktion strukturiert ist und eine Beziehung zum Patienten hergestellt wird. 3. *Perceptual skills* (Wahrnehmung) betreffen Problemlösung, Interpretation und Differentialdiagnosen ebenso wie den Umgang mit eigenen Emotionen und Einstellungen und denen der Patienten.^[8, 40]

Wie wird geprüft? Im Review-Artikel von Aspegren zum Lehren und Lernen kommunikativer Fähigkeiten in der Medizin werden bis zu zehn verschiedene Methoden genannt, mit denen kommunikative Fähigkeiten gemessen wurden.^[3] Dazu gehören Kursevaluationen, bei denen Kursteilnehmer ihre Meinung über den Nutzen des Trainings äußern, kognitive Tests des Wissens über medizinische Gespräche, Selbstbeurteilungen der Studierenden, direkte Observation durch externe Beobachter, Beurteilung von Videoaufnahmen von Interviews, OSCE-Prüfungen, Patientenbewertung der Studierenden-Leistung bis hin zum Gesundheitsergebnis der Patienten. Bei Performance-basierten Prüfungen klinischer Kompetenzen, in denen die praktische Umsetzung des theoretischen Wissens der Studierenden überprüft werden soll, ist der Einsatz von Simulationspatientinnen/-en weit verbreitet.^[17] Die Beurteilung der Interviews mit SP oder deren Videoaufnahmen erfolgt anhand von Checklisten oder auch Global Ratings (s. Kap. 1.4.1). Videoaufnahmen haben den Vorteil, dass bei unklaren Bewertungen das Interview nochmals betrachtet werden kann bzw. eine Bewertung durch mehrere Beurteiler unabhängig voneinander möglich ist. Die Studierenden haben außerdem die Möglichkeit, sich selbst zu betrachten und einzuschätzen.

Wer prüft? Häufig übernehmen Mitglieder der Fakultät (Hochschullehrer, Dozenten) die Prüferrolle, da deren Erfahrung als wichtige Voraussetzung für valide und reliable Prüfungen gesehen wird. In summativen Prüfungen werden jedoch aus juristischen Gründen primär Fakultätsmitglieder eingesetzt.^[28] Neben Fakultätsmitgliedern übernehmen in formativen Prüfungen aber auch Pflegende, Angehörige, Simulationspatienten, echte Patienten und Kommilitonen die Rolle des Beurteilers.^[41] Auch die Selbsteinschätzung der Studierenden zur Evaluierung kommunikativer Fähigkeiten wird genutzt.^[42, 43] Eine Arbeit, in der Studierende in eigenen Videoaufnahmen spezifische Verhaltensweisen untersuchten, um Prioritäten für eigenes Lernen zu entwickeln, konnte eine positive signifikante Korrelation zwischen Studierendenbewertung und Prüferbewertung nachgewiesen werden.^[44]

Einen wichtigen Stellenwert nimmt die Bewertung durch *Simulationspatientinnen/-en* ein. Sie bewerten die Leistung der Studierenden aus einer anderen Perspektive – der Perspektive der direkt Betroffenen. In zahlreichen Publikationen wird von Maßnahmen berichtet, SP auszubilden, detailliertes Feedback zu geben und die Leistung der Studierenden zu beurteilen.^[20, 45, 46, 47] Vu et al. weisen SP gute bis sehr gute Genauigkeit im Ausfüllen von Checklisten nach, wobei ein wesentlicher Faktor die Item-Anzahl der Checkliste ist. Je kürzer die Checkliste, desto genauer die Bewertungen. Weitere wichtige Faktoren waren Typ und klare Formulierung der Checklisten-Items.^[46] Wenn Fehler in der SP-Bewertung vorkommen, geschehen diese eher

zugunsten der Studierenden.^[48] Finlay et al. kommen zum Ergebnis, dass SP im Vergleich mit Experten zur Vergabe höherer Punktzahlen neigen.^[49] Dennoch korrelieren SP-Bewertungen signifikant mit Prüferbewertungen^[49, 50, 51] und es werden Reliabilitäten vergleichbar mit denen unabhängiger Prüfer erzielt.^[52, 53] Die Unterschiede in der Wahrnehmung (Patientenrolle oder unabhängiger Betrachter) führen jedoch dazu, dass unterschiedliche Aspekte der Kommunikation fokussiert werden.^[54] Cooper plädiert deshalb dafür, dass SP-Bewertungen Expertenbewertungen ergänzen, diese aber nicht ersetzen sollen.

Bei der Vielzahl der beschriebenen Instrumente zur Überprüfung kommunikativer Fähigkeiten sind neben Zielsetzung und Fragestellung die Kriterien Reliabilität (s. Kap. 1.3.1), Validität (s. Kap. 1.3.2), Integrationsfähigkeit in das Studium und die Abstimmung auf die Lehrformen neben Kosten und Zeitaufwand wesentliche Aspekte. Es sollte ein möglichst hohes Maß an Reliabilität und Validität bei noch vertretbarem Aufwand an Zeit und Geld erreicht werden.

Die anschließenden Kapitel beschreiben die Diskussion über Vor- und Nachteile der Anwendung von Checklisten oder Global Ratings bei der Bewertung von Interviews, die Problematik der Fallspezifität und die Einflüsse des Geschlechts der Studierenden auf kommunikative Fähigkeiten.

1.4.1 Detaillierte Checkliste versus Global Rating

Im Zusammenhang mit der Bewertung der Leistung von Studierenden in Interviews wurde in den neunziger Jahren die Anwendung detaillierter Checklisten oder Global Ratings mit Vor- und Nachteilen mehrfach diskutiert. Forschung und Entwicklung von Instrumenten zur Erfassung kommunikativer Fähigkeiten in der medizinischen Ausbildung führten zunächst zu sehr detaillierten Instrumenten wie beispielsweise der „Maastricht History-taking and Advice Checklist“ (MAAS)^[55] oder dem „Cambridge-Calgary-Observation-Guide“ (CCOG)^[81]. Sowohl die Ansicht, dass Lernende einen hohen Spezifizierungsgrad brauchen, um explizite Anforderungen im medizinischen Interview zu verinnerlichen als auch das Argument, dass detaillierte Checklisten die Objektivität erhöhen und Fehler durch Ungenauigkeit von Prüfern reduzieren, unterstützten diesen Trend.

In den neunziger Jahren wurden zunehmend wieder globalere Bewertungen (Global Rating) angewandt. Die Annahme, dass detaillierte Checklisten objektiver und bezüglich ihrer psychometrischen Eigenschaften den Global Ratings überlegen sind, wurde widerlegt.^[33, 56] Global Ratings sind zeitsparender beim Ausfüllen, benötigen jedoch erfahrene und gut trainierte Prüfer. Global Ratings führen zu größeren Schwankungen zwischen unterschiedlichen Prüfern in deren Bewertung. Die Reliabilität des gesamten Testes wird jedoch kaum beeinflusst, wenn viele

unterschiedliche Prüfer an unterschiedlichen Stationen prüfen. Der individuelle Prüferfehler wird über mehrere unterschiedliche Stationen ausgeglichen.^[57] Dies bestätigen mehrere Studien, bei denen Experten mit kürzeren Global Ratings über mehrere Stationen gleichwertige oder gar höhere Reliabilitäten erzielten als mit langen und detaillierten Checklisten.^[33, 58, 59, 60, 61, 62] Für Bewertungen durch Simulationspatienten wurden ähnliche Ergebnisse erzielt.^[47, 50, 63]

Untersuchungen zeigen auch, dass Global Ratings validere Ergebnisse erzielen können als detaillierte Checklisten.^[50, 61, 64] Checklisten scheinen eher Vollständigkeit als Kompetenz zu testen. Hodges beschreibt in einer Arbeit, dass Checklisten nicht unbedingt ein höheres Level an Erfahrung abbilden können. Allgemeinmediziner (family physicians) erzielten weniger Punkte als Studierende höherer Semester (clerks) in der Checklistenbewertung. Beim Global Rating war dies umgekehrt. Hodges begründete dies mit dem veränderten Fragemuster bei Allgemeinmedizinern, die sich auf das Wesentliche konzentrieren. Ihr Fragemuster entsprach nicht mehr der Urform, wie es im Studium vermittelt und in der Checkliste abgefragt wurde.^[64]

Durch die Detaillierung von Checklisten kann es zur Trivialisierung der Prüfung interpersoneller Fähigkeiten kommen, indem die Wahrnehmung der Patienten in der Gesamtheit der Begegnung negiert wird. Eine Atomisierung sollte deswegen vermieden werden. Diese kann zum Verlust der Validität der Ergebnisse führen, indem die Prüfung auf spezifisches Verhalten (z. B. Augenkontakt) eingeschränkt wird.^[27, 33] Das gemessene Phänomen wird inkomplett.

In den Niederlanden wurde außerdem ein verändertes Lernverhalten beobachtet, weswegen unter anderem von detaillierten Checklisten abgewichen wurde. Die Studierenden beschäftigten sich in der Vorbereitung auf Prüfungen nicht mehr mit der Kommunikation an sich, sondern konzentrierten sich darauf, einzelne Punkte des Bewertungsbogens zu üben.^[65]

Global Ratings scheinen in der Bewertung der Leistung von Studierenden im Interview Vorteile gegenüber Checklisten zu besitzen. Da Global Ratings anfälliger für Schwankungen bei unterschiedlichen Prüfern sind, ist eine gute Schulung der Prüfer und das Prüfen über mehrere Fälle notwendig. Neben den prüferabhängigen Schwankungen (*Interrater*-Reliabilität) hat vor allem die Fallspezifität (*Intercase*-Reliabilität) einen wesentlichen Einfluss auf die Erfassung kommunikativer Fähigkeiten.

1.4.2 Fallspezifität

Einige Autorinnen/-en untersuchten, inwieweit sich der fachlich medizinische Kontext als Störfaktor eliminieren lässt. Im Rahmen einer Studie von Hodges zur Erfassung kommunikativer Fähigkeiten im OSCE-Format wurden mehrere Kommunikationsstationen entwickelt. Zusätzlich wurde das Wissen der Studierenden über den klinischen Inhalt bzw. über das klinische Problem

erfasst. Es zeigte sich eine deutliche Korrelation zwischen den Ergebnissen der Überprüfung kommunikativer Fähigkeiten und den Ergebnissen der Überprüfung des inhaltlichen medizinischen Wissens.^[66] Neben der Tatsache, dass bei unbefriedigender Kommunikation weniger Inhalt in Erfahrung gebracht werden kann und ohne inhaltliches Wissen die Kommunikation problematisch wird, wird auch auf die Möglichkeit des Halo-Effekt hingewiesen. Der Halo-Effekt (Ausstrahlungseffekt) liegt vor, wenn beispielsweise eine Fähigkeit (medizinisch fachliche Kompetenz) die Beurteilung einer anderen Fähigkeit (kommunikative Kompetenz) beeinflusst.

Eine Studie von Thomson untersuchte die Korrelation der Ergebnisse kommunikativer Fähigkeiten an zwei unterschiedlichen Stationen. Die Prüfungskandidaten wurden dabei im Vorhinein über das medizinische Problem und Lösungsansätze informiert, um Einflüsse durch fehlendes fachliches Wissen zu verringern. Dennoch ergaben sich nur geringe Korrelationen zwischen den Stationsergebnissen.^[67] Einige Zeit wurde davon ausgegangen, dass allgemeine kommunikative Kompetenzen existieren, die erfasst werden könnten. Die Studie von Thomson zeigte jedoch, dass die Leistung der Studierenden von präzisen Details des Stations-Settings abhängen. Die *Intercase*-Reliabilität, die Zuverlässigkeit der Ergebnisse eines Prüflings über verschiedene Fälle hinweg, ist somit ein wichtiger limitierender Faktor bei der Messung kommunikativer Fähigkeiten.^[68] Die *Interrater*-Reliabilität ist ihr gegenüber weniger relevant, da bei genügend hoher Anzahl an geprüften Fällen die Anzahl an Prüferinnen/-n und Simulationspatientinnen/-en ausreicht, um reproduzierbare Ergebnisse zu erzielen.^[69] Voraussetzung ist, dass an jeder Station andere Prüfer/-innen prüfen. Van der Vleuten folgert, dass an einer Station nicht mehr als ein/-e Prüfer/-in benötigt wird. Bei einer größeren Anzahl zur Verfügung stehender Prüfer/-innen sollten diese eher dazu verwendet werden, die Fallzahl (Stationen) zu erhöhen.

Wie viele Fälle werden jedoch benötigt, um reproduzierbare Ergebnisse zu erzielen? In einigen Publikationen wird mit Hilfe der *Generalizability Theory*^[70] ein *Generalizability Coefficient* bestimmt, der die Reproduktion von Ergebnissen widerspiegelt. Als Gold-Standard gilt ein Wert von $g \geq 0,8$. In der Umkehrung kann mit Hilfe der *Generalizability Theory* die Zeit bzw. Anzahl von Fällen und Prüferinnen/-n berechnet werden, die benötigt werden, um einen Generalizability Coefficient von $g \geq 0,8$ zu erreichen. Schätzungen reichen von 4 bis 25 mit einem Konsensus von 12 bis 15 Stationen, um diesen Wert zu erzielen.^[39] Van der Vleuten und Swanson kommen in einem Review über die Prüfung klinischer Fertigkeiten mit Hilfe von SP zu einer mindestens benötigten Test-Zeit von 3 bis 4 Stunden.^[69] Van Thiel und Kollegen reichen mit einer neueren Version der „Maastricht History-Taking and Advice Checklist“, dem MAAS-R, 2 bis 2½

Stunden bei 8 bis 10 Fällen. Sie stellen außerdem fest, dass die Reliabilität sich bedeutend verbessert, wenn die Prüfung sich auf *basic interviewing skills* beschränkt.^[68] Mit einer weiterentwickelten Version des Messinstrumentes, dem MAAS-Global, bei dem die Checkliste in ein Global Rating umgewandelt wurde, reichten der gleichen Arbeitsgruppe sogar nur 2½ Konsultationen (ca. 40min) aus.^[58] Cohen et al. verglichen die Intercase-Reliabilität einer Checkliste mit der eines Global Rating zur Prüfung interpersoneller und kommunikativer Fähigkeiten. Die Checkliste benötigte 13 Fälle (4h 20min) und das Global Rating 8 Fälle (2h 40min).^[47]

Die Spannweite der Anzahl der benötigten Stationen, um eine ausreichende Intercase-Reliabilität zu erzielen, spiegelt die Unterschiede zwischen den Prüfungsdesigns wieder. Die benötigte Fallzahl ist jedoch niedriger, wenn weniger Kriterien im Prüfungsinstrument benutzt werden.^[39]

1.4.3 Einfluss des Geschlechts der Studierenden

Hat der Faktor Geschlecht einen Einfluss auf die kommunikativen Fähigkeiten von Studierenden? Marteau et al. untersuchten Faktoren, die die kommunikativen Fähigkeiten von Medizinstudenten im ersten Jahr beeinflussen. Es zeigte sich, dass das Geschlecht der Studierenden einen signifikanteren Vorhersagewert über das Level der kommunikativen Fähigkeiten besaß als die Teilnahme oder Nicht-Teilnahme an einem Kommunikationstraining. Frauen wurden bezüglich Empathie, Wärme und Kompetenz signifikant besser bewertet als Männer.^[71] Eine Arbeit von Furman, die den Effekt von Studierenden-Geschlecht und SP-Geschlecht im Simulationspatientengespräch untersuchte, konnte keine Interaktion zwischen Studierenden-Geschlecht und SP-Geschlecht feststellen. Es ergab sich jedoch ein signifikanter Haupteffekt für das Studierenden-Geschlecht bezüglich der Bewertung kommunikativer Fähigkeiten mit besseren Bewertungen bei Studentinnen.^[72] Ein ähnliches Ergebnis lieferte eine Arbeit von Chambers et al., die die Interaktion zwischen Studenten- und SP-Geschlecht beim CSA untersuchte.^[73] Aspegren kommt in seinem Review über Lehre und Lernen kommunikativer Fähigkeiten zu dem Schluss, dass Männer nach Kommunikations-Trainingskursen schlechtere Ergebnisse erzielen als Frauen und belegt dies mit vier weiteren Studien.^[3]

Diese Ergebnisse stehen im Einklang mit Publikationen, die die Einstellung (attitudes) von Studierenden bezüglich der Arzt-Patientenbeziehung betreffen. Batenburg und Small erfassten die professionellen Einstellungen von Medizinstudenten anhand von Fragebögen. Studentinnen zeigten hierbei mehr patienten-, psychosozial- und „care“-orientierte Einstellungen als Studenten.^[74] Rees und Sheard kamen bei der Befragung von 491 Studierenden mittels der

„Communications Skills Attitude Scale“ zum Ergebnis, dass Studierende mit positiven Einstellungen gegenüber kommunikativen Fähigkeiten eher weiblich sind.^[75] Hojat et al. befragten über 300 Medizinstudentinnen/-en anhand der „Jefferson Scale of Physician Empathy“ – ein Fragebogen, der entwickelt wurde, um die Empathie von Medizinstudentinnen/-en und Ärztinnen/-en in Bezug auf die Patientenversorgung zu untersuchen. Frauen erzielten bei dieser Umfrage signifikant höhere Punktzahlen als Männer.^[76]

Die zitierten Arbeiten beschreiben, dass Studentinnen offener gegenüber Kommunikationskursen zu sein scheinen als Studenten, was sich auch in der Leistungsbewertung einiger Studien niederschlägt.

1.5 Synopsis und Fragestellung

Fähigkeiten zu Kommunikation und Interaktion sind als explizites Ausbildungsziel des RSM in der Studienordnung aufgeführt. Um dieses Ziel zu erreichen, findet in 14-tägigem Rhythmus die Übung Interaktion als fester Bestandteil des Stundenplans der Studierenden statt. Die regelmäßig stattfindenden MCQ- und OSCE-Prüfungen beziehen sich jedoch vor allem auf kognitive und anwendungsbezogene Lernziele des RSM, während der kommunikative Bereich in den MCQ-Prüfungen bisher nicht und in OSCE-Prüfungen nur unsystematisch geprüft wurde. Um den Stellenwert kommunikativer Fähigkeiten im Curriculum zu verdeutlichen und eine Rückmeldung über die Effektivität des Kommunikationstrainings zu bekommen, ist eine Überprüfung der kommunikativen Fähigkeiten der Studierenden notwendig. Das Studierverhalten soll positiv beeinflusst und ein zusätzliches Feedback gegeben werden.

Die Ausbildungsforschung im medizinischen Bereich zeigt, dass eine Überprüfung kommunikativer Fähigkeiten möglich und sinnvoll ist. Eine große Anzahl an Methoden und Instrumenten wird in verschiedenen Publikationen beschrieben. Eine mehrfach angewandte Methode ist die Prüfung kommunikativer Fähigkeiten in Simulationsspatientengesprächen mit Hilfe von Checklisten bzw. Global Ratings.

Vor diesem Hintergrund wurde im Wintersemester 2001/2002 im Rahmen der in allen drei Semestern stattfindenden OSCE-Prüfungen ein spezielles Augenmerk auf die kommunikativen Fähigkeiten der Studierenden gerichtet. Einerseits sollte eine Bestandsaufnahme der kommunikativen Fähigkeiten der Studierenden des RSM erfolgen und andererseits mit den gewonnenen Erfahrungen dieser Studie die Entwicklung eines sinnvollen Konzeptes für die Integration des kommunikativen Bereiches in das Prüfungssystem vorangetrieben werden.

Ausgehend von der Annahme, dass die regelmäßig stattfindende Übung Interaktion zu einem Anstieg der kommunikativen Fähigkeiten der Studierenden im Verlauf des Studiums führt, stellt

sich bei der Erfassung kommunikativer Fähigkeiten in den drei Jahrgängen des RSM im Wintersemester 2001/2002 für diese Arbeit folgende zentrale Frage:

„Gibt es einen qualitativen Unterschied bezüglich der kommunikativen Fähigkeiten der Studierenden des Reformstudienganges Medizin in Abhängigkeit vom aktuellen Semester der Studierenden?“

Als Ausgangshypothese gilt, dass zwischen den Semestern ein signifikanter Unterschied bezüglich der kommunikativen Fähigkeiten der Studierenden besteht. In der Annahme vergleichbarer Populationen der einzelnen Semester und eines validen und reliablen Instrumentes müsste sich mit höherem Semester ein Zuwachs an kommunikativen Fähigkeiten ergeben.