

Aus der Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Entwicklung und Erprobung einer videobasierten Prüfung zur Messung ärztlicher
kommunikativer Kompetenz von Medizinstudierenden im Vergleich zu einer OCSE
sowie Betrachtung des Zusammenhangs der Prüfungsleistungen
mit Familienkommunikation, Alter und Geschlecht

Inauguraldissertation
zur Erlangung des Doktorgrades der
physiologischen Wissenschaften
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Dipl.-Psych. Stephanie Ludwig

Mainz, 2024

Wissenschaftlicher Vorstand:

Univ.-Prof. Dr. Philipp Drees

1. Gutachter:

2. Gutachter:

3. Gutachter:

Tag der Promotion:

10.11.2025

Nachnutzungslizenz:

InC-1.0

Inhaltsverzeichnis

Abkürzungsverzeichnis	I
Abbildungsverzeichnis	II
Tabellenverzeichnis	III
Hinweise	IV
1. Einleitung	1
2. Literaturdiskussion	3
2.1 Grundlegende Kommunikationsmodelle und die Bedeutung kommunikativer Kompetenz für das Fach Medizin	3
2.2 Praktizierte Formen von Lehre und Prüfung kommunikativer Kompetenzen	6
2.3 Prüfungsform Objective Structured Clinical Examination (OSCE)	8
2.4 Überprüfung von Fertigkeiten unter Bezugnahme auf Kompetenzmodelle	9
2.5 Vor- und Nachteile OSCE und schriftliche Prüfungen	12
2.6 Digitale Prüfungsformate	13
2.7 Zusammenhang OSCE und Key-Feature-Prüfung	17
2.8 Constructive Alignment und AGIL-Schema	18
2.9 Konzept videobasierte Prüfung und dazugehörige Hypothesen der vorliegenden Arbeit	19
2.10 Konstrukt Familienkommunikation und dazugehörige Hypothesen der vorliegenden Arbeit	22
2.11 BMBF-gefördertes Forschungsprojekt	27
3. Methoden	28
3.1 Studienteilnehmer- und ablauf	28
3.2 Entwicklung Video-based Single-Choice Examination (VSE)	29
3.3 OSCE	31
3.4 Evaluation	32
3.5 Entwicklung des Fragebogens zur Familienkommunikation	33
3.6 Statistische Auswertung	35
4. Ergebnisse	37
4.1 Itemcharakteristiken VSEs	37
4.2 Ergebnisse VSEs und OSCE	37
4.3 Ergebnisse Evaluation	41
4.4 Ergebnisse in Bezug auf Familienkommunikation, Alter und Geschlecht.....	42
5. Diskussion	47
5.1 Diskussion der Ergebnisse in Bezug auf VSEs und OSCE	47
5.2 Diskussion zu den Ergebnissen der Evaluation.....	51

5.3 Diskussion zu den Ergebnissen in Bezug auf Familienkommunikation, Alter und Geschlecht.....	52
5.4 Bewertung und Limitationen in Bezug auf Videos und das Prüfungsformat.....	55
5.5 Abschließende Betrachtung der Prüfungsformen	58
6. Zusammenfassung.....	60
7. Literaturverzeichnis	62
8. Anhang.....	70
8.1 Fragebogen Familienkommunikation	70
Danksagung.....	71
Lebenslauf	72

Abkürzungsverzeichnis

AGIL-Schema	Schema zur Erfassung der Qualität von Lehrveranstaltungen (A: Adaption, G: Goal Attainment, I: Integration, L: Latent Pattern Maintenance)
NURSE-Modell	Kommunikationsmodell zum Umgang mit den Gefühlen von Patienten (N: Naming, U: Understanding, R: Respect, S: Support, E: Exploration)
OSCE	Objective Structured Clinical Examination
OSVE	Objective Structured Video Exam
SPIKES-Modell	Kommunikationsmodell zur Übermittlung schlechter Nachrichten vom Behandler an den Patienten (S: Setting of the interaction, P: Patient's perception and preparation, I: Invitation and information need, K: Knowledge of the information, E: Empathy and exploration, S: Summary and strategic planning)
SS	Sommersemester
VSE	Video-based Single-Choice Examination
WS	Wintersemester

Abbildungsverzeichnis

Abb. 1: Vier Gruppen psychologischer Kommunikationsmodelle	3
Abb. 2: Pyramide der Kompetenzebenen	10
Abb. 3: Taxonomie Tabelle der Kompetenzebenen	11
Abb. 4: Theoretische Herleitung Prüfungskonzeption und OSCE-Vergleich.....	21
Abb. 5: Theorie der Familienkommunikation.....	23
Abb. 6: Videoausschnitt VSE.....	30

Tabellenverzeichnis

Tab. 1: Studienteilnehmer	28
Tab. 2: Itemcharakteristiken der VSE im SS18	38
Tab. 3: Itemcharakteristiken der VSE im WS18/19.....	39
Tab. 4: Hauptachsenanalysen der VSEs im SS18 und WS18/19	40
Tab. 5: Ergebnisse Evaluation	42
Tab. 6: Median und IQR der Items der Fragebögen zur Familienkommunikation im SS18 und WS18/19.....	43
Tab. 7: Hauptachsenanalysen der Fragebögen zur Familienkommunikation im SS18 und WS18/19.....	43
Tab. 8: Rangkorrelationskoeffizienten nach Spearman zwischen VSEs, OSCE und dem Fragebogen zur Familienkommunikation	45
Tab. 9: Ergebnisse des linearen gemischten Modells VSE SS18.....	46
Tab. 10: Ergebnisse des linearen gemischten Modells VSE WS18/19	46
Tab. 11: Ergebnisse des linearen gemischten Modells OSCE WS18/19	46

Hinweise

Aus Gründen der besseren Lesbarkeit wird in dieser Arbeit weitestgehend das generische Maskulinum verwendet. Die in dieser Arbeit verwendeten Personenbezeichnungen beziehen sich - sofern nicht anders kenntlich gemacht - auf alle Geschlechter.

Die Inhalte und Ergebnisse dieser Arbeit wurden teilweise publiziert:

Ludwig S, Behling L, Schmidt U, Fischbeck S. Development and testing of a summative video-based e-examination in relation to an OSCE for measuring communication-related factual and procedural knowledge of medical students. *GMS J Med Educ.* 2021;38(3):Doc70.

1. Einleitung

Der Kontakt zwischen Arzt und Patient stellt eine komplexe Interaktion dar, in der viele Aspekte eine Rolle spielen. Vordergründig relevant ist vor allem die fachliche Expertise des Arztes, die zunächst zur Konsultation durch den Patienten führt. Doch schon an diesem Punkt stellt sich die Frage, wie diese fachliche Expertise im Kontakt angewandt und vermittelt wird. Gelingt es dem Arzt beispielsweise eine Diagnose verständlich zu erklären? Werden Behandlungsoptionen dargelegt und wird gemeinsam mit dem Patienten eine Entscheidung getroffen? Zudem kann ein Gespräch abhängig von gezeigter Empathie und Authentizität sehr unterschiedlich verlaufen. Fühlt sich ein Patient verstanden und ernst genommen? Hat er den Eindruck, er kann Fragen stellen? Wird auf seine Ängste und Sorgen eingegangen (Bewusstsein und Mündigkeit des Patienten jeweils vorausgesetzt)?

Diese und weitere Fragen beschäftigen u.a. das Fach Medizinische Psychologie und Medizinische Soziologie in Forschung und Lehre. Der Stellenwert der in diesem Fach gelehrt und abgeprüften ärztlichen kommunikativen Kompetenz hat in den vergangenen Jahren immer weiter zugenommen. Maßgeblich dafür war in Deutschland u.a. die Änderung der Approbationsordnung für Ärzte. In dieser wurde festgelegt, dass die Medizinstudierenden „Kenntnisse und Fähigkeiten“ in der „ärztlichen Gesprächsführung“ nachweisen müssen (1). Auch der „Masterplan Medizinstudium 2020“ betont die Relevanz der Thematik Arzt-Patient-Kommunikation, die „maßgeblich die Arzt-Patienten-Beziehung, den Behandlungserfolg und das Wohlbefinden der Patientinnen und Patienten beeinflusst“ (2). Aber auch international nimmt der Bereich ärztliche kommunikative Kompetenz einen hohen Stellenwert ein und ist Teil der medizinischen Ausbildung (3, 4).

In der vorliegenden Arbeit wird die Entwicklung und Erprobung einer videobasierten Prüfung zur Messung dieser kommunikativen Kompetenz bei Medizinstudierenden vorgestellt und mit der Prüfungsform Objective Structured Clinical Examination (OSCE) verglichen. Bei der OSCE handelt es sich um eine Prüfungsform, in der die Lerninhalte in simulierten Arzt-Patienten-Gesprächen abgeprüft werden. Dieser Prüfungsform steht im Hochschulalltag in der Regel die klassische schriftliche Prüfung gegenüber. Beide Prüfungsformen weisen Vor- und Nachteile auf, besonders bedeutsam für diese Arbeit sind dabei der Blick auf die Höhe des praktischen Bezuges der Prüfungen, die Testgütekriterien Objektivität und Reliabilität sowie den Ressourcenaufwand. In der Literaturdiskussion soll darauf genauer eingegangen werden. Vorwegnehmend soll aber bereits herausgestellt werden, dass die Höhe des praktischen Bezuges bei einer OSCE augenscheinlich höher ausfällt als bei einer rein schriftlichen Prüfung und ein rein schriftliches Verfahren durch die Unabhängigkeit von

Ratern augenscheinlich objektiver ist. In dem Bestreben eine möglichst praktische, objektive und ressourcenschonende Prüfungsform zu entwickeln, entstand unter Bezugnahme der Möglichkeiten der wachsenden Digitalisierung die Idee einer Prüfungsform, die schriftliche Prüfungsfragen mit Videos zu Arzt-Patienten-Gesprächen verbindet. Soweit bekannt, wurde eine Prüfung in dieser Form erstmalig für den deutschsprachigen Raum erarbeitet. Die genaue Ausgestaltung, Umsetzung und ersten Erfahrungen mit dieser Prüfung sollen in der vorliegenden Arbeit nachfolgend vorgestellt werden. Auch der Vergleich zur OSCE wurde gezogen und wird dargelegt.

Zudem wurde für diese Arbeit die Fragestellung verfolgt, inwieweit die gemessene ärztliche kommunikative Kompetenz der Studierenden mit der Kommunikationsqualität in ihrer Ursprungsfamilie zusammenhängt. Können Studierende, die in ihrer Ursprungsfamilie beispielsweise eine offene Kommunikation kennengelernt haben und denen dort zugehört wurde, von diesen Erfahrungen im Arzt-Patienten-Kontakt profitieren? Schneiden diese Studierenden in den Prüfungen der ärztlichen kommunikativen Kompetenz besser ab als Studierende, die in ihrer Ursprungsfamilie weniger Offenheit erlebt haben und denen dort weniger zugehört wurde?

Auch für diesen Aspekt der vorliegenden Arbeit soll dargelegt werden, wie der Bereich Familienkommunikation zu definieren und zu messen ist und ob und in welcher Höhe sich ein Zusammenhang mit den Prüfungsergebnissen ergab. Zudem sollen Alter und Geschlecht als weitere mögliche Determinanten der Prüfungsleistungen neben der Familienkommunikation betrachtet werden.

2. Literaturdiskussion

2.1 Grundlegende Kommunikationsmodelle und die Bedeutung kommunikativer Kompetenz für das Fach Medizin

Viele Wissenschaftler haben versucht, den Begriff „Kommunikation“ zu definieren (5). Nach Rossmann (5) stammt die am häufigsten genutzte Definition von Maletzke (6), nach der Kommunikation die „Bedeutungsvermittlung zwischen Lebewesen“ ist. Rossmann zeigt auf, dass die Thematik Kommunikation in vielen Fachgebieten wie Kommunikationswissenschaft, Medienwissenschaft, Psychologie, Soziologie und weiteren Disziplinen eine hohe Relevanz besitzt und somit als interdisziplinäres Thema angesehen werden kann.

Röhner und Schütz (7) legen dar, dass es eine Vielzahl an Kommunikationsmodellen gibt, die je nach Fachrichtung und wissenschaftlicher Tradition Unterschiede aufweisen. Im Bereich psychologischer Kommunikationsmodelle konnten Krauss und Fussel (8) durch Literaturrecherche vier Gruppen ausmachen: „Encoder-/Decoder-Modelle“, „Intentionsorientierte Modelle“, „Perspektivübernahmemodelle“ und „Dialogmodelle“ (s. Abb. 1).

<p style="text-align: center;"><u>Encoder-/Decoder-Modelle</u> (Shannon und Weaver, 9, Schulz von Thun, 10)</p>	<p style="text-align: center;"><u>Intentionsorientierte Modelle</u> (Grice, 11)</p>
<p style="text-align: center;"><u>Perspektivübernahmemodelle</u> (Rogers, 12)</p>	<p style="text-align: center;"><u>Dialogmodelle</u> (Watzlawick et al., 13)</p>

Abbildung 1: Vier Gruppen psychologischer Kommunikationsmodelle (nach Krauss und Fussel, 8, aufgearbeitet von Röhner und Schütz, 7)

In die Gruppe „Encoder-/Decoder-Modelle“ fallen die sehr bekannten Kommunikationsmodelle von Shannon und Weaver (9) und von Schulz von Thun (10).

Das Modell von Shannon und Weaver ist ursprünglich vor einem technischen Hintergrund entstanden und beschreibt, wie durch einen Sender ein Signal/eine Nachricht an einen Empfänger übertragen und entschlüsselt wird. Bei diesem Prozess kann es zu Störungen kommen.

Beim Modell von Schulz von Thun wird menschliche Kommunikation anhand von vier Aspekten betrachtet, auf die sich eine Äußerung beziehen kann: den Sachinhalt (Sachinformation, die vermittelt werden soll), den Appell (Aufforderung etwas zu tun, zu denken oder zu fühlen), die Selbstoffenbarung (gewollte Selbstdarstellung und unbeabsichtigte Selbstenthüllung) und die Beziehung (Art oder Inhalt der Äußerung lässt Rückschlüsse auf die Beziehung zu). Auch hierbei kann es zu Missverständnissen kommen, wenn eine Nachricht beispielsweise vom Sagenden auf der Sachebene „gemeint“ war, aber vom Hörenden auf der Beziehungsebene verstanden wird.

Zur Kategorie „Intentionsorientierte Modelle“ gehören die Maxime der Kommunikation nach Grice (11). Hierbei steht das Kooperationsprinzip im Fokus. Danach muss ein gemeinsames Kommunikationsziel vorliegen, damit Kommunikation gelingen kann. Es werden vier Maxime umschrieben: Maxime der Quantität („sei präzise“), der Qualität („sei wahr“), der Relevanz („sei relevant“) und der Klarheit („sei klar“).

Zum Bereich der „Perspektivübernahmemodelle“ zählen die Regeln der gelingenden Kommunikation nach Rogers, die im Kontext seines Ansatzes der klientenzentrierten Gesprächspsychotherapie zu sehen sind (12). Dabei sind drei wesentliche Merkmale gelingender Kommunikation einführendes Verstehen, Echtheit und emotionale positive Wertschätzung. Für Rogers sind insbesondere das Verstehen und Einfühlen in sein Gegenüber bedeutsam und wesentlich für eine gute und seinem Gegenüber angepasste Kommunikation.

In der letzten Kategorie „Dialogmodelle“ befindet sich das Kommunikationsmodell nach Watzlawick et al. (13). Wesentlicher Teil dieses Modells sind 5 Axiome: „man kann nicht nicht kommunizieren“, „jede Kommunikation hat einen Inhalts- und einen Beziehungsaspekt“, „Kommunikationsabläufe unterliegen einer Struktur, in der die Gesprächspartner ständig aufeinander reagieren“, „Kommunikation bedient sich digitaler (Gesagtes) und analoger (z.B. Mimik und Gestik) Modalitäten“ und „Kommunikationsabläufe sind entweder symmetrisch (gleichwertige Partner) oder komplementär (sich ergänzende Partner)“. Dabei gilt, dass Kommunikation kein reiner Austausch von Informationen, sondern interessengetrieben ist.

Nach Röhner und Schütz (7) eint alle psychologischen Kommunikationsmodelle die Annahme, dass eine getätigte Aussage nicht unbedingt wie beabsichtigt vom Empfänger verstanden wird.

Der Bereich der Kommunikation zwischen Arzt und Patient kann als ein Teilgebiet des Themas Kommunikation gelten und hat vor allem für Mediziner und angrenzende Berufsgruppen eine hohe Bedeutung (14).

Wie in Kapitel 1 bereits dargestellt, zeigt sich eine Bedeutungszunahme der Arzt-Patient-Kommunikation im Fachbereich Medizin anhand von politischen Bestrebungen und Vorgaben zur Umsetzung entsprechender Inhalte in der Lehre (1, 2).

In etwa einem Drittel bis zu der Hälfte der Zeit besteht die ärztliche Tätigkeit nach einer Untersuchung von Kempf (15) darin, mit Patienten, Angehörigen und Kollegen zu sprechen. Nach Geisler (16) befindet sich ein Arzt in der Klinik 50 %, ein niedergelassener Arzt sogar 60 bis 80 % seiner Zeit in Patientengesprächen. Diese Zahlen verdeutlichen, welchen Raum Kommunikation im medizinischen Kontext einnimmt und wie wichtig eine gelingende Kommunikation ist.

In Untersuchungen konnte hierzu wiederholt gezeigt werden, dass eine gelungene Arzt-Patient-Kommunikation einen positiven Einfluss auf Patienten, aber auch auf Ärzte hat. Di Blasi et al. (17) konnten in ihrer Übersichtsarbeit beispielsweise feststellen, dass Behandler, die eine beruhigende und unterstützende Kommunikation praktizierten, in drei der betrachteten Studien einen signifikanten Einfluss auf den Behandlungserfolg ihrer Patienten hatten. In einer anderen Arbeit gaben Frauen, die an Brustkrebs erkrankt waren, an, dass die Fähigkeit des Arztes zuzuhören am hilfreichsten für sie in der Zeit der Diagnosestellung gewesen ist (18). Gleichzeitig zeigte sich, dass Frauen mit Brustkrebs, die die Kommunikation mit Ärzten als negativ empfanden, eine schlechtere Lebensqualität aufwiesen (19). In einer weiteren Arbeit konnte festgestellt werden, dass sich folgende positive Effekte durch gelungene Kommunikation ergeben: eine geringere wahrgenommene Belastung durch die Erkrankung und eine höhere Compliance bezüglich der Behandlung und Verhaltensänderungen auf Seite der Patienten, aber auch ein geringeres Stressempfinden und eine höhere Berufszufriedenheit bei den Behandlern auf der anderen Seite (20). Dies kann u.a. dadurch vermittelt werden, dass durch eine gute Kommunikation die Probleme des Patienten besser erfasst und auf psychologischer Ebene besser auf diese eingegangen werden kann, wodurch in der Regel die Zufriedenheit von Patient und Behandler steigt.

Es gilt zudem als sehr gut untersucht, dass kommunikative Kompetenzen erlernbar sind (20, 21). Kurtz et al. (21) beschreiben, dass kommunikative Kompetenz keine Persönlichkeitseigenschaft darstellt, sondern aus einer Reihe von zu erlernenden Fertigkeiten besteht.

2.2 Praktizierte Formen von Lehre und Prüfung kommunikativer Kompetenzen

Aspegren (22) legt in seiner Überblicksarbeit dar, dass sich kommunikative Fertigkeiten auf verschiedene Aspekte in einem Arzt-Patienten-Gespräch beziehen. Auch ein Artikel von Baribeau et al. (23) hebt die Vielschichtigkeit ärztlicher kommunikativer Kompetenz hervor, wodurch die Entwicklung eines Lehrplans und einer Prüfung der entsprechenden Fähigkeiten eine Herausforderung darstellt.

Im „Kursbuch ärztliche Kommunikation“ (24) wird auf einige der psychologischen Kommunikationsmodelle Bezug genommen und diese werden auf die Arzt-Patienten-Beziehung angewandt. Aus dem Modell von Shannon und Weaver (9) wird abgeleitet, dass das Gespräch zwischen Arzt und Patient u.a. durch eine deutliche Sprechweise, eine an den Empfänger angepasste Wortwahl und eine ungestörte Umgebung gekennzeichnet sein sollte. „Gesprächsförderer“ und „Gesprächsstörer“ werden definiert. Entsprechend dem Modell von Schulz von Thun (10) wird zudem erarbeitet, dass ein Arzt neben der Sachebene auch die anderen Ebenen der Kommunikation bedenken und je nach Kontext adressieren sollte. Es gibt hierzu auch praktische Übungsanleitungen wie „auf welchem ‚Ohr‘ [welcher Ebene] hören Sie die Aussage des Patienten?“ (24). In Anlehnung an Grice (11) wird zudem als wichtiges Prinzip ärztlicher kommunikativer Kompetenz erarbeitet, dass die vom Arzt vermittelten Inhalte relevant sein sollten (beispielsweise angepasst an einen Kollegen, an den Bildungsstand oder an das Vorwissen des Patienten). Auch Ansätze aus der Psychotherapie wie Empathie und Wertschätzung fließen in das Kurskonzept ein. Diese sollen vor allem dann zum Einsatz kommen, wenn Patienten beispielsweise emotional reagieren, sehr belastet sind oder im Erstkontakt/zum Beziehungsaufbau. Ergänzend zu den klassischen psychologischen Kommunikationsmodellen wird zudem ein Modell für die Arzt-Patienten-Beziehung vorgestellt. Dabei wird die Arzt-Patienten-Beziehung von der Persönlichkeit des Arztes, des Patienten, der Professionalität und Expertise des Arztes sowie von krankheitsbedingten Faktoren beeinflusst. Zusätzlich lässt sich die Arzt-Patienten-Beziehung in eine paternalistische (autoritär, Arzt fällt die Entscheidung), dienstleistungsorientierte (Arzt berät, Patient entscheidet) oder partnerschaftliche (gemeinsame Entscheidungsfindung unter der Expertise des Arztes) Beziehungsgestaltung

unterteilen. Aus dem Anspruch einer Beziehungsgestaltung im partnerschaftlichen Modell leiten sich Modelle wie die Partizipative Entscheidungsfindung (25) ab, um im Arzt-Patienten-Gespräch einer hohen Patientenorientierung gerecht werden zu können.

Das vorgestellte Kurskonzept ist beispielhaft zu sehen und nicht allgemein verbindlich. Die Lehre ärztlicher kommunikativer Kompetenz kann an den Hochschulen variieren, Schwerpunkte können unterschiedlich gesetzt, Modelle weggelassen oder ergänzt werden.

Bei Betrachtung der praktizierten Formen von Lehre und Prüfung ärztlicher kommunikativer Kompetenz über verschiedene Hochschulen hinweg zeigte sich in einer Übersichtsarbeit von Härtl et al. (26), dass an Hochschulen im deutschsprachigen Raum vorwiegend mit Lernzielkatalogen und Modellen gearbeitet wird. Bei den Lernzielkatalogen kommen u.a. selbst entwickelte Lernzielkataloge, das Baseler Consensus Statement (27) und der Nationale Kompetenzbasierte Lernzielkatalog Medizin (28) zum Einsatz. Bei den Modellen werden sehr häufig das NURSE-Modell (29) für den Umgang mit Emotionen und das SPIKES-Modell (30) zum Überbringen schwieriger Nachrichten gelehrt. Bei den Prüfungsformen stellt die praktische Prüfung OSCE die häufigste Form vor schriftlichen Prüfungen als zweithäufigste Form dar.

International ergibt sich hinsichtlich der Prüfungsformen ein ähnliches Bild. Auch in Großbritannien kommt neben schriftlichen Prüfungen vor allem die Prüfungsform OSCE zum Einsatz (31). Auch in anderen Ländern wie Australien, Kanada und den USA zeigt sich, dass neben schriftlichen Prüfungen die Prüfungsform OSCE Anwendung findet (32-34).

Im Fach Medizinische Psychologie und Medizinische Soziologie der Universitätsmedizin Mainz wurden folgende Lernziele für die Lehre und Überprüfung kommunikativer Kompetenz gesetzt: Kenntnis und Anwendung des biopsychosozialen Krankheitsmodells im Arzt-Patienten-Kontakt, Basisfertigkeiten erlernen in patientenorientierter ärztlicher Gesprächsführung in Bezug auf Anamnese, Compliance, Informationen verständlich vermitteln, Visite und Mitteilen einer Krebsdiagnose sowie Sensibilisierung gegenüber schwerkranken und sterbenden Patienten (35). Es werden sowohl theoretische Inhalte vermittelt als auch praktische Übungen absolviert. Teil der theoretischen Inhalte sind die bereits vorgenannten Modelle NURSE (29) und SPIKES (30), bei den praktischen Übungen wird mit Rollenspielen und Lehrvideos gearbeitet (35). Im Kursus finden sich u.a. die unter 2.1 vorgestellten psychologischen Kommunikationsmodelle nach Watzlawick et al. (13) und Rogers (12) wieder. Auch die Gesprächsförderer und -störer in Bezug zu Grice (11) sind ein Bestandteil der Lehrinhalte. Das Kurskonzept wurde von Fischbeck entwickelt (36).

2.3 Prüfungsform Objective Structured Clinical Examination (OSCE)

Die Objective Structured Clinical Examination (OSCE) ist wie unter 2.2 dargelegt, die häufigste Prüfungsform zur Messung ärztlicher kommunikativer Kompetenz. Sie wurde in ersten Zügen 1975 in Schottland entwickelt und durchgeführt (37) und erstmals 1979 vollständig vorgestellt und beschrieben (38). Von den Autoren, vordergründig von Ronald Harden, wurde der bis dahin vorliegende Mangel an einer Prüfungsform erkannt, die klinisch-praktische Fähigkeiten der Medizinstudierenden misst und dabei als objektiv und strukturiert gelten kann. Die Überprüfung klinisch-praktischer Fähigkeiten ist bis dahin, wenn überhaupt, über unstrukturierte, nicht objektivierte Aufgaben im Klinikalltag erfolgt, wobei es auch keine festgelegten Kriterien für die Einschätzung der gezeigten Leistungen gab. Harden und Kollegen entwickelten damals das Prinzip, den Studierenden standardisierte Aufgaben an verschiedenen Stationen zu stellen, die diese rotierend durchlaufen. Inhaltlich bezogen sich die Aufgaben u.a. auf die Durchführung einer Anamnese, die Durchführung einer körperlichen Untersuchung und die Interpretation von Laborergebnissen in Bezug auf den dargestellten Patientenfall. Der Einsatz von Checklisten und trainierten Simulationspatienten wurde zudem eingeführt. Teilweise kommen mittlerweile bei einer OSCE aber auch Stationen mit Multiple-Choice-Formaten zum Einsatz (39).

Seit der Einführung der OSCE findet diese in vielen Ländern weltweit in verschiedenen Disziplinen der Medizin und bei unterschiedlichen Graden des Ausbildungsstandes Anwendung (40). Mit Stand aus dem Jahr 2016 wurden mehr als 1600 Artikel zur OSCE veröffentlicht (40). Harden selbst resümiert in diesem Artikel, dass die Prinzipien der OSCE bis heute aktuell sind.

Die meisten durchgeführten OSCEs beziehen sich inhaltlich nicht primär auf kommunikative Kompetenzen, sondern diese sind eher als zusätzliche Aspekte bei der Prüfung der Lerninhalte anderer Disziplinen im Medizinstudium zu sehen (35, 41). Auch aus der Übersichtsarbeit von Härtl et al. (26) geht hervor, dass die OSCE im deutschsprachigen Raum zwar die häufigste Form bei der Überprüfung ärztlicher kommunikativer Kompetenz ist, allerdings nur an wenigen Standorten einzelne Stationen komplett auf diese Thematik bezogen sind. Dennoch wurden auch OSCEs speziell für die Überprüfung ärztlicher kommunikativer Kompetenz entwickelt (35, 41).

Für den Standort Mainz wurde angelehnt an die etablierten Konzepte von Harden und Gleeson sowie Jünger und Nikendei (38, 42) eine OSCE für die Überprüfung kommunikativer Kompetenzen im vorklinischen Abschnitt des Studiums erarbeitet und erprobt (35). Die OSCE erreichte eine interne Konsistenz von Cronbachs $\alpha = .50$ ($n = 181$). Bei den

Studierenden kam die Prüfung sehr gut an, mehr als 90 % der Studierenden sprachen sich für eine Beibehaltung der OSCE aus.

Das bestehende Konzept wurde für diese Arbeit verwendet und bestehende OSCE-Prüfungsaufgaben kamen zum Einsatz (vgl. 3.3).

Da die OSCE als häufig genutzte Prüfungsform bereits zahlreich untersucht wurde, gibt es auch eine Vielzahl an Arbeiten zur Frage nach der Testgüte (wie Objektivität, Reliabilität, Validität). Blume und Kollegen (43) kamen zu dem Schluss, dass die OSCE unter Einhaltung bestimmter Voraussetzungen „im Allgemeinen“ ausreichende Ergebnisse erzielt. Auch eine Arbeit, die sich auf eine umfangreiche Literatursuche zu dieser Fragestellung für den Zeitraum 1975 bis 2003 bezieht (39), kommt zu einem entsprechenden Ergebnis. Demnach können die Testgütekriterien bei sorgfältiger Planung und Durchführung erreicht werden. Zu den Voraussetzungen zählt dabei u.a., dass die Reliabilität (interne Konsistenz) und Inhaltsvalidität einer OSCE maßgeblich von der Anzahl der Stationen bestimmt wird (39).

Bei der internen Konsistenz wird im Rahmen der Testkonstruktion davon ausgegangen, dass die Items dasselbe Merkmal messen (44). Das am meisten verwendete Maß zur Ermittlung der internen Konsistenz ist Cronbachs Alpha. Je höher die Korrelationen zwischen den Items im Durchschnitt sind, desto höher ist die interne Konsistenz des Tests. Die Höhe von Cronbachs Alpha hängt u.a. von der Anzahl der Items ab (44, 45). Enthält ein Test viele Items kann in der Regel eine höhere interne Konsistenz erzielt werden.

Um eine möglichst hohe Reliabilität und Inhaltsvalidität zu erzielen, sind demnach in der Regel eine Vielzahl an OSCE-Stationen notwendig. Es zeigte sich zudem, dass die Interrater-Reliabilität und damit die Objektivität in einigen Fällen als unzureichend einzustufen ist (39). Auch hier wird in der Arbeit darauf verwiesen, dass sich diese Problematik mit einer hohen Anzahl an Stationen ausgleichen lässt.

Für den Zusammenhang zwischen Objektivität und Reliabilität gilt, dass die Objektivität (Standardisierung der Testsituation, -durchführung und -auswertung) eine Voraussetzung für die Reliabilität darstellt (44). Eine hohe Reliabilität kann also dann erzeugt werden, wenn der Test hinsichtlich Testsituation, Durchführung und Auswertung möglichst objektiv gestaltet wird und abläuft.

2.4 Überprüfung von Fertigkeiten unter Bezugnahme auf Kompetenzmodelle

Miller stellte heraus, dass die Komplexität der Leistung eines Arztes nicht durch eine einzelne, einfache Messmethode erfasst werden kann (46). Seiner Arbeit nach lassen sich klinische Fertigkeiten auf verschiedenen Ebenen betrachten. Er unterscheidet Fertigkeiten

auf kognitiver Ebene von Fertigkeiten auf verhaltensbezogener Ebene. Zur kognitiven Ebene zählen Faktenwissen („knows“) und Handlungswissen („knows how“), zur verhaltensbezogenen Ebene gehören praktisch gezeigte Leistungen unter „Laborbedingungen“ („shows how“) und Leistungen am „realen“ Patienten („does“). Dieser Systematik folgend kann die OSCE als Prüfung auf verhaltensbezogener Ebene im Bereich „shows how“ eingeordnet werden. Schriftliche Prüfungen bewegen sich hingegen im Bereich der kognitiven Ebene mit der Abfragung von Faktenwissen und Handlungswissen („knows“ und „knows how“).

Das Modell nach Miller arbeitet mit einem hierarchischen oder gestuften Ansatz. Das Maß an Praxisbezug steigt mit jeder Stufe oder Ebene an (s. Abb. 2).

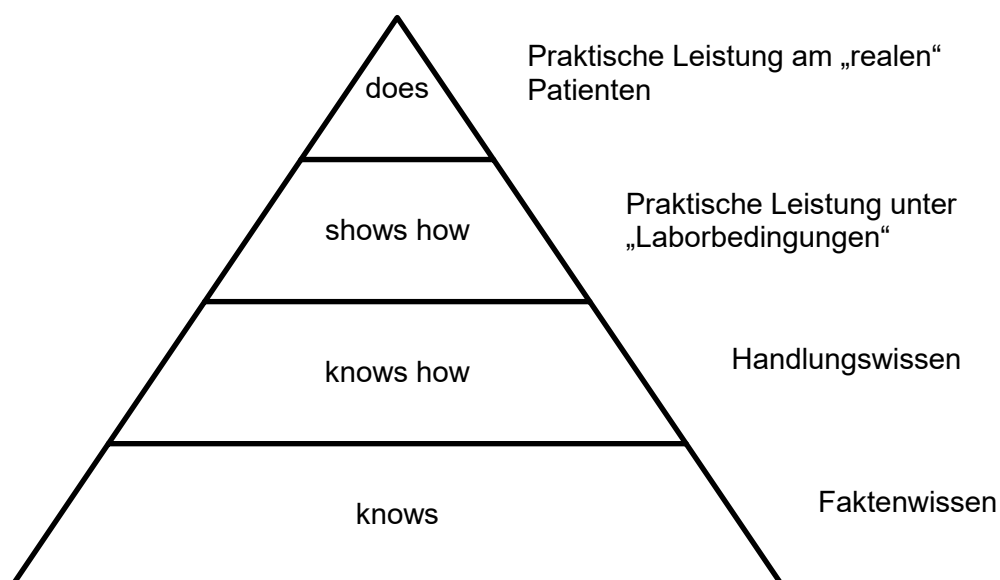


Abbildung 2: *Pyramide der Kompetenzebenen (nach Miller, 46)*

Fakten- und Handlungswissen können u.a. über Single/Multiple-Choice Fragen, Key-Feature Konzepte (Treffen kritischer Entscheidungen) und Situational Judgement Tests abgeprüft werden (47-50).

Bei einem Situational Judgement Test kommen schriftliche oder videobasierte Szenarien zum Einsatz, die zu bewältigende, typische Situationen im Berufsalltag abbilden. Die Aufgabe besteht darin, sich anhand des dargestellten Szenarios für eine darauffolgende Handlungsoption zu entscheiden. Die Antwortform kann variieren.

Untersuchungen ergaben zumeist gute interne Konsistenzen (47).

Im Kontext medizinischer Ausbildung wurden Situational Judgement Tests bereits angewandt (49).

In einer Arbeit aus dem Jahr 2019 legen Witheridge et al. (51) dar, dass die Einteilung der Kompetenzebenen nach Miller immer noch aktuell ist. Sie weisen darauf hin, dass bei Prüfungsvorhaben auch weiterhin die Einschätzung, welche WissensEbene mit den Prüfungsinhalten abgefragt wird, wichtig ist. Zudem soll kritisch hinterfragt werden, ob die Wahl des Prüfungstools für diese WissensEbene geeignet ist. Die Autoren sehen es beispielsweise kritisch, dass in OSCEs zum Teil auch kognitive Wissensbereiche integriert werden, für die sie das Tool OSCE entsprechend der Einteilung nach Miller nicht als geeignete Prüfungsform ansehen. Sie schlagen hier entsprechende Modifikationen der Aufgabenstellungen vor, um dem Anspruch, das jeweils passende Tool für die dazugehörige WissensEbene einzusetzen, auch gerecht werden zu können.

Auch andere Kompetenzmodelle arbeiten mit verschiedenen Ebenen oder Stufen von Kompetenzen. Ein weiteres, sehr verbreitetes Modell stammt von Bloom (52), meist in der modifizierten Form von Krathwohl (53) angewandt.

	<u>Kognitiver Prozess</u>					
	Erinnern	Verstehen	Anwenden	Analysieren	Evaluieren	Erstellen
<u>Wissensebene</u>						
A Faktisches Wissen						
B Konzeptionelles Wissen						
C Prozedurales Wissen						
D Metakognitives Wissen						

Abbildung 3: *Taxonomie Tabelle der Kompetenzebenen nach Bloom (52), modifiziert nach Krathwohl (53)*

In der ursprünglichen Form beschrieb Bloom sechs Kategorien kognitiver Kompetenzen, die von einfach zu komplex und von konkret zu abstrakt kumulativ hierarchisch aufgebaut sind. Dies bedeutet, die Kompetenzen auf weiter unten liegenden Ebenen bedingen den Erfolg auf höheren Ebenen. Die Kategorien sind Wissen, Verständnis, Anwendung, Analyse, Synthese und Evaluation. In der Modifikation von Krathwohl erfolgt, als eine der Anpassungen, auf der WissensEbene eine Unterteilung in faktisches Wissen, konzeptionelles Wissen, prozedurales

Wissen/Handlungswissen und metakognitives Wissen. Zudem erfolgte eine Überarbeitung der kognitiven Komponenten.

In Abbildung 3 ist dargestellt, dass die Leistungsinhalte sowohl auf der Achse der Wissensebene als auch auf der Achse des kognitiven Prozesses eingeordnet werden können. So ergeben sich zahlreiche Kombinationen und Zuordnungen. Krathwohl beschreibt, dass seine Taxonomie Lehrende und Prüfende bei der Einschätzung unterstützen kann, welche Ebenen sie ansprechen und prüfen wollen, um Lehre und Prüfung entsprechend zu planen.

2.5 Vor- und Nachteile OSCE und schriftliche Prüfungen

Wie bereits dargelegt besteht einer der großen Vorteile der Prüfungsform OSCE darin, dass sie einen vergleichsweise hohen Praxisbezug realisiert und die Leistung auf direkter Verhaltensebene betrachtet (46). Gleichzeitig bedeutet die Durchführung einer OSCE allerdings in der Regel auch einen hohen Zeitaufwand und erfordert erhebliche finanzielle (39) und personelle Ressourcen (u.a. hinsichtlich der Anzahl an benötigten Prüfern, dem Einsatz von Simulationspatienten, nur wenige Studierende können die Prüfung gleichzeitig ablegen). Dabei gilt es auch zu beachten, dass die Reliabilität und Inhaltsvalidität einer OSCE maßgeblich von der Anzahl der Stationen abhängt (39). Soll eine möglichst hohe Reliabilität und Inhaltsvalidität erzielt werden, sind demnach in der Regel eine Vielzahl an Stationen notwendig, wodurch Aufwand und Kosten nochmals steigen. Zudem stellt sich bei der Anwendung einer OSCE die Frage nach der Objektivität der Bewertung, die in der Regel durch verschiedene Prüfer vorgenommen wird. In Untersuchungen zeigte sich, dass die Interrater-Reliabilität und damit die Objektivität in einigen Fällen unzureichend ausfällt (39).

Schriftliche Prüfungsformen beziehen sich im Gegensatz zur OSCE auf die kognitive Ebene der Leistung (46), erzielen also augenscheinlich einen geringeren Praxisbezug. Theoretischen Überlegungen folgend sollte zudem die Objektivität einer schriftlichen Prüfung insbesondere im Falle derselben Items, Antwortmöglichkeiten und einer standardisierten, ggf. sogar computergestützten Auswertung einen höheren Grad erreichen als eine OSCE. Des Weiteren erscheint es deutlich ressourcenschonender, wenn im Rahmen einer schriftlichen Prüfung eine hohe Anzahl Studierender zur selben Zeit durch eine deutlich geringere Anzahl an Prüfern und ohne Einsatz zusätzlicher Simulationspatienten geprüft werden kann.

Die dargelegten Vor- und Nachteile der beiden häufigsten Prüfungsformen ärztlicher kommunikativer Kompetenz führten zu der Frage nach alternativen Prüfungsformen, die sowohl der Relevanz eines hohen Praxisbezuges als auch den Anforderungen an ein ressourcenschonendes Vorgehen und eine hohe Testgüte Rechnung tragen.

2.6 Digitale Prüfungsformate

Nach Michel et al. (54) kommen digitale Prüfungsformate an deutschen Hochschulen seit Mitte der 1990er Jahre zum Einsatz. Die Autoren legen dar, dass entsprechende Konzepte auf Einzelinitiativen oder einzelnen Förderprojekten beruhen, Standards oder übergreifende Ansätze existieren ihnen zufolge bislang nicht.

Bei bestehensrelevanten Prüfungen wird nach Michel et al. in der Regel das Konzept einer klassisch schriftlichen Prüfung auf die digitale Prüfung übertragen, wobei die Abfrage der Prüfungsinhalte über PCs oder Tablets/Notebooks statt über „Papier und Stift“ erfolgt.

Hinsichtlich des Prüfungskontextes legen sie dar, dass sich digitale Prüfungsformate sehr gut im Bereich naturwissenschaftlicher, mathematischer und technischer Fachdisziplinen aufgrund ihrer meist eindeutigen Auswertbarkeit eignen. Zudem können sie auch in geistes-, sozial- und gesellschaftswissenschaftlichen Fächern sowie in künstlerischen Disziplinen Anwendung finden, hierbei ist die Auswertung durch individuellere Formen der Lösung von Prüfungsaufgaben mit freien Texten, Zeichnungen oder auch Videos allerdings komplexer und herausfordernder.

Zudem beschreiben die Autoren, dass digitale Formate eine gute Möglichkeit darstellen, eine hohe Anzahl an Prüfungen mit möglichst geringem Zeitaufwand und automatisierter Auswertung effizient gestalten zu können.

Auf die Einbettung von Videos in digitale Prüfungsformate gehen Michel et al. in ihrer Übersichtsarbeit nicht genauer ein. Es erscheint allerdings plausibel, dass sich je nach Fachdisziplin und Prüfungsinhalt der Einsatz von Videos im Zuge digitaler Prüfungen anbietet und eignet. Zu dieser Einschätzung kommen auch Ehlers et al. (55). Sie stellen ebenfalls dar, dass es sich bei digitalen Prüfungen in der Regel lediglich um eine Übertragung klassisch schriftlicher Prüfungsformate handelt, jedoch weitaus mehr Gestaltungsmöglichkeiten bei digitalen Formaten durch den Einsatz von Bild- und Tonmaterial sowie fallbasiertem Prüfen gegeben sind.

Unter anderem im Bereich kommunikative Kompetenz erscheint es naheliegend, digitale Formate mit Videos zu nutzen. Gesprächsführungstechniken, Gesprächsabläufe, Aspekte wie Mimik, Gestik und Tonfall können dadurch sehr gut abgebildet und praxisnah gezeigt werden.

Baribeau et al. (23) beschreiben in ihrer Arbeit, dass mit Hilfe von Videos die Möglichkeit besteht, die Fähigkeit zum Erkennen, Verstehen und Kritisieren kommunikativer Kompetenzen messen zu können.

Mayer (56) erarbeitete auf Grundlage empirischer Befunde und Erkenntnisse aus der Forschung im Bereich Gedächtnis und allgemeiner Lerntheorien eine kognitive Theorie des multimedialen Lernens. Danach werden einerseits Text- und Bild-/Videoinformationen verknüpft und diese wiederum mit gespeichertem Wissen in Verbindung gebracht. Die Präsentation der Lehrinhalte in verbaler und visueller Form unterstützt nach Mayer die Erstellung eines kohärenten mentalen Modells.

Trost et al. (57) legen dar, dass eine Videodarstellung den Arzt-Patienten-Kontakt mit einem höheren Realitätsbezug abbilden kann als ein schriftliches Szenario. Dabei beschreiben sie, dass verbale und non-verbale Informationen so in die Prüfung eingeschlossen werden können. Zudem ordnen sie ein, dass über Prüfungen mit Videodarstellungen und den Einsatz von Multiple-Choice-Antwortformaten Faktenwissen und Handlungswissen nach Miller (46) abgefragt werden kann. Auch Rottmann et al. (58) schätzen ein, dass bei computergestützten Prüfungen die Einbettung von Videos eine höhere plastische und realistische Darstellung von Inhalten zur Erzielung einer hohen Handlungsorientierung ermöglicht. Eine hohe Handlungsorientierung zeichnet sich für sie durch einen hohen Bezug zum späteren beruflichen Vorgehen und Alltag aus.

Entsprechend diesen Darlegungen melden die Studierenden im Kursus an der Universitätsmedizin Mainz zurück, dass in Videos dargestellte Arzt-Patienten-Kontakte besser geeignet sind, die Lerninhalte praxisnah zu erleben, als der Einsatz rein schriftlich dargelegter Patientenfälle und Arzt-Patienten-Dialoge. Durch den Einsatz von Videos kann es zudem vermutlich leichter gelingen, sich mit der Arztrolle zu identifizieren und am direkten Modell zu lernen (vgl. sozialkognitive Lerntheorie, Modelllernen, Bandura, 59, und Findeisen et al., 60).

Von einem höheren Praxisbezug durch den Einsatz von Videos im Vergleich zu rein schriftlichen Prüfungen kann vermutlich auch dann ausgegangen werden, wenn bei den schriftlichen Prüfungen Arzt-Patienten-Dialoge und Kasuistiken genutzt werden.

Da Videos in Prüfungssoftware integriert werden können, scheint sich durch die Verbindung schriftlicher Prüfungen mit Videos eine Möglichkeit zu ergeben, die jeweiligen Vorteile praktischer und schriftlicher Prüfungen zu nutzen und die jeweiligen Nachteile zu minimieren oder zu vermeiden. Dies bedeutet, ein höherer Praxisbezug wäre durch die Videos gegeben und gleichzeitig kann die digitale Prüfung durch den Einsatz von standardisierten Aufgaben und vorgegebenen Antwortmöglichkeiten ressourcenschonend mit wenig Zeit- und Prüferaufwand, der Prüfung hoher Studierendenzahlen gleichzeitig und mit vermutlich höherer Testgüte die Lerninhalte abprüfen.

Für den deutschsprachigen Raum konnten 2017, als mit der vorliegenden Arbeit begonnen wurde, videogestützte Übungen zum Erlernen (61) kommunikativer Kompetenzen recherchiert werden. In Bezug auf Prüfungen fand sich lediglich in der Übersichtsarbeit von Härtl et al. (26) eine wenig ausführliche Nennung, dass an zwei Universitäten im deutschsprachigen Raum Videos in Prüfungen genutzt wurden.

International ergab die Recherche hingegen mehrere Studien, die sich mit der Entwicklung und Erprobung videobasierter Prüfungen zur Messung ärztlicher kommunikativer Kompetenz auseinandersetzten (23, 62, 63).

In den genannten Studien wurde mit einer als OSVE (Objective Structured Video Exam) bezeichneten Prüfung gearbeitet.

Bei Humphris und Kaney (62) wurden Videos produziert, in denen ein Arzt-Patienten-Gespräch dargestellt wurde, wobei ein Elternteil nach Diätmöglichkeiten für sein Kind fragt. Den Studierenden wurden die Videos präsentiert und sie sollten die kommunikativen Fertigkeiten benennen und beschreiben, die sie im Video erkennen konnten. Zudem sollten sie in einem zweiten Aufgabenteil die Konsequenzen der jeweiligen gezeigten Fertigkeiten diskutieren und in einem dritten Teil Alternativen nennen und begründen. Die Auswertung erfolgte durch mehrere Prüfer (Ratingverfahren schriftlicher Antworten), wobei eine Intra-Klassen-Korrelation von $r = .94$ ermittelt wurde. Es ergab sich zudem eine signifikante Korrelation von $r = .17$ zwischen OSVE und OSCE.

Bei Baribeau et al. (23) wurden den Studierenden bei der OSVE ebenfalls Videos gezeigt und zudem sollte ein Transkript der Videos in offener schriftlicher Form kommentiert werden (Welche kommunikativen Fertigkeiten wurden erkannt? Welche wurden nicht gezeigt oder was hätte der Arzt besser machen können?). Auch hier schätzten unterschiedliche Prüfer die

Leistungen ein, wobei eine gewisse Anzahl an Prüfern mindestens zum gleichen Urteil gelangen musste, damit eine positive Wertung vorgenommen wurde.

Im späteren Verlauf der Arbeit an diesem Projekt wurde zudem ein Artikel veröffentlicht, der sich mit der Einbettung von Videodarstellungen im Kontext psychiatrischer Prüfungen im deutschsprachigen Raum in einer Pilotstudie befasst (57), auf den bereits weiter oben unter 2.6 eingegangen wurde. In dieser Arbeit wurden Medizinstudierenden Videosequenzen zu psychopathologischen Fragestellungen im Fach Psychiatrie gezeigt. Als Antwortformat wurde dabei Multiple-Choice gewählt. Die Testgütekriterien der Prüfungsfragen bewerteten die Autoren als heterogen. Von den Studierenden wurde diese Prüfungsform als geeignete Vorbereitung für den späteren Berufsalltag eingeschätzt. Die Autoren sehen in ihrem Konzept ein Prüfungstool, das vielversprechend ist und kommunikative Kompetenzen in ihrer Vielschichtigkeit gut abprüfen kann. Kritisch sehen sie den Mehraufwand im Vergleich zu rein schriftlichen Prüfungen (Videoproduktion, Kostenaufwand).

Zudem veröffentlichte Ludwig (64) 2021 eine Arbeit, die die Entwicklung einer videogestützten Key-Feature-Klausur in Bezug auf die Thematik der klinischen Entscheidungsfindung (bezogen auf Diagnostik und Therapie) im Medizinstudium darstellt. In der Arbeit sollte untersucht werden, ob ein Unterschied im Erinnern hinsichtlich textbasierter und videobasierter Key-Feature-Lehrtools und darauffolgender entsprechender Prüfungen besteht und ob die Übereinstimmung der Prüfungsform in der Lern- und Prüfungsphase einen Einfluss auf die Erinnerungsfähigkeit hat. Bei einem Key-Feature handelt es sich um ein Szenario, bei dem der zu Prüfende vor eine kritische Entscheidung gestellt wird, die er treffen muss, um ein Problem zu lösen. Es zeigte sich, dass videogestützte Key-Feature-Fälle kurz- und tendenziell auch mittelfristig eine bessere Erinnerung an die Lehrinhalte erzeugen als textbasierte Fälle. Ein Unterschied in der Erinnerungsleistung hinsichtlich verschiedenen und identischen Lehr- und Prüfungsformaten konnte nicht festgestellt werden.

Der Einsatz von Videos wird in dieser Arbeit als Chance, aber mit Blick auf den Lernerfolg auch kritisch betrachtet. Problematisch könnte es sein, dass Videos für den Sachverhalt zu viele unwesentliche Informationen enthalten, die den Lernerfolg behindern könnten. Es wird diskutiert, dass Videos ggf. reduziert gestaltet und produziert werden sollen. Gleichzeitig könnte hierbei die Realitätsnähe leiden. Der Vorteil einer hohen Realitätsnähe wird in der Arbeit auch nochmal herausgestellt und von den Studierenden entsprechend eingeschätzt. Die Mehrheit der Studierenden gab an, dass die Videos realistisch waren und dass ihnen die Videofälle besser gefallen haben als die Textfälle.

Des Weiteren befasst sich die Arbeit auch mit der Fragestellung, welche Kompetenzebene mit welchem Prüfungsformat abgebildet werden kann. Zur Abfrage der Thematik klinische Entscheidungsfindung wurde das Key-Feature-Format zur Prüfung prozeduralen Wissens (Handlungswissen) eingesetzt.

Die Darlegung bei Ludwig, wonach sich zu viele Informationen in einem Video als überladend und überlastend herausstellen können, scheint der vorgestellten Theorie nach Mayer (56, siehe weiter oben unter 2.6) zu widersprechen. Die Diskussion bei Ludwig bezieht sich auf die Theorie der kognitiven Belastung beim Lernen (Chandler und Sweller, 65), wonach die Kapazität des Arbeitsgedächtnisses begrenzt ist. Kok (66) arbeitet diese Diskrepanz in ihrer Arbeit auf und legt unter Bezugnahme verschiedener gesichteter Literatur u.a. dar, dass ein wichtiger vermittelnder Aspekt darin besteht, dass die verbalen und visuell präsentierten Inhalte zueinander passend sein müssen, um einer möglichen Überlastung entgegenwirken zu können (u.a. Mousavi et al., 67).

2.7 Zusammenhang OSCE und Key-Feature-Prüfung

Neben den Ergebnissen zum Zusammenhang von OSCE und OSVE durch Humphris und Kaney (62) liegen noch Ergebnisse einer Untersuchung zum Zusammenhang von OSCE und einer Key-Feature-Prüfung vor (68). Die Fragestellung dieser Untersuchung lautete, inwieweit die Ergebnisse der Studierenden in der OSCE mit den Ergebnissen in einer Key-Feature-Prüfung zusammenhängen. Die Fragestellung entwickelte sich vor dem Hintergrund, dass Fischbeck et al. die OSCE als sehr kostenintensiv einschätzten. In der betrachteten Untersuchung wurden Medizinstudierenden schriftlich Fälle vorgelegt, wobei sie in einem offenen schriftlichen Antwortformat bezüglich dieser Fälle bestimmte Aufgaben/Probleme lösen mussten. Beispielsweise sollte nach SPIKES-Protokoll einer Patientin die Diagnose Brustkrebs eindeutig, aber einfühlsam mitgeteilt werden. Es zeigte sich ein signifikanter Zusammenhang zwischen OSCE und Key-Feature-Prüfung von $r = .32$. Die Autoren schlussfolgerten, dass aufgrund der Höhe der Korrelation offenbar verschiedene Wissens Ebenen und Fertigkeiten durch die beiden Prüfungen adressiert werden (Handlungswissen und praktische Leistung unter „Laborbedingungen“).

2.8 Constructive Alignment und AGIL-Schema

Eine weitere wichtige Fragestellung in Bezug auf Lehre und Prüfungen an Hochschulen ist die Frage nach der Deckung zwischen Lehr- und Prüfungsinhalten. Idealerweise deckt die Prüfung die gelehrten Inhalte ab und misst den Lernerfolg der Studierenden. Biggs (69) befasste sich intensiv mit dieser Fragestellung und legte dar, dass sich Lehrende darüber im Klaren sein müssen, was die Studierenden erlernen sollen und wie sie sicherstellen können, dass die Studierenden diese Lernziele auch erreicht haben. In Prüfungen soll idealerweise gezeigt werden, dass das Erlernete nicht nur wiedergegeben werden kann, sondern auch verstanden wurde. Entsprechend empfiehlt er möglichst anwendungsrelevante Aufgaben zu erstellen, die dies abprüfen können.

Wildt und Wildt (70) beschreiben bezugnehmend auf Biggs, dass der Lehr-/Lernprozess, das Prüfungsformat und die späteren beruflichen/alltäglichen Anforderungen aufeinander abgestimmt sein müssen. Sie geben an, dass diese Forderung in der Hochschuldidaktik allgemein anerkannt wird und die Realisierung dieses Ziels grundsätzlich erforderlich ist.

Das vorgestellte Prinzip wird als Constructive Alignment bezeichnet (69).

Zur Erfassung der Qualität von Lehrveranstaltungen kann zudem ein Modell herangezogen werden, das sich auf vier Dimensionen guter Lehre bezieht (Parson, 71, in der überarbeiteten Form von Schmidt et al., 72).

Diese vier Dimensionen sind nach Parson „Adaption“, „Goal Attainment“, „Integration“ und „Latent Pattern Maintenance“ (AGIL-Schema).

In der Modifikation beschreiben Schmidt et al. die Dimension „Adaption“ als Bereich der „Ressourcen und Strukturen“. Darunter fassen sie beispielsweise die Qualität und Eignung der Dozenten, zur Verfügung stehende Räumlichkeiten oder die Ausstattung.

Die Dimension „Goal Attainment“ bezieht sich auf die Aspekte „Zielerreichung und Ergebnisse“, wobei die in der Lehre vermittelten Fertigkeiten und das vermittelte Wissen dieser Dimension zugeordnet werden.

Gute Arbeitsmaterialien oder hilfreiche Rückmeldungen durch Dozenten sind zudem Inhalte der Dimension „Integration“ bzw. in der Modifikation „Integration und Prozesse“.

Die letzte der vier Dimensionen, die nach Parson als „Latent Pattern Maintenance“ bezeichnet wird, und die von Schmidt et al. als Dimension „Kultur und Ziele“ angepasst wurde, bezieht sich beispielsweise auf die Transparenz von Zielen oder die Menge und das Tempo des Lernstoffes.

Die vier Dimensionen wurden von den Autoren empirisch überprüft.

Auf das Schema wird im Hinblick auf die Evaluation der Videoprüfungen dieser Arbeit in Abschnitt 3.4 Bezug genommen.

2.9 Konzept videobasierte Prüfung und dazugehörige Hypothesen der vorliegenden Arbeit

Die dargestellten Vor- und Nachteile der beiden häufigsten Prüfungsformen ärztlicher kommunikativer Kompetenz im deutschsprachigen Raum sowie die vorgestellten Überlegungen und Vorarbeiten zum Einsatz von Videos führten zu dem Vorhaben eine videobasierte Prüfung zur Messung der Leistungen der Studierenden im Kursus des Fachs Medizinische Psychologie und Medizinische Soziologie zu entwickeln, zu erproben und wissenschaftlich zu untersuchen.

Entsprechend dem Konzept Constructive Alignment (69, 70) sollten bei der Prüfungsentwicklung möglichst viele Lernziele und Lehrinhalte des Kursus mit seinen theoretischen Inhalten, zahlreichen (Handlungs-)Modellen und praktischen Übungen abgebildet werden. Abgesehen von praktischen Übungen wie Rollenspielen wurden die Kursinhalte ansonsten nach den Kompetenzmodellen nach Miller (46) und Krathwohl (53), in Anlehnung an Bloom (52), auf den Ebenen Fakten- und Handlungswissen eingeordnet. Entsprechend sollten die Items der Prüfung Fakten- und Handlungswissen abbilden und dahingehend konzipiert werden. Die Ebene von praktischen Übungen im Kursus konnte entsprechend den bisherigen Darlegungen in dieser Form nicht in die geplante Prüfung integriert werden (Kompetenzebene „shows how“).

Trost et al. (57) gehen des Weiteren davon aus, dass eine Prüfung mit Videosequenzen und in ihrem Fall Multiple-Choice-Fragen Fakten- und Handlungswissen abfragen kann. Ebenso schätzen Rottmann et al. (58) ein, dass sich Videos dazu eignen, eine hohe Handlungsorientierung zu erzielen, womit Items auf der Ebene von Handlungswissen neben Items auf der Ebene von Faktenwissen mit der geplanten Videoprüfung gut realisierbar erscheinen. Auch Ludwig (64) arbeitet mit der Einordnung der videobasierten Key-Feature-Prüfung als Abfrage prozeduralen Wissens.

In einer ersten Projektphase wurde entschieden, die videobasierte Prüfung in Bezug auf die Items zu Handlungswissen mit Aufgaben im Sinne eines Situational Judgement Tests zu gestalten und als bestehensrelevante Prüfung abzuhalten. Als Antwortformat wurde das Single-Choice-Format festgelegt. Die Prüfung wurde als Video-based Single-Choice Examination, VSE, benannt. Es sollten 30 Items zu Fakten- und Handlungswissen entsprechend des Prüfungsschemas der Institution erstellt werden (für die genaue Ausgestaltung vgl. 3.2).

In einer zweiten Projektphase, die nicht Teil der vorliegenden Arbeit ist, sollte zudem basierend auf den Erfahrungen dieser ersten videobasierten Prüfung eine Prüfung mit verzweigten Antworten entwickelt werden.

Entsprechend der theoretischen Einordnungen der Lehrinhalte und der geplanten Videos nach Wissenssebenen sowie der entsprechenden Konzeption der geplanten videobasierten Prüfung VSE stellte sich die Frage, ob sich die theoretisch angenommene Einteilung in Fakten- und Handlungswissen auch entsprechend in einer zweifaktoriellen Lösung der Prüfungssitems zeigt. Diese Fragestellung sollte untersucht werden, es wird angenommen, dass den Prüfungssitems eine zweifaktorielle Struktur zugrunde liegt (Hypothese 1).

Die VSE sollte zudem zunächst inhaltlich mit der OSCE verglichen werden. Zwar beziehen sich beide Prüfungen auf denselben Kursus, allerdings wird aufgrund der Einschätzung, dass die VSE der Theorie nach Fakten- und Handlungswissen abbildet und die OSCE die Kompetenzebene der praktischen Leistung unter „Laborbedingungen“ bedient, angenommen, dass die beiden Prüfungsformen zwar zusammenhängen, der Zusammenhang allerdings gering ausfällt (Hypothese 2). Diese Annahme entspricht auch den Ergebnissen von Humphris und Kaney (62) sowie Fischbeck et al. (68) zum Zusammenhang von OSCE und OSVE und OSCE und Key-Feature-Prüfung.

Die wesentlichen Annahmen und zugrundeliegenden Theorien der Prüfungskonzeption und des Vergleichs mit der OSCE werden mit dem Ziel eines guten Überblicks nochmal in Abbildung 4 skizziert.

Neben der Annahme, dass VSE und OSCE grundsätzlich zusammenhängen, sollte auch untersucht werden, ob es Unterschiede im Zusammenhang zwischen den beiden Prüfungen in Bezug auf die unterschiedlichen Wissenssebenen gibt. Wird nochmal Bezug genommen auf die Einteilung nach Miller (46) und Krathwohl (53) hinsichtlich der Kompetenzebenen, gehen diese von einer hierarchischen Struktur der Ebenen aus, wonach Faktenwissen weniger nah an die tatsächliche Handlung heranreicht als Handlungswissen. Dieser Einordnung folgend stellt sich die Frage, ob die OSCE, die die praktische Handlung unter „Laborbedingungen“ abprüft, in einem höheren Zusammenhang mit den Teilen der VSE steht, die sich auf Handlungswissen beziehen, als mit der Gesamtprüfungsleistung, die auch Faktenwissen beinhaltet. Es wird die Hypothese formuliert, dass ein höherer Zusammenhang zwischen der Prüfungsleistung in den auf Handlungswissen bezogenen Items der VSE und der OSCE-Prüfungsleistung als zwischen der Gesamtprüfungsleistung der VSE (Fakten- und Handlungswissen) und der OSCE-Prüfungsleistung besteht (Hypothese 3).

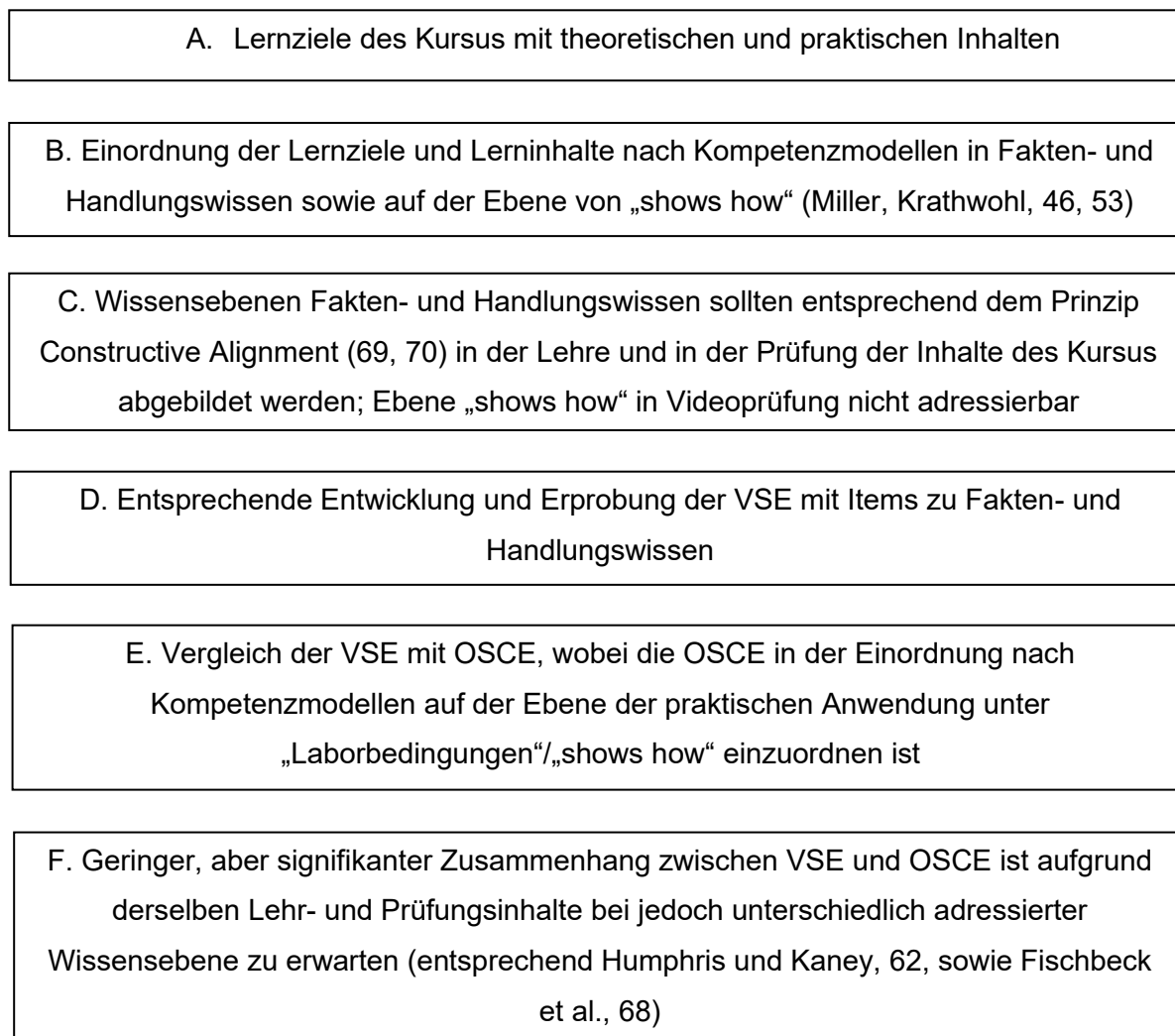


Abbildung 4: *Theoretische Herleitung Prüfungskonzeption und OSCE-Vergleich*

Des Weiteren sollen die internen Konsistenzen der beiden Prüfungen VSE und OSCE untersucht werden. Davon ausgehend, dass mit einer höheren Anzahl an Items die Höhe der internen Konsistenz zunimmt (44, 45), wird angenommen, dass die VSE eine gute interne Konsistenz erzielen kann. Die VSE wird mit 30 Items geplant, eine Itemanzahl, die ausreichend hoch erscheint, um eine konsistente Skala konstruieren zu können. Die eingesetzte OSCE besteht hingegen aus 5 Stationen, eine Anzahl die als eher gering und damit vermutlich als weniger reliabel einzuschätzen ist (vgl. 39). Zudem stellt eine hohe Objektivität eine wichtige Voraussetzung für eine hohe Reliabilität dar (44). Davon ausgehend, dass die VSE eine hohe Objektivität durch Softwaresteuerung der Prüfung und Rater unabhängige, ebenfalls softwaregesteuerte Auswertung der Ergebnisse erzielen kann,

wird angenommen, dass dies zusätzlich einen positiven Effekt auf die Höhe der Reliabilität haben wird.

Folglich wird die Annahme getroffen, dass die VSE eine hohe interne Konsistenz aufweist (Hypothese 4).

Die vorliegende Arbeit hat sich bereits in verschiedener Hinsicht mit der Frage beschäftigt, in welchem Ausmaß die unterschiedlichen Prüfungsformen einen praktischen Bezug zum Arbeitsalltag der angehenden Ärzte herstellen. Es wurden Überlegungen zum Einsatz von Videos und empirische Ergebnisse dargestellt. Auch in Bezug auf die für diese Arbeit entwickelte, videobasierte Prüfung stellte sich die Frage, wie hoch der Praxisbezug durch Videos ausfällt. Die Annahme bei Projektbeginn bestand darin, dass sich der Praxisbezug durch den Einsatz von Videos im Vergleich zu schriftlichen Prüfungen erhöhen lässt. Zu dieser Annahme kommen wie berichtet auch Trost et al. (57) und Rottmann et al. (58), die davon ausgehen, dass Videos den Arzt-Patienten-Kontakt mit einem höheren Realitätsbezug abbilden können als in schriftlichen Szenarien möglich bzw. dass die Einbettung von Videos eine höhere plastische und realistische Darstellung von Inhalten zur Erzielung einer hohen Handlungsorientierung ermöglicht. Im Rahmen der Arbeiten von Trost et al. und Ludwig (64) erfolgte die Einschätzung der Höhe des Praxisbezugs jeweils über Studierendenbefragungen. Bei beiden Untersuchungen schätzten die Studierenden die Videos als realistisch ein bzw. sahen diese als geeignete Vorbereitung für den späteren Berufsalltag an.

Für die vorliegende Arbeit wurde ebenfalls entschieden, die Höhe des Praxisbezuges über eine Studierendenbefragung bewerten zu lassen. Hierbei wird entsprechend bisherigen Darlegungen davon ausgegangen, dass der Einsatz von Videos einen von den Studierenden in der Evaluation zurückgemeldeten hohen erlebten Praxisbezug der Prüfung (und einen höheren Praxisbezug als bei rein schriftlichen Prüfungen) erzielt (Hypothese 5).

2.10 Konstrukt Familienkommunikation und dazugehörige Hypothesen der vorliegenden Arbeit

Neben der Entwicklung und Überprüfung der VSE wurde für diese Arbeit noch ein weiterer Untersuchungsbereich hinzugenommen, der sich mit einer möglichen Determinante der Prüfungsleistung der Studierenden beschäftigt. Bezugnehmend auf die Arbeit von Fischbeck (73) steht die Kommunikationsqualität in der Ursprungsfamilie in Zusammenhang mit der Kommunikationsqualität der Medizinstudierenden in der Arzt-Patienten-Beziehung (gemessen durch die Prüfungsleistung).

Nach Epstein et al. (74) stellt Familienkommunikation einen Prozess dar, bei dem Informationen zwischen Familienmitgliedern in verbaler und non-verbaler Form ausgetauscht werden.

Koerner und Fitzpatrick (75) beschreiben unter Bezugnahme gesichteter Literatur, dass die Definition von Familie zunehmend vom biologischen Fokus abkommt und stärker betont, dass eine Familie Gemeinschaft, ein Zuhause, Gruppenidentität und eine geteilte Vergangenheit und Zukunft bedeutet. Sie erarbeiteten eine Theorie zur Familienkommunikation auf Grundlage von Vorarbeiten von Fletcher (76) und Fletcher und Thomas (77) zu Beziehungsschemata (s. Abb. 5).

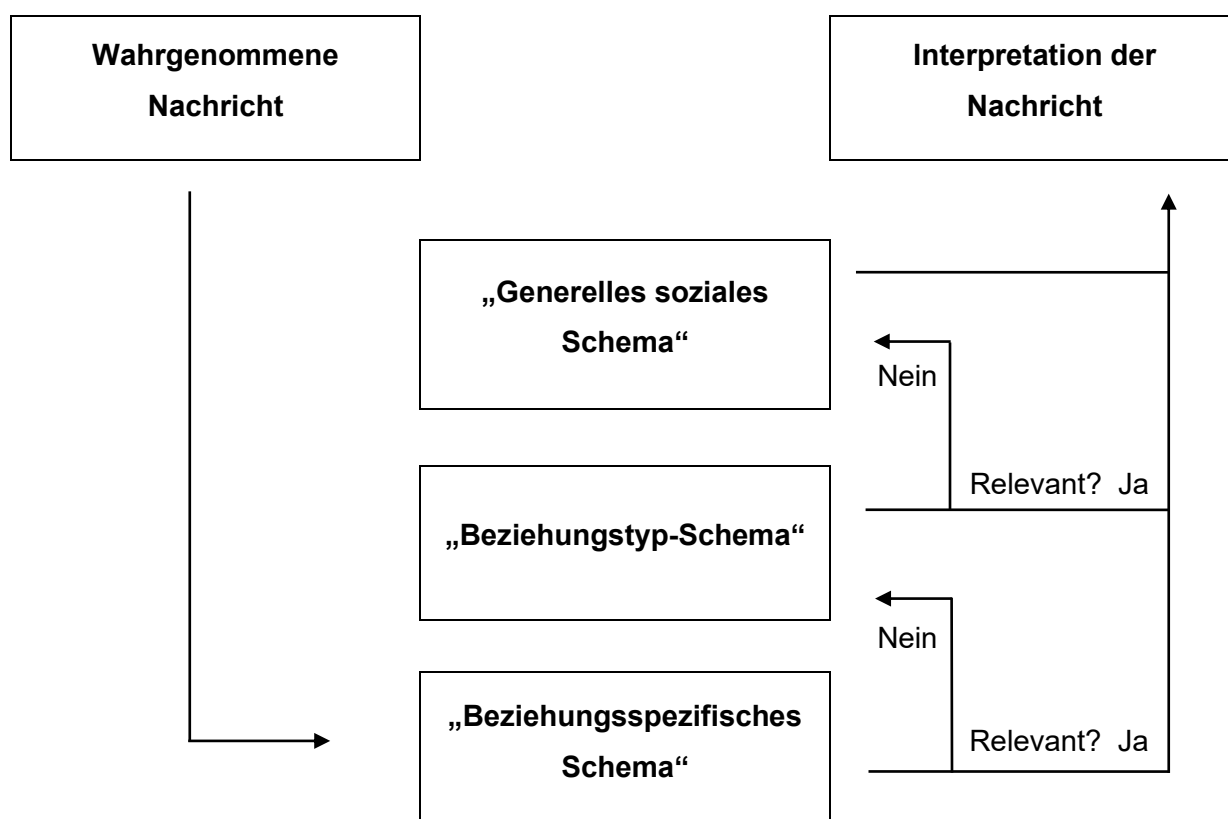


Abbildung 5: Theorie der Familienkommunikation nach Koerner und Fitzpatrick (75)

Die Grundannahme besteht zunächst darin, dass es verschiedene Schemata zu Beziehungswissen gibt. Es wird unterschieden in „Generelles soziales Schema“, „Beziehungstyp-Schema“ und „Beziehungsspezifisches Schema“. Im „Beziehungsspezifischen Schema“ ist Beziehungswissen zu einer bestimmten Person abgespeichert, im „Beziehungstyp-Schema“ ist Wissen über die Art der Beziehung enthalten (z.B. Paarbeziehung, Freundschaftsbeziehung, kollegiale Beziehung) und das „Generelle soziale Schema“ enthält allgemeines Wissen, über das Menschen über alle ihre

Beziehungen hinweg verfügen. Die Schemata sind hierarchisch organisiert, d.h. als erstes greift das „Beziehungsspezifische Schema“, dann das „Beziehungstyp-Schema“ und zuletzt das „Generelle soziale Schema“. Sie legen des Weiteren dar, dass das in verschiedene Schemata unterteilbare Beziehungswissen in seinem Aufbau, neben anderen Faktoren der Sozialisierung, maßgeblich durch die Familie bestimmt wird, die in der Regel im ersten Abschnitt des Lebens dominant und prägend ist. Sie nehmen an, dass Menschen ihr abgespeichertes Wissen aus der Familie in Bezug auf andere Beziehungen abrufen und anwenden.

Eine weitere, in der Einschätzung von Koesten (78) weniger beachtete Theorie, stammt von Chaffee et al. (79) und bezieht sich auf eine Vier-Felder-Typologie. Nach dieser Typologie lernen Kinder viel mehr von der Struktur der Familienkommunikation als von ihrem Inhalt. Die Einordnung von verschiedenen Typen der Familienkommunikation erfolgt zweidimensional anhand der Dimensionen „soziale Orientierung“ und „Konzept-Orientierung“, wobei im Falle einer hohen „sozialen Orientierung“ der Fokus der Kommunikation auf Konfliktvermeidung liegt und im Falle einer hohen „Konzept-Orientierung“ auf der Betrachtung aller „Seiten“ und dem Vertreten der eigenen Meinung, auch wenn sie anderen widerspricht. Anhand einer Dichotomisierung (niedrig/hoch) dieser beiden Dimensionen ergeben sich vier Typen der Struktur von Familienkommunikation. Die vier Typen benannten sie mit „laissez faire“, „pluralistisch“, „protektiv“ und „übereinstimmend“. In einer Familie des Typs „laissez faire“ wird kein Kommunikationstyp favorisiert oder bestärkt. Beim „pluralistischen“ Typ werden Kinder stark dazu angehalten Meinungen in Frage zu stellen und sich mit ihnen auseinanderzusetzen. Die soziale Anpassung steht beim „protektiven“ Typ im Vordergrund. Beim „übereinstimmenden“ Typ werden hingegen beide Dimensionen betont und gefördert. Sie gehen darüber hinaus davon aus, dass die Erfahrungen in der Familie mit Kommunikation im Sinne einer „kognitiven Mappe“ auf andere soziale Situationen übertragen werden. Sie halten diese Übertragung bis ins Erwachsenenalter für stabil.

Nach Koesten (78) waren Stone und Chaffee (80) die ersten, die die beiden vorgenannten Dimensionen von Familienkommunikation vorstellten.

Fitzpatrick und Ritchie (81, 82) beziehen sich auf diese zwei Dimensionen der Familienkommunikation, die sie als Teil des vorgestellten Beziehungswissens, geordnet in verschiedene Beziehungsschemata (75), ansehen. Sie benannten die Dimensionen um in „Gesprächsorientierung“ und „Übereinstimmungsorientierung“.

Erstere bezieht sich in deren Darlegung auf das Ausmaß, in dem es einer Familie gelingt, dass sich alle Mitglieder ermutigt fühlen uneingeschränkt an Interaktionen zu verschiedenen

Themen teilzunehmen. In Familien mit einer hohen Gesprächsorientierung wird viel Zeit miteinander verbracht und viel sowie offen diskutiert.

Die zweite Dimension bezieht sich auf das Ausmaß eines Klimas von Homogenität bezüglich Einstellungen, Werten und Überzeugungen. In Familien mit einer hohen Übereinstimmungsorientierung beziehen sich Interaktionen auf Konformität, Konfliktvermeidung und gegenseitige Abhängigkeit.

Auf Grundlage dieser Dimensionen und unter Bezugnahme und Verbesserung/Modernisierung eines Fragebogens von 1972 (83) wurde von den Autoren der Fragebogen „The Revised Family Communication Pattern Instrument“ in zwei Versionen für Eltern und Kinder entwickelt. Er besteht aus 26 Items und die Selbsteinschätzung erfolgt auf einer siebenstufigen Likert-Skala. Im Methodenteil wird auf diesen Fragebogen nochmal verwiesen und dieser wird für die Erstellung eines eigenen Fragebogens genutzt.

In einer Metaanalyse untersuchten Schrodtt et al. (84) anhand von 56 eingeschlossenen Studien den Einfluss von Familienkommunikation auf Informationsverarbeitungsprozesse wie kognitive Flexibilität, Einstellung zu politischen Themen oder den Umgang mit Medien, zudem auf Verhaltensweisen wie Aggression oder Alkoholkonsum und zuletzt noch in Bezug auf psychosoziale Aspekte wie psychische und physische Gesundheit, Selbstwert, Ängste und weitere. Es zeigten sich signifikante Zusammenhänge zwischen Familienkommunikation und allen drei Aspekten im Gesamten sowie einzeln betrachteten Zusammenhängen (Korrelationen zwischen $r = .29$ und $r = .46$). Am höchsten war der Zusammenhang zwischen der Dimension „Übereinstimmungsorientierung“ des Konstrukts Familienkommunikation und psychosozialen Aspekten. Die Autoren schließen aus den Ergebnissen, dass die Familienkommunikation mit vielen Aspekten des Lebens zusammenhängt und einen hohen Einfluss auf verschiedenste Bereiche des Lebens ausübt.

In einer empirischen Arbeit untersuchte Koesten (78) an $n = 299$ Probanden, inwieweit die Qualität der Familienkommunikation (gemessen mit „The Revised Family Communication Pattern Instrument“) mit der Fähigkeit zu gelungener Kommunikation in Freundschafts- und Paarbeziehungen (gemessen anhand einer Selbsteinschätzung) zusammenhängt. Es zeigte sich, dass mit einer höheren Qualität der Familienkommunikation auch bessere Kommunikationskompetenzen in den anderen beiden Beziehungen einhergehen.

Hojat et al. (85) untersuchten zudem $n = 1710$ Medizinstudierende und betrachteten u.a. den Zusammenhang zwischen ihrer klinischen Kompetenz/Leistung während ihrer Famulatur und der wahrgenommenen Qualität der früheren Beziehung zu ihren Eltern. Es zeigte sich, dass

die Studierenden, deren Kompetenz geringer ausfiel, auch eine geringere Zufriedenheit in Bezug auf ihre frühe Beziehung zu ihren Eltern angaben.

Bei Fischbeck (73) wurde die Kommunikationsqualität in der Ursprungsfamilie anhand einer Selbsteinschätzung über die zwei Items „Wie intensiv wurde in Ihrer (Ursprungs-) Familie miteinander gesprochen?“ (Aspekt der Intensität) und „Wie offen können Sie in Ihrer (Ursprungs-)Familie über Ihre Anliegen und Probleme sprechen?“ (Aspekt der Offenheit für Problemmunikation), eingeschätzt auf einer fünfstufigen Likert-Skala von „sehr wenig“ bis „sehr viel“ bzw. „nicht“ bis „sehr offen“, untersucht. Bei einer Stichprobengröße von $n = 130$ ergaben sich signifikante Korrelationen von $r = .24$ zwischen der Leistung der Studierenden in der OSCE und der Intensität der Familienkommunikation sowie $r = .29$ zwischen der Leistung der Studierenden in der OSCE und der Offenheit für Problemmunikation in der Familie. Es zeigte sich somit: je intensiver Kommunikation in der Ursprungsfamilie stattfand und je offener in der Ursprungsfamilie über Anliegen und Probleme gesprochen wurde, desto besser schnitten die Medizinstudierenden in den praktischen Aufgaben der OSCE ab, bei denen Sie adäquat mit Patienten in Bezug auf unterschiedliche Problemstellungen kommunizieren sollten.

Fischbeck untersuchte zudem den Einfluss von Alter und Geschlecht auf die Prüfungsleistung. Es zeigte sich kein signifikanter Zusammenhang zwischen dem OSCE-Ergebnis und dem Alter. Sie diskutiert, dass die Stichprobe bezogen auf das Alter homogen war und der fehlende Effekt ggf. darauf zurückzuführen sein könnte. Hinsichtlich des Geschlechts ergab sich ein signifikanter Unterschied. Weibliche Studierende schnitten besser ab als männliche Studierende.

Auch in einer Übersichtsarbeit (86) zeigte sich, dass das Geschlecht einen Einfluss auf den Studienerfolg von Medizinstudierenden hat.

Sowohl Fischbeck als auch Erdel (87) legen dar, dass Forschungen im schulischen und im universitären Kontext immer wieder bestätigen können, dass Mädchen und Frauen bessere Leistungen erzielen.

Dabei nehmen weibliche Studierende das Studium ernster, gehen formaler vor und lernen mehr auswendig (88).

Bezüglich des Alters kommt auch eine weitere Untersuchung neben der vorgenannten von Fischbeck zu dem Ergebnis, dass der Studienerfolg von Medizinstudierenden vom Geschlecht, nicht aber vom Alter abhängt (89).

Anhand der dargestellten Theorien und empirischen Befunde wird in dieser Arbeit angenommen, dass Studierende im Kontext der Arzt-Patient-Kommunikation angeeignetes

Beziehungswissen aus ihrer Ursprungsfamilie einbringen und auf die Beziehung und Kommunikation zwischen Arzt und Patient übertragen. Diese Übertragung soll sich in einem Zusammenhang zwischen der selbst eingeschätzten Familienkommunikation in der Ursprungsfamilie und den Prüfungsergebnissen zur Messung ärztlich kommunikativer Kompetenz widerspiegeln.

Es wird vor allem entsprechend der Ergebnisse von Fischbeck (73) erwartet, dass eine von Offenheit und intensivem Austausch geprägte Familienkommunikation mit einem besseren Abschneiden in den Prüfungen VSE und OSCE zusammenhängt (Hypothese 6a).

Zudem wird erwartet, dass weitere Facetten der Familienkommunikation (wie respektvoller Umgangston, gemeinsame Entscheidungsfindung in Anlehnung an „The Revised Family Communication Pattern Instrument“) ebenfalls mit den Prüfungsleistungen zusammenhängen (Hypothese 6b).

Zusätzlich soll der Einfluss von Alter und Geschlecht, neben der Familienkommunikation, auf die Prüfungsergebnisse betrachtet werden. Entsprechend bisherigen Befunden (73, 86-89) wird angenommen, dass das Geschlecht neben der Familienkommunikation mit dem Ergebnis der Prüfungen zusammenhängt und diese mitbestimmt (weibliche Studierende schneiden in den Prüfungen besser ab als männliche Studierende), das Alter hingegen nicht (Hypothese 6c).

2.11 BMBF-gefördertes Forschungsprojekt

Die Inhalte dieser Arbeit sind Teil des Forschungsprojekts „digiRole“, das vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wurde. Die Ethikkommission des Psychologischen Instituts der Universität Mainz hatte keine Einwände bezüglich der Durchführung des Projekts (2017-JGU-psychEK-007).

Während sich die vorliegende Arbeit hauptsächlich auf die Entwicklung und Erprobung der videobasierten Prüfung konzentriert, werden in einer weiteren Dissertation im Rahmen des Forschungsprojekts vorwiegend zusätzliche Einflussfaktoren untersucht, die den Erwerb kommunikativer Fertigkeiten im Medizinstudium bedingen (wie Lernprozess und Studierverhalten, Eingangsvoraussetzungen, Persönlichkeit). Der Fachbereich dabei ist die Soziologie, aus deren Blickwinkel diese Einflussfaktoren betrachtet werden.

3. Methoden

3.1 Studienteilnehmer und -ablauf

Im Sommersemester 2018 (SS18) und im Wintersemester 2018/2019 (WS18/19) wurde die VSE jeweils bei den Medizinstudierenden des zweiten Semesters an der Universitätsmedizin Mainz als bestehensrelevante Prüfung zum Semesterende eingesetzt. Im WS18/19 wurden die Lerninhalte zusätzlich mit der OSCE überprüft.

Die VSEs wurden technisch über die Prüfungssoftware ILIAS realisiert und die Abnahme der Prüfungen erfolgte in den Räumlichkeiten des Zentrums für Datenverarbeitung. Zum Anhören der Videos wurden die Studierenden gebeten, eigene Kopfhörer mitzubringen. Zudem standen auch Kopfhörer zur Verfügung, die ausgegeben werden konnten. Bei der Durchführung kam es vereinzelt zu Problemen mit der Tonwiedergabe und den Ladezeiten der Videos, die allerdings rasch behoben werden konnten. Ansonsten verliefen die Prüfungen reibungslos.

Inhaltlich wurden die Studierenden im Kursus mit Beispielaufgaben auf die Prüfungen vorbereitet. Im Anschluss an den Unterricht erfolgten die Studierendenbefragungen mit „Papier und Stift“. Die Evaluation wurde jeweils nach den Prüfungen ausgehändigt. Angaben zu Anzahl, Alter und Geschlecht der Studienteilnehmer sind in Tabelle 1 aufgeführt.

Tabelle 1: *Studienteilnehmer*

	VSE SS18	VSE WS18/19	OSCE WS18/19	OSCE und VSE WS18/19
n	226	193	195	192
Alter	Mdn = 22 IQR = 6 (18-45)	Mdn = 25 IQR = 4 (19-49)*	Mdn = 25 IQR = 4 (19-49)*	Mdn = 25 IQR = 4 (19-49)*
Geschlecht weiblich	58 %	58 %	57 %	58 %
Geschlecht männlich	42 %	42 %	43 %	42 %

Anmerkung. *2 missing.

3.2 Entwicklung Video-based Single-Choice Examination (VSE)

In der VSE sollten die grundlegenden Konzepte aus dem Kursus abgedeckt sein. Es flossen daher u.a. das NURSE-Modell (29), Gesprächsführungstechniken wie aktives Zuhören (90), das Prinzip der Partizipativen Entscheidungsfindung (25) und das SPIKES-Protokoll (30) in die Prüfung ein.

Bezugnehmend auf die Lern- und Prüfungsinhalte vorangegangener Jahre wurden Patientenfälle und Arzt-Patienten-Dialoge erarbeitet. Die Patientenfälle und Arzt-Patienten-Dialoge beinhalteten die Bereiche Medikamentencompliance prüfen und fördern, Anamnesegespräch führen, Krebsdiagnose nach SPIKES-Protokoll mitteilen, Visitengespräch führen, mit Patient Therapie bei Hypertonie nach dem Prinzip der Partizipativen Entscheidungsfindung absprechen, bei Patientin mit Adipositas eine Verhaltensänderung herbeiführen, eine Analyse der Stressreaktion vornehmen und Informationen vermitteln.

Die Fälle und Dialoge wurden in einem ersten Schritt von Kursdozenten gelesen und überarbeitet und dann einer Expertenbefragung unterzogen. Bei den Experten handelte es sich um Allgemeinmediziner und Psychologen, d.h. Personen die den praktischen Berufsalltag als Arzt kennen, die Inhalte lehren oder beides. Zudem wurde eine Expertin befragt, die bereits an der Entwicklung der OSCE mitgewirkt hatte. Anhand schriftlicher Rückmeldungen der Experten (orientiert an gestellten Fragen, beispielsweise Bitte um Einschätzung der Glaubwürdigkeit von Setting, Krankengeschichte, Gesprächsablauf, Patientendarstellung usw.) wurden die Dialoge überarbeitet und angepasst.

Die Produktion der Videos erfolgte in Zusammenarbeit mit dem Medienzentrum (Universität Mainz) und der Lernklinik (Universitätsmedizin Mainz). Die Rollen wurden von Laienschauspielern und einem Allgemeinmediziner verkörpert.

Die Darstellung des Bildmaterials wurde mit einem Split-Screen-Verfahren realisiert, d.h. die Studierenden sahen in der Prüfung jeweils die Totale und die Frontalaufnahmen (s. Abb. 6). Dadurch sollten Schnitte und Perspektivwechsel als Einflussgröße ausgeschlossen werden.

Bezugnehmend auf die Erfahrungen von Hoppe-Seyler et al. (91), wie wichtig auch die von den Studierenden wahrgenommene Glaubwürdigkeit des Bildmaterials ist, wurde vor den Hauptdreharbeiten eine Probeproduktion realisiert. Bei einer Evaluation dieser Probeproduktion durch $n = 8$ Studierende (über einen E-Mailverteiler von Medizinstudierenden verschiedener Semester) schätzte die Mehrzahl der Studierenden die

dargestellte Arztrolle (63 %), Patientenrolle (75 %), Praxis (88 %) und Interaktion (88 %) als authentisch ein.



Abbildung 6: *Videoausschnitt VSE*

Jede entwickelte Prüfungsfrage stand in Bezug zu einem Videoausschnitt und es gab jeweils fünf Antwortmöglichkeiten, die im Single-Choice-Format gehalten waren. Die Prüfungsfragen orientierten sich an Prüfungsfragen vorangegangener Semester.

Die Prüfungsaufgaben bezogen sich, wie im Kapitel 2 dargestellt, entweder auf Faktenwissen, das in Bezug zum Video steht, oder auf eine geforderte Handlungsentscheidung/Auswahl einer geeigneten Formulierung im Sinne eines Situational Judgement Tests.

Inhaltlich wurde sich zudem für das Abfragen einer weiten Bandbreite an Themenfeldern aus dem Kursus entschieden, um möglichst viele Lernziele abprüfen zu können. In der ersten Prüfung lag der Schwerpunkt etwas mehr auf Fakten- und in der zweiten Prüfung etwas mehr auf Handlungswissen (Aufteilung in Fakten- und Handlungswissen siehe Tabelle 2 und 3). Die in einem ersten Entwurf erarbeiteten Fragen wurden von Kursdozenten durchgesehen und angepasst.

Da die Inhalte der Prüfung weiterhin im Fachbereich Anwendung finden, können die Prüfungssitems nicht veröffentlicht werden. Allerdings soll im Folgenden eine Beispielaufgabe vorgestellt werden, die der Veranschaulichung dient:

Im Video wird der Fall einer 55-jährigen Patientin mit Magenschmerzen und Schlafstörungen dargestellt. Sie beklagt, dass viele Sorgen sie vom Schlafen abhalten würden. Die

Prüfungsfrage lautet: „Welche Aussage wäre vor dem Hintergrund der Ihnen bekannten Gesprächsstörer und Gesprächsförderer als Reaktion auf die Äußerung der Patientin am ehesten gesprächsfördernd?“ und die fünf Antwortmöglichkeiten sind „Versuchen Sie im Bett nicht zu grübeln, da sollten Sie dann besser aufstehen.“, „Sie sind also ein Mensch, der sich schnell viele Sorgen macht?“, „Frau Hamann, ganz langsam, es ist doch meist nicht so schlimm, wie man denkt.“, „Sie liegen also nachts wach und grübeln über Ihre Sorgen.“ (richtige Antwort) und „Beim Grübeln hilft es, wenn Sie sich mit etwas anderem ablenken, vielleicht lesen Sie z.B. etwas.“

Die fünf Antwortmöglichkeiten wurden bei der Auswertung der VSE dichotomisiert („richtig“/„falsch“). Die richtigen Antworten wurden mit einem Punkt bewertet und das Prüfungsergebnis stellte den Summenwert aller richtigen Antworten dar.

Aufgrund der noch weiter unten berichteten Ergebnisse der Hauptachsenanalyse wurden fünf Items sowohl im SS18 als auch im WS18/19 verwendet.

3.3 OSCE

Im Fach Medizinische Psychologie und Medizinische Soziologie in Mainz liegen bereits langjährige Erfahrungen mit einer Kommunikations-OSCE vor (35), auf die für diese Arbeit zurückgegriffen wurde.

Die im Fach entwickelte OSCE besteht aus fünf verschiedenen Stationen, welche die Zentralthemen des Kurses Anamnese, Analyse der Stressreaktion, Compliance überprüfen und fördern, Partizipative Entscheidungsfindung und Mitteilung einer Krebsdiagnose nach SPIKES-Protokoll betreffen.

Die Bewertung der Studierenden erfolgt sodann anhand von Checklisten, bei denen die Prüfer vorab festgelegte Items als „gezeigt“ oder „nicht gezeigt“ einschätzen. Ein solches dichotomes Vorgehen wurde gewählt, um den Prüfungsinhalt des jeweiligen Items im Gespräch zwischen dem Studierenden und dem Simulationspatienten unter dem Gesprächsfluss möglichst leicht erfassen zu können. Dafür eignet sich ein dichotomes Vorgehen besser als ein globales Rating (35). Die Punkte beziehen sich dabei auf die jeweiligen Inhalte der einzelnen Stationen und zusätzlich auf Aspekte der Gesprächsführung wie Empathie, aktives Zuhören, Blickkontakt und offene Körperposition.

Mit den Simulationspatienten fand im Vorfeld der Prüfung ein Treffen zur Vorbereitung statt. Dabei wurde explizit auf ein standardisiertes Vorgehen hingearbeitet, d.h. die Simulationspatienten sollten keine Variation in der Darstellung ihres Patienten über die Durchgänge mit den Studierenden hinweg vornehmen, sondern stets in gleicher Form agieren und reagieren.

An drei aufeinanderfolgenden Prüfungstagen wurden die Studierenden von insgesamt 11 Prüfern bewertet (Mitarbeiter der Universitätsmedizin und der Universität Mainz, ein Allgemeinmediziner als externer Prüfer).

Zusätzlich wurde der Prüfungsablauf von wissenschaftlichen Hilfskräften unterstützt, die beispielsweise die Studierenden in Empfang nahmen und die Prüfungs- und Wechselzeiten an den einzelnen Stationen takteten. Je Station hatten die Studierenden fünf Minuten Zeit, eine Minute Zeit gab es für den Stationswechsel.

Im Vergleich zur VSE sind die geprüften Inhalte bei der OSCE durch die Begrenzung auf fünf Stationen weniger breit, die Themen der fünf Stationen finden sich allerdings auch alle in der VSE wieder.

Für die OSCE konnte eine maximale Punktzahl von 35 Punkten mit maximalen 7 Punkten je Station erzielt werden. Die Punkte der Checklisten waren für die Bewertung unterschiedlich gewichtet, z.B. gab es für manche Kriterien 0,5 Punkte und für andere z.B. 1 Punkt.

Die genauen Prüfungsinhalte der OSCE können in Analogie zur VSE nicht veröffentlicht werden.

3.4 Evaluation

Die Evaluation wurde jeweils als Paper-Pencil-Befragung im Anschluss an die Prüfungen durchgeführt. Nachdem die Studierenden die VSEs absolviert hatten, füllten sie die Evaluation im Prüfungssaal an ihrem Sitzplatz aus. Nach der OSCE erfolgte die Evaluation auf Sitzplätzen vor den Prüfungsräumen.

Die gesamte Evaluation wurde unter Bezugnahme des AGIL-Schemas (71, 72, vgl. 2.8) ausgearbeitet, d.h. die Inhalte der vier Dimensionen wurden mit den Items der Evaluation abgedeckt. Für die vorliegende Arbeit sind allerdings nur einzelne Items aus der Evaluation relevant, die im Folgenden dargestellt werden.

Da die videobasierte Prüfung neuartig war, interessierte vor allem die Frage, ob sich die Studierenden für die Beibehaltung der Prüfung aussprechen und wie sie die Prüfung im Gesamten sowie hinsichtlich einzelner Aspekte beurteilen. Dabei sollte abgefragt werden, wie der Praxisbezug und der Einsatz von Videos bewertet wird. Die für diese Arbeit relevanten Items beziehen sich somit vor allem auf die Dimension „Integration und Prozesse“ des AGIL-Schemas.

Es wurden konkret folgende Punkte abgefragt:

- 1) Die Studierenden sollten mit den dichotomen Antwortmöglichkeiten „ja“ und „nein“ die Frage „Sollte die VSE aus Ihrer Sicht beibehalten werden?“ beantworten.
- 2) Zudem sollten sie eine Gesamtnote für die VSE im Schulnotensystem von eins bis sechs vergeben.
- 3) Des Weiteren sollten die Studierenden das Item „In der videobasierten Überprüfung kommunikativer Fertigkeiten haben die Videos im Vergleich zu herkömmlichen schriftlichen Prüfungen ohne Videoeinsatz einen höheren Praxisbezug der Inhalte erzeugen können“ auf einer siebenstufigen Likert-Skala von 1 = „stimme gar nicht zu“ bis 7 = „stimme vollkommen zu“ bewerten.
- 4) Ebenfalls auf der vorgenannten Likert-Skala sollte zudem das Item „In der videobasierten Überprüfung kommunikativer Fertigkeiten wurde durch die Videos der Bezug zur Praxis des ärztlichen Gesprächs gut hergestellt“ gerankt werden.

Da die OSCE bereits bei vorangegangenen Untersuchungen erprobt und evaluiert worden war, wird für die vorliegende Arbeit nur in Kürze berichtet, dass die Studierenden auch für die OSCE eine Bewertung im Schulnotensystem von eins bis sechs abgeben sollten.

3.5 Entwicklung des Fragebogens zur Familienkommunikation

Zur Erfassung der Qualität der Familienkommunikation in der Ursprungsfamilie der Studierenden wurde ein eigener Fragebogen entwickelt.

Dies wurde als notwendig erachtet, da bei der Literaturrecherche nur Fragebögen ermittelt werden konnten, die nicht genau zum vorliegenden Untersuchungsgegenstand passten.

Bestehende Fragebögen wurden beispielsweise vor dem Hintergrund spezifischer Fragestellungen entwickelt und waren entsprechend auf die Bedürfnisse der jeweiligen Untersuchung/Fragestellung zugeschnitten oder erfassten teilweise nur ähnliche Konstrukte, aber nicht genau Familienkommunikation.

In den Fragebogen für den Kontext dieser Arbeit flossen die beiden Items aus der Untersuchung von Fischbeck (73) ein.

Zudem wurden einige Items aus dem Fragebogen von Fitzpatrick und Ritchie (81, 82, 92) entnommen bzw. in Anlehnung an diesen entwickelt.

Items, die inhaltlich den Aspekten Intensität und Offenheit für Problemmunikation nach Fischbeck zuzuordnen sind, finden sich auch in dem Fragebogen von Fitzpatrick und Ritchie wieder. Darüber hinaus beziehen sich die Items nach Fitzpatrick und Ritchie auf die Themen „Gefühle frei zum Ausdruck bringen“, „ernst genommen werden“, „gemeinsames Treffen von Entscheidungen“ und „Zuhören“ (Dimension „Gesprächsorientierung“). Alle genannten Aspekte sind in die Items für den Fragebogen der vorliegenden Arbeit eingeflossen.

Die Aspekte „Empathie“, ein „respektvoller und wertschätzender Umgangston“, die „angemessene Äußerung von Kritik“, das „Ausmaß an Streit und Auseinandersetzungen“ sowie das „offene Austragen von Konflikten“ wurden als Items zudem für den in dieser Arbeit genutzten Fragebogen hinzugenommen. In dieser Form sind sie nicht Teil des Fragebogens nach Fitzpatrick und Ritchie. Die Anpassungen erschienen erforderlich, da die Items im Fragebogen von Fitzpatrick und Ritchie in Bezug auf „Übereinstimmungsorientierung“ stark auf die Abfrage eines traditionellen Familientyps zugeschnitten sind und die Frage, inwieweit in der Ursprungsfamilie Hierarchien vorlagen nicht in allen Facetten optimal die Kommunikationsaspekte erfasst, die für diese Untersuchung und die Arzt-Patient-Kommunikation relevant sind.

Somit wurde in Anlehnung an die Lerninhalte des Kursus „Arzt-Patient-Kommunikation“ gezielt nach den oben genannten Aspekten „Empathie“, ein „respektvoller und wertschätzender Umgangston“, die „angemessene Äußerung von Kritik“, das „Ausmaß an Streit und Auseinandersetzungen“ sowie das „offene Austragen von Konflikten in der Familie“ gefragt.

Für den Fragebogen dieser Untersuchung wurde sich für eine vierstufige Likert-Skala mit den Antwortmöglichkeiten „trifft nicht zu“, „trifft etwas zu“, „trifft eher zu“ und „trifft genau zu“ entschieden. Es wurde keine mittlere Kategorie genutzt, um eine Tendenz des Raters zu forcieren. Der Fragebogen besteht aus 12 Items, die genauen Formulierungen sind im Anhang aufgeführt.

3.6 Statistische Auswertung

In Bezug auf die beiden videobasierten Prüfungen wurden zunächst die Häufigkeiten richtiger und falscher Antworten, die Schwierigkeitsindizes p_i , die Trennschärfen r_{it} sowie zudem Median und Interquartilsabstand über alle Items und Studierenden hinweg ermittelt.

Cronbachs α als Maß der internen Konsistenz der Skala wurde für alle Prüfungen (VSEs und OSCE) und die Fragebögen zur Familienkommunikation berechnet.

Um die Zusammenhänge zwischen den Prüfungsleistungen der Studierenden in der VSE und der OSCE sowie zwischen den Prüfungsleistungen und den Einschätzungen zur Familienkommunikation der Studierenden zu ermitteln, wurden die Rangkorrelationskoeffizienten nach Spearman betrachtet. Kolmogorov-Smirnov-Tests wurden zur Prüfung der Normalverteilung der Variablen eingesetzt.

Eine iterative Hauptachsenanalyse wurde jeweils mit den Prüfungsitens der VSEs beider Semester durchgeführt.

Dabei galt es Folgendes zu berücksichtigen: die Prüfungsitens sind einerseits dichotome Variablen, andererseits ergaben sich teilweise sehr geringe oder sehr hohe Itemschwierigkeiten (vgl. Tab. 2 und 3) und somit ungleiche Randsummenverteilungen. Daher wurden Odds Ratios zwischen allen Items berechnet. Da Odds Ratios Werte zwischen Null und Unendlich aufweisen, wurden sie anhand der von Yules vorgeschlagenen Formel auf Werte zwischen -1 bis +1 transformiert. Die transformierten Odds Ratios bildeten die Matrix, mit der die iterative Hauptachsenanalyse durchgeführt wurde. Yules Y wurde mit der Formel $Y = (\sqrt{OR} - 1) / (\sqrt{OR} + 1)$ ermittelt (93-95).

Für die bereits bewährte und erprobte OSCE wurden nur in Kürze der Median und der Interquartilsabstand über alle Items und Studierenden hinweg ermittelt.

In Bezug auf die Items der Evaluation erfolgte entweder jeweils die Betrachtung von Median und Interquartilsabstand oder die prozentuale Häufigkeitsangabe der dichotomen Variable.

Für die Fragebögen zur Familienkommunikation wurden in Bezug auf die Items jeweils der Median und Interquartilsabstand ermittelt und zudem wurde für beide Semester eine iterative Hauptachsenanalyse gerechnet.

Zudem wurde mit den Variablen Familienkommunikation, Alter und Geschlecht ein lineares gemischtes Modell in Bezug auf die VSE im SS18 und in Bezug auf die VSE und die OSCE im WS18/19 gerechnet. Die Normalverteilung der Residuen wurde anhand von Kolmogorov-Smirnov-Tests überprüft. Da teilweise keine Normalverteilung vorlag, wurde in diesen Fällen Bootstrapping angewandt.

4. Ergebnisse

4.1 Itemcharakteristiken VSEs

Tabelle 2 und 3 zeigen die Itemcharakteristiken der beiden VSEs. Die meisten Studierenden beantworteten die Mehrzahl der Items richtig. Die Trennschärfen der Items der VSEs bewegten sich in einem Bereich von -.04 bis .44.

4.2 Ergebnisse VSEs und OSCE

Zur Untersuchung der Faktorenstruktur der Videoprüfung wurde in Bezug auf die Kriterien von Rost und Schermer (96) die Anzahl relevanter Faktoren extrahiert (Hypothese 1).

Für das Sommersemester 18 wurden 13 Eigenwerte, die größer als eins sind, ermittelt. Bei Betrachtung des Screeplots ergab sich allerdings nur ein einziger Faktor. Bei einer Beschränkung auf Variablen für diesen Faktor, für deren Kommunalität $h^2 \geq .16$ gilt und deren Absolutbetrag $a \geq .40$ ist (vgl. Tabelle 4), verblieben noch 8 Items. Inhaltlich sind diese allerdings kaum ausreichend sinnvoll in Bezug auf einen Faktor hin zu interpretieren. Hingegen lassen 5 der Items (Item 4, 11, 13, 16, 19) eine gemeinsame inhaltliche Ausrichtung unter dem Aspekt „patientenorientiertes, einfühlsames Agieren und Verstehen anhand der Situation des Patienten und der Gesprächsinhalte“ (Cronbachs $\alpha = .62$, $n = 226$) erkennen.

Im WS18/19 lagen 12 Eigenwerte über 1 und es zeigte sich ebenfalls nur ein Faktor anhand des Screeplots. 14 Items erfüllten dann noch die Kriterien nach Rost und Schermer (vgl. Tabelle 4). 10 der 14 Items passen dabei inhaltlich zu dem Thema „patientenorientiertes, einfühlsames Agieren und Verstehen anhand der Situation des Patienten und der Gesprächsinhalte“ aus dem vorherigen Semester (3, 5, 7, 9, 12, 23, 24, 25, 26, 30; Cronbachs $\alpha = .60$, $n = 193$). Zudem überschneiden sich vier weitere Items mit dem Thema, allerdings finden sich auch Überschneidungen zu anderen Themen, sodass diese Items nicht noch zusätzlich eindeutig dem Faktor zugeordnet werden können. 4 der 5 Items aus dem SS18 laden im WS18/19 erneut auf dem Faktor (Items 23-26 WS18/19).

In Bezug auf beide Semester soll zudem noch herausgestellt werden, dass sich eine hohe Vielfalt an Themen und Aspekten zeigte, sodass die dargelegten Zuordnungen nicht als vollständig und komplett scharf beschrieben werden können.

Tabelle 2: Itemcharakteristiken der VSE im SS18 (n = 226)

VSE SS18				
	richtig	falsch	Schwierigkeitsindex p_i	Trennschärfe r_{it}
Item (Oberthemen)				
Items Handlungswissen				
4 Patientenorientierung/ Empathie (bei Non-Compliance) *	202	24	89.38	.30
11 Patientenorientierung/ Empathie *	220	6	97.35	.33
16 Patientenorientierung/ Empathie *	213	13	94.25	.31
19 Patientenorientierung/ Empathie (bei Visite) *	214	12	94.69	.27
1 Verständliche Informationsvermittlung	212	14	93.81	.16
7 Stressanalyse	109	117	48.23	.14
9 SPIKES-Protokoll	208	18	92.04	.10
12 Partizipative Entscheidungsfindung	159	67	70.35	.18
15 Verhaltensmodifikation	139	87	61.50	.08
20 Visite, Struktur, Informierung	224	2	99.12	.14
25 Umgang mit Nebenwirkungen	135	91	59.73	.08
27 Gesprächsförderer	225	1	99.56	-.04
30 Verständliche Informationsvermittlung	57	169	25.22	-.01
Items Faktenwissen				
13 Patientenorientierung/ Empathie (Schulz von Thun) *	219	7	96.90	.17
2 Anamnese	187	39	82.74	.15
3 Fragetypen	169	57	74.78	.20
5 Anamnese	150	76	66.38	.14
6 Metakommunikation	100	126	44.25	.21
8 Aktives Zuhören	204	22	90.27	-.01
10 Nonverbale Kommunikation	185	41	81.86	.26
14 Modell „Stages of Change“	160	66	70.80	.18
17 Beurteilungsfehler	220	6	97.35	.04
18 Partizipative Entscheidungsfindung	187	39	82.74	.19
21 NURSE-Modell	210	16	92.92	.09
22 Asymmetrisches Vermeidungsverhalten	195	31	86.28	.21
23 Theorie zu Empathie	20	206	8.85	.11
24 Transaktionales Stressmodell	126	100	55.75	.13
26 Patientenorientierung/ Verhaltensmodifikation	193	33	85.40	.26
28 ICF-Modell der WHO	113	113	50.00	.21
29 Motivational Interviewing	115	111	50.88	.12

Anmerkung. Die mit einem * markierten Items wurden in beiden Prüfungen (SS18 und WS18/19) eingesetzt. Bei den anderen Items handelt es sich prüfungsübergreifend teilweise zudem um dasselbe Oberthema, allerdings ist die inhaltliche Ausgestaltung der Aufgabe und der Frage verschieden. Richtig und falsch sind Häufigkeitsangaben.

Tabelle 3: Itemcharakteristiken der VSE im WS18/19 (n = 193)

VSE WS18/19				
Item (Oberthemen)	richtig	falsch	Schwierigkeitsindex p_i	Trennschärfe r_{it}
Items Handlungswissen				
22 Patientenorientierung/ Empathie (bei Non-Compliance) *	174	19	90.16	.06
23 Patientenorientierung/ Empathie *	189	4	97.93	.17
25 Patientenorientierung/ Empathie *	190	3	98.45	.30
26 Patientenorientierung/ Empathie (bei Visite) *	181	12	93.78	.26
2 Risikokommunikation	89	104	46.11	.15
3 Patientenorientierung/ Empathie (Positive Wertschätzung)	184	9	95.33	.36
5 Patientenorientierung/ Empathie	175	18	90.67	.16
7 Patientenorientierung/ Empathie (Motivational Interviewing)	155	38	80.31	.29
8 Verständliche Informationsvermittlung	169	24	87.56	.07
9 Patientenorientierung/ Empathie (Asymmetrisches Vermeidungsverhalten)	147	46	76.17	.32
10 Partizipative Entscheidungsfindung	32	161	16.58	.19
14 Anamnese	155	38	80.31	.18
15 Anamnese	166	27	86.01	.22
20 Compliance	179	14	92.75	.24
21 Motivational Interviewing	182	11	94.30	.30
29 Aktives Zuhören	122	71	63.21	.24
30 Patientenorientierung/ Empathie (‘schwieriger’ Patient)	182	11	94.30	.44
Items Faktenwissen				
24 Patientenorientierung/ Empathie (Schulz von Thun) *	186	7	96.37	.26
1 SPIKES-Protokoll	86	107	44.56	.08
4 Operante Konditionierung	145	48	75.13	.26
6 Modell „Stages of Change“	150	43	77.72	.04
11 Patientenrolle im Allgemeinkrankenhaus	147	46	76.17	.26
12 Patientenorientierung/ Empathie	179	14	92.75	.34
13 Arzt-Patienten-Beziehung	180	13	93.26	.24
16 Beurteilungsfehler	185	8	95.85	.23
17 NURSE-Modell	189	4	97.92	.12
18 Fragetypen	161	32	83.41	.14
19 Nonverbale Kommunikation	157	36	81.35	.11
27 Verhaltensmodifikation	119	74	61.66	.05
28 Konzept Kontrollüberzeugung	101	92	52.33	.04

Anmerkung. Die mit einem * markierten Items wurden in beiden Prüfungen (SS18 und WS18/19) eingesetzt. Bei den anderen Items handelt es sich prüfungsübergreifend teilweise zudem um dasselbe Oberthema, allerdings ist die inhaltliche Ausgestaltung der Aufgabe und der Frage verschieden. Richtig und falsch sind Häufigkeitsangaben.

Tabelle 4: Hauptachsenanalysen der VSEs im SS18 und WS18/19, Einfaktorenlösungen, Markieritems

Item	VSE SS18		VSE WS18/19	
	Ladung a	Kommunalität h ²	Ladung a	Kommunalität h ²
	Faktor: Patientenorientierung/ Empathie		Faktor: Patientenorientierung/ Empathie	
1	.56	.31		
2				
3			.75	.56
4	.64	.40		
5			.42	.18
6				
7			.44	.19
8				
9			.54	.29
10	.45	.21		
11	.85	.72		
12			.65	.42
13	.66	.44	.50	.25
14				
15				
16	.81	.65	.63	.40
17				
18				
19	.66	.44		
20			.52	.27
21			.61	.37
22				
23			.50	.25
24			.58	.34
25			.76	.58
26	.47	.22	.60	.36
27				
28				
29				
30			.77	.59

Anmerkung. Angegeben sind die Ladungen der Items für die gilt $h^2 \geq .16$ und Absolutbetrag $a \geq .40$.

Zur Betrachtung des Zusammenhangs zwischen VSE und OSCE (Hypothese 2) wurde der Rangkorrelationskoeffizient nach Spearman ermittelt. Dabei ergab sich ein Wert von $r = .21$ ($p \leq .01$).

Die beiden Variablen waren nicht normalverteilt (Kolmogorov-Smirnov-Test OSCE $D(192) = .09$ und VSE WS18/19 $D(192) = .15$, jeweils $p \leq .01$).

Bei der Betrachtung des Zusammenhangs der Items der VSE, die zu Handlungswissen gehörig konzipiert waren, und dem OSCE-Ergebnis (Hypothese 3), ergab sich zudem ein Rangkorrelationskoeffizient nach Spearman von $r = .25$ ($p \leq .01$).

Die Variable Handlungswissen war ebenfalls nicht normalverteilt (Kolmogorov-Smirnov-Test $D(192) = .19$, $p \leq .01$).

In Bezug auf die Fragestellung nach der Höhe der internen Konsistenz wurde für die VSE im SS18 ein Cronbachs-Alpha-Koeffizient von $\alpha = .55$ ($n = 226$) und im WS18/19 von $\alpha = .62$ ($n = 193$) ermittelt (in Bezug auf Hypothese 4).

Für die OSCE lag die interne Konsistenz bei $\alpha = .60$ ($n = 195$).

Der Median der VSE im SS18 betrug $Mdn = 23$, $IQR = 3$ für $n = 226$, für die VSE im WS18/19 lagen die Werte bei $Mdn = 25$, $IQR = 3$ für $n = 192$ und für die OSCE bei $Mdn = 26$, $IQR = 3$ für ebenfalls $n = 192$.

4.3 Ergebnisse Evaluation

In Tabelle 5 sind die Ergebnisse der Evaluation der VSEs und damit die Ergebnisse zu Hypothese 5 aufgeführt.

Es zeigte sich, dass sich die Studierenden in beiden Semestern mehrheitlich für eine Fortführung der VSE aussprachen.

Im Schulnotensystem wurde die VSE im SS18 mit einer Gesamtnote von $Mdn = 2$, $IQR = 1$ und damit als „gut“ sowie im WS18/19 mit einer Gesamtnote von $Mdn = 3$, $IQR = 1$ und damit als „befriedigend“ bewertet. Der Praxisbezug der VSEs wurde als relativ hoch eingeschätzt.

Die Evaluation der OSCE im WS18/19 ergab zudem eine Gesamtnote von $Mdn = 2$, $IQR = 1$ und wurde somit von den Studierenden als „gut“ bewertet ($n = 145$).

Tabelle 5: *Ergebnisse Evaluation*

Item	SS18				WS18/19			
	n	Mdn	IQR	Ja	n	Mdn	IQR	Ja
Sollte die VSE aus Ihrer Sicht beibehalten werden?	142			78 %	153			74 %
Gesamtnote der VSE (Schulnotensystem 1-6)	143	2	1		138	3	1	
In der videobasierten Überprüfung kommunikativer Fertigkeiten haben die Videos im Vergleich zu herkömmlichen schriftlichen Prüfungen ohne Videoeinsatz einen höheren Praxisbezug der Inhalte erzeugen können.	143	6	1		138	5	2	
In der videobasierten Überprüfung kommunikativer Fertigkeiten wurde durch die Videos der Bezug zur Praxis des ärztlichen Gesprächs gut hergestellt.	143	6	1		138	5	2	

Anmerkung. Insgesamt n = 147 Teilnehmer der Evaluation im SS18 und n = 192 im WS18/19. Item 3 und 4 in der Tabelle wurden anhand einer siebenstufigen Likert-Skala von 1 = „stimme gar nicht zu“ bis 7 = „stimme vollkommen zu“ bewertet.

4.4 Ergebnisse in Bezug auf Familienkommunikation, Alter und Geschlecht

In Tabelle 6 werden zunächst jeweils der Median und der Interquartilsabstand der Items der beiden Fragebögen zur Familienkommunikation aus dem SS18 und dem WS18/19 aufgeführt.

Eine iterative Hauptachsenanalyse des Fragebogens zur Familienkommunikation im SS18 ergab nur einen Eigenwert größer 1 und auch nach dem Screeplot zeigte sich nur ein Faktor. In Tabelle 7 sind die Ladungen und Kommunalitäten aufgeführt. Da Item 10 als einziges Item mit einer im Vergleich zu allen anderen Items geringeren Ladung und Kommunalität herausstach und dabei auch inhaltlich am wenigsten passend erschien, wurde dieses Item nicht weiter genutzt (für die weiteren Berechnungen und die Befragung im WS18/19). Für

das SS18 lag die interne Konsistenz bei $\alpha = .94$ ($n = 132$) und für das WS18/19 bei $\alpha = .94$ ($n = 150$).

Tabelle 6: Median und IQR der Items der Fragebögen zur Familienkommunikation im SS18 und WS18/19

Item	Fragebogen SS18		Fragebogen WS18/19	
	Mdn	IQR	Mdn	IQR
1	4	1	3	1
2	4	1	4	1
3	3	1	3	2
4	3	1	3	2
5	3	1	4	1
6	3	2	3	1
7	4	1	4	1
8	4	1	3	1
9	4	1	4	1
10	3	1	-	-
11	3	2	3	2
12	3	1	3	2

Anmerkung. Insgesamt $n = 132$ Teilnehmer der Befragung im SS18 und $n = 150$ im WS18/19. Alle Items in der Tabelle wurden anhand einer vierstufigen Likert-Skala mit 1 = „trifft nicht zu“, 2 = „trifft etwas zu“, 3 = „trifft eher zu“ und 4 = „trifft genau zu“ bewertet.

Tabelle 7: Hauptachsenanalysen der Fragebögen zur Familienkommunikation im SS18 und WS18/19, Einfaktorenlösungen

Item	Fragebogen SS18		Fragebogen WS18/19	
	Ladung a	Kommunalität h^2	Ladung a	Kommunalität h^2
1	.74	.55	.73	.53
2	.81	.66	.79	.63
3	.70	.48	.68	.46
4	.75	.56	.77	.59
5	.83	.69	.84	.70
6	.77	.59	.73	.53
7	.81	.65	.76	.58
8	.85	.73	.79	.62
9	.70	.48	.71	.51
10	.54	.29	-	-
11	.75	.57	.80	.64
12	.71	.50	.77	.60

Eine iterative Hauptachsenanalyse des Fragebogens zur Familienkommunikation im WS18/19 ergab zwei Eigenwerte größer 1, wobei der höchste Eigenwert 6.80 betrug und der zweite Eigenwert über 1 bei 1.02 lag. Entsprechend dieses deutlich geringeren zweiten Wertes, der knapp an der Kriteriumsgrenze von 1 lag, ergab sich allerdings weder auf mathematischer (bei Betrachtung der Korrelationen der Items in Bezug auf die zwei Faktoren) noch auf inhaltlicher Ebene ein klarer zweiter Faktor, sodass der einfaktoriellen Lösung Vorrang gegeben wurde. Die Ladungen und Kommunalitäten für die einfaktorielle Lösung sind in Tabelle 7 aufgeführt.

Zur Untersuchung des Zusammenhangs zwischen VSE und der Familienkommunikation (vgl. Hypothese 6a und 6b) wurde der Rangkorrelationskoeffizient nach Spearman in Bezug auf die VSE und den Fragebogen zur Familienkommunikation im SS18 ermittelt. Es ergab sich ein Wert von $r = .06$ ($p = .48$).

Für die beiden Items von Fischbeck (73) einzeln betrachtet zeigte sich für das Item „wurde intensiv miteinander gesprochen“ ein Rangkorrelationskoeffizient nach Spearman von $r = .06$ ($p = .47$) und für das Item „konnte ich offen über meine Probleme und Anliegen sprechen“ von $r = .11$ ($p = .23$).

Die Variable Familienkommunikation im SS18 war nicht normalverteilt (Kolmogorov-Smirnov-Test $D(132) = .15$, $p \leq .01$).

Der Rangkorrelationskoeffizient nach Spearman in Bezug auf die VSE und den Fragebogen zur Familienkommunikation im WS18/19 ergab einen Wert von $r = .01$ ($p = .92$). In Bezug auf die OSCE und den Fragebogen zur Familienkommunikation im WS18/19 wurde zudem ein Rangkorrelationskoeffizient nach Spearman von $r = .10$ ($p = .31$) ermittelt.

Für die beiden Items von Fischbeck (73) einzeln betrachtet zeigte sich für das Item „wurde intensiv miteinander gesprochen“ ein Rangkorrelationskoeffizient nach Spearman von $r = .06$ ($p = .52$) und für das Item „konnte ich offen über meine Probleme und Anliegen sprechen“ von $r = .13$ ($p = .18$) in Bezug auf die OSCE. In Bezug auf die VSE wurde für das Item „wurde intensiv miteinander gesprochen“ ein Rangkorrelationskoeffizient nach Spearman von $r = .06$ ($p = .56$) und für das Item „konnte ich offen über meine Probleme und Anliegen sprechen“ von $r = .10$ ($p = .31$) ermittelt.

Die Variable Familienkommunikation im WS18/19 war nicht normalverteilt (Kolmogorov-Smirnov-Test $D(100) = .13$, $p \leq .01$).

Eine Übersicht über alle vorgenannten Rangkorrelationskoeffizienten nach Spearman ist in Tabelle 8 zusammengestellt.

Tabelle 8: Rangkorrelationskoeffizienten nach Spearman zwischen VSEs, OSCE und dem Fragebogen zur Familienkommunikation (Fragebogen gesamt sowie Item 1 und 2 einzeln)

	VSE SS18	VSE WS18/19	OSCE WS18/19
Fragebogen gesamt	r = .06 p = .48	r = .01 p = .92	r = .10 p = .31
Item 1 „wurde intensiv miteinander gesprochen“	r = .06 p = .47	r = .06 p = .56	r = .06 p = .52
Item 2 „konnte ich offen über meine Probleme und Anliegen sprechen“	r = .11 p = .23	r = .10 p = .31	r = .13 p = .18

Anmerkung. Item 1 und 2 sind für eine übersichtlichere Präsentation der Ergebnisse verkürzt formuliert.

Das lineare gemischte Modell ergab jeweils mit den abhängigen Variablen VSE und OSCE sowie Alter, Geschlecht und Familienkommunikation als festen Effekten und den Studierenden, aufgeteilt in unterschiedliche Kurse, als zufälligem Effekt die in Tabelle 9 bis 11 aufgeführten Werte (vgl. [Hypothese 6c](#)).

Aufgrund von verletzten Annahmen zur Normalverteilung wurde in Bezug auf die VSE im SS18 und in Bezug auf die OSCE im WS18/19 Bootstrapping angewandt. Die Ergebnisse der Kolmogorov-Smirnov-Tests waren $D(69) = .12$ ($p = .02$) für die Residuen in Bezug auf die VSE im SS18, für die VSE im WS18/19 $D(97) = .07$ ($p = .20$) und für die OSCE im WS18/19 $D(97) = .10$ ($p = .02$).

Für das SS18 ergab ein Nullmodell zunächst ein Akaike Informationskriterium von $AIC = 332.38$, unter Hinzunahme der Variablen Familienkommunikation, Alter und Geschlecht von $AIC_{\text{gesamtes Modell}} = 325.75$. Für das WS18/19 in Bezug auf die VSE wurde im Nullmodell zudem zunächst ein Akaike Informationskriterium von $AIC = 955.88$ ermittelt und unter Hinzunahme der Variablen Familienkommunikation, Alter und Geschlecht reduzierte sich der Wert auf $AIC_{\text{gesamtes Modell}} = 474.16$. In Bezug auf die OSCE ergab sich im Nullmodell $AIC = 853.94$, mit den Variablen $AIC_{\text{gesamtes Modell}} = 470.53$.

Anhand von Streudiagrammen wurden Linearität und Homoskedastizität geprüft. Die Voraussetzungen zur Anwendung des linearen gemischten Modells waren erfüllt.

Tabelle 9: *Ergebnisse des linearen gemischten Modells VSE SS18 (mit Bootstrapping)*

	Estimate	SE	p
Intercept	21.73	2.95	≤.01
Fragebogen zur Familienkommunikation	-.03	.04	.50
Alter	-.003	.09	.98
Geschlecht	1.49	.69	.04

Anmerkung. Die signifikanten Ergebnisse sind hervorgehoben.

Tabelle 10: *Ergebnisse des linearen gemischten Modells VSE WS18/19*

	Estimate	SE	df	T	p
Intercept	26.61	2.20	92.92	12.08	≤.01
Fragebogen zur Familienkommunikation	-.05	.04	91.34	-1.25	.22
Alter	-.16	.07	91.56	-2.34	.02
Geschlecht	1.79	.59	92.96	3.05	≤.01

Anmerkung. Die signifikanten Ergebnisse sind hervorgehoben.

Tabelle 11: *Ergebnisse des linearen gemischten Modells OSCE WS18/19 (mit Bootstrapping)*

	Estimate	SE	p
Intercept	24.69	2.07	≤.01
Fragebogen zur Familienkommunikation	-.005	.03	.88
Alter	.01	.07	.84
Geschlecht	.78	.61	.22

Anmerkung. Die signifikanten Ergebnisse sind hervorgehoben.

5. Diskussion

Das Ziel der vorliegenden Arbeit, eine videobasierte Prüfung zur Messung ärztlicher kommunikativer Fertigkeiten von Medizinstudierenden zu entwickeln und zu erproben, konnte erreicht werden. Auch der Vergleich mit der OSCE sowie die Zusammenhänge mit der Qualität der Familienkommunikation sowie dem Alter und Geschlecht der Studierenden konnten untersucht werden. Die videobasierte Prüfung wurde u.a. anhand von Aufgaben im Stil eines Situational Judgement Tests (47, 49, 50) ausgearbeitet, bei der OSCE wurde auf bereits bewährte Prüfungsstationen zurückgegriffen (35). Im SS18 absolvierten die Studierenden im zweiten Semester die videobasierte Prüfung, im WS18/19 die videobasierte Prüfung und die OSCE. Zudem erfolgten jeweils die Studierendenbefragungen und die Evaluationen der Prüfungen. Im Folgenden sollen die erworbenen Ergebnisse diskutiert sowie Limitationen und mögliche Anpassungen betrachtet werden.

5.1 Diskussion der Ergebnisse in Bezug auf VSEs und OSCE

Die erste Annahme dieser Arbeit, dass eine zweifaktorielle Lösung der beiden VSEs entsprechend dem theoretischen Konzept von Fakten- und Handlungswissen nach Miller (46) und nach Krathwohl (53), basierend auf Bloom (52), und in Bezug auf videobasierte Ansätze entsprechend nach Trost et al. (57), Rottmann et al. (58) und Ludwig (64) denkbar ist (Hypothese 1), konnte nicht bestätigt werden.

Die statistischen Ergebnisse lassen auf ein sehr heterogenes Themenfeld schließen. Mit Einschränkungen (Überschneidungen mit anderen Themen, zum Teil keine optimale inhaltliche Passung) kann in beiden Semestern allenfalls ein Bereich hervorgehoben werden, der inhaltlich grob die Themen Patientenorientierung und Empathie umfasst. Ansonsten beziehen sich die Items anscheinend auf viele verschiedene Aspekte.

Setzt man diese Ergebnisse nochmal in Bezug zu den Kursinhalten und der Konstruktion der VSE erscheint dies auch durchaus nachvollziehbar. Bereits das Konstrukt ärztliche kommunikative Kompetenz wurde eingangs als sehr vielschichtig und komplex beschrieben (22, 23). Es wurde dargelegt, dass in das Konzept des Kursus verschiedene psychologische Kommunikationsmodelle (vgl. 2.1) und verschiedene Lernziele eingeflossen sind. Entsprechend dem Prinzip Constructive Alignment (69, vgl. 2.8) spiegelten sich die Kursinhalte/Lernziele auch in der Prüfung wider, sodass die Inhalte der Prüfung breit gefasst sind und in einem gelehrten und geprüften Oberthema vermutlich oft verschiedene Wissensinhalte und verschiedene Fertigkeiten benötigt werden.

Es ist denkbar, dass das heterogene Themenfeld eine Einteilung in Fakten- und Handlungswissen überlagert. Dies bedeutet, Items, die zu Faktenwissen gezählt werden können, beziehen sich auf so unterschiedliche Themen, Inhalte, Lernziele, Konstrukte und Modelle innerhalb der ärztlichen kommunikativen Kompetenz, dass diese bei der Zuordnung zu Faktoren entscheidend sind und nicht deren zusätzlich theoretisch mögliche Zuordnung zu Fakten- und Handlungswissen.

Gleichzeitig erscheint es plausibel, dass die Themen Patientenorientierung und Empathie sehr basale Aspekte des Kursus und des Konstrukts ärztliche kommunikative Kompetenz darstellen und in mehreren Aufgaben und Themen wiederzufinden sind. Fischer (97) beschreibt hierzu, dass in der Fachwelt Einigkeit darüber besteht, wie essenziell Empathie in der Patientenversorgung ist. Auch Pander et al. (98) stellen Empathie als bedeutenden Teil ärztlicher kommunikativer Kompetenz heraus. In einer Studie von Herzig und Koerfer (99) wurden zudem ärztliche Schlüsselkompetenzen untersucht. Dabei erfolgte eine Zusammenstellung von neun Schlüsselkompetenzen, wobei Fachkompetenz, Handlungskompetenz, Empathie und Patientenorientierung 2/3 der Gesamtkompetenz ausmachten. Von diesen vier ragten wiederum Fachkompetenz und Empathie nochmal heraus. Diese Darlegungen und Befunde passen auch zu den Regeln nach Rogers (12), nach denen Empathie wesentlich für eine gelingende Kommunikation ist.

Angemerkt werden soll zudem, dass im WS18/19 deutlich mehr Items dem bereits dargelegten Faktor zugeordnet werden konnten, was vermutlich dadurch zu erklären ist, dass auch der Schwerpunkt der VSE im WS18/19 etwas mehr auf Handlungswissen bzw. Aufgaben zur Gesprächsführung gelegt wurde, die auch konzeptuell einen höheren Bezug zu Patientenorientierung und Empathie haben sollten.

Für zukünftige Prüfungen bedeuten diese Ergebnisse, dass nochmal sehr bewusst entschieden werden sollte, wie genau die Schwerpunktsetzung der Themen gewählt wird.

Eine weitere Anmerkung soll an dieser Stelle noch in Bezug auf die statistischen Kennwerte der VSEs erfolgen. Es zeigte sich, dass die Items insgesamt eher leicht waren und eine niedrige Trennschärfe aufwiesen. Um eine bessere Differenzierung in der Kompetenzbewertung der Studierenden realisieren zu können, sollten die Items in zukünftigen Prüfungen hinsichtlich ihres Schweregrades besser verteilt sein.

In Bezug auf den angestrebten Vergleich zwischen VSE und OSCE zeigte sich wie erwartet ein leicht positiver, signifikanter Zusammenhang zwischen der Prüfungsleistung der Studierenden in der VSE und der OSCE (Hypothese 2). Dieses Ergebnis deckt sich mit

früheren Ergebnissen von Humphris und Kaney (62) sowie von Fischbeck et al. (68) zum Zusammenhang von OSCE und OSVE und OSCE und Key-Feature-Prüfung.

Somit ist anzunehmen, dass die Inhalte der VSE und der OSCE zu einem gewissen Grad Überschneidungen aufweisen und offenbar durchaus in Teilen dasselbe Konstrukt messen. Gleichzeitig zeigt die Höhe der Korrelation aber auch an, dass sich scheinbar auch ein größerer Teil der gemessenen Inhalte bzw. des gemessenen Konstrukts unterscheidet.

Zu diesem Ergebnis würde die vorgestellte Einteilung und theoretische Einordnung der Prüfungen in Wissens Ebenen nach Miller (46) und nach Krathwohl (53) passen, wonach die OSCE auf der Ebene von „shows how“/„Laborbedingungen“ anzusiedeln ist und sich die VSE auf die Ebene von Fakten- und Handlungswissen bezieht. Besonders relevant könnte dieser Unterschied bei Items der OSCE sein, die von einem Rater direkt beobachtet werden müssen und Aspekte wie Körperposition oder Blickkontakt abdecken. Diese Items konnten in der VSE nicht umgesetzt werden und sind somit nicht Teil dieser Prüfungsform.

Bei anderen Items der VSE ist durchaus eine größere Annäherung an Items der OSCE möglich. Sowohl bei der VSE als auch bei der OSCE muss der Studierende beispielsweise die Schritte im Prozess der Partizipativen Entscheidungsfindung kennen. Während bei der VSE die Wahl des nächsten Schrittes nach einem gezeigten Video aus verschiedenen Antworten heraus erfolgen muss, zeigen die Studierenden im Rollenspiel mit dem Simulationspatienten, dass sie diesen Schritt ebenfalls kennen, führen ihn aber zusätzlich aus. Somit gibt es eine Schnittmenge, aber auch einen Unterschied.

Des Weiteren kommt in Bezug auf die gemessene Korrelation zum Tragen, dass die Bandbreite der Themen der VSE größer angelegt war als die Bandbreite der Themen der OSCE und auch dies die geringe Höhe der Korrelation mitbedingen kann. Diesbezüglich wäre eine erneute Untersuchung mit einer VSE, die sich nur auf die Inhalte der fünf OSCE-Stationen beschränkt, interessant. Möglicherweise ließe sich die Korrelation dadurch steigern.

Sich an die Betrachtung des Zusammenhangs zwischen VSE und OSCE anschließend, soll der Zusammenhang zwischen der Prüfungsleistung in den auf Handlungswissen bezogenen Items der VSE und der OSCE-Prüfungsleistung in die Diskussion hinzugenommen werden. Es wurde erwartet, dass ein höherer Zusammenhang zwischen der Prüfungsleistung in den auf Handlungswissen bezogenen Items der VSE und der OSCE-Prüfungsleistung als zwischen der Gesamtprüfungsleistung der VSE (Fakten- und Handlungswissen) und der OSCE-Prüfungsleistung besteht (Hypothese 3).

Wie bereits dargelegt, hat sich eine zweifaktorielle Lösung mit den beiden Variablen Fakten- und Handlungswissen nicht ergeben. Unter Berücksichtigung der theoretisch zugrunde gelegten Zuordnung in Fakten- und Handlungswissen war der Zusammenhang zwischen den theoretisch für Handlungswissen konzipierten Items der VSE und der OSCE-Prüfungsleistung nur unwesentlich höher als der Zusammenhang zwischen allen Items der VSE und der OSCE.

Dieses Ergebnis soll erneut vor dem Hintergrund der hohen Heterogenität der Themen des Kursus, der unterschiedlichen Bandbreite an Themen der Prüfungen und der Komplexität des Konstrukts ärztliche kommunikative Kompetenz eingeordnet werden. Wie bereits dargestellt, beziehen die Items vermutlich verschiedene Themen, Fertigkeiten und Aspekte des Kursus und des Konstrukts mit ein. Es erscheint, wie bereits weiter oben diskutiert, gut vorstellbar, dass die Breite und Heterogenität der Items eine Einteilung in Fakten- und Handlungswissen überlagert. Damit wäre es nachvollziehbar, wieso zwischen den theoretisch Handlungswissen zugeordneten Items und der OSCE-Prüfungsleistung kein höherer Zusammenhang besteht als zwischen allen Items der VSE und der OSCE.

Als weiterer Punkt der vorliegenden Arbeit wurde die Höhe der internen Konsistenz der Prüfungen untersucht. Entgegen der Annahme (Hypothese 4) fiel die interne Konsistenz der VSEs jeweils gering aus (Einordnung Cronbachs $\alpha > .7$ akzeptabel, darunter $> .6$ fragwürdig und $< .5$ inakzeptabel, 100). Auch die Annahme, dass die VSE aufgrund ihres Konzepts mit standardisierten Aufgaben, standardisierter, softwaregestützter Auswertung und der Anzahl von 30 Items der OSCE hinsichtlich der internen Konsistenz überlegen sein könnte, bestätigte sich nicht.

Die Diskussion dieses Ergebnisses soll sich an die bisherige Diskussion anschließen. Da das Themenfeld ärztliche kommunikative Kompetenz sehr breit aufgestellt ist und die Items wie dargestellt offenbar vergleichsweise wenig inhaltliche Schnittmengen haben, sind sehr hohe interne Konsistenzen als Maß für die Zusammenhänge der Items vor diesem Hintergrund offenbar grundsätzlich schwierig zu erzielen.

Erfolgt auch hier nochmal eine eingehende Betrachtung der jeweiligen Prüfungsinhalte wird deutlich, dass die VSE in der umgesetzten Form und entsprechend ihrem Konzept möglichst breit die Kursthemen abdeckt, wohingegen die OSCE mit der Begrenzung auf die fünf Stationen in der Breite limitierter ist. Bei der VSE sollten sich Lernziele und Prüfungsinhalte möglichst jeweils wiederfinden lassen (vgl. Constructive Alignment, 69). Die Anzahl an Items war aber offensichtlich nicht hoch genug, um ausreichend viele Items zu bestimmten

Oberthemen in die Prüfung inkludiert zu haben, die zumindest theoretisch in Bezug auf verschiedene Faktoren eine höhere interne Konsistenz erzielen können.

Eine Erhöhung der internen Konsistenz der VSEs wäre somit theoretisch über eine Erhöhung der Anzahl der Items zu bestimmten Oberthemen oder wie bei der OSCE ebenfalls über eine Begrenzung auf weniger Themenfelder umsetzbar.

Dies kann auch als grundsätzliche Frage in Bezug zu Schecker (101) verstanden werden, ob bei der Konstruktion eines Tests der Schwerpunkt eher auf die Bandbreite oder eher auf die Spezifität gelegt wird. Schecker legt dar, dass eine konsistente Skala leichter zu erreichen ist, wenn die Items nur aus einem einzigen Themengebiet stammen. Gleichzeitig sollte mit Blick auf die Fachdidaktik bei einem komplexen Themengebiet möglichst eine größere Bandbreite abgedeckt sein.

Es wird somit deutlich, dass bei der Zusammenstellung zukünftiger videobasierter Prüfungen diese Entscheidung erneut bewusst abgewogen und gefällt werden sollte. Nach Schecker wären die erzielten geringen internen Konsistenzen der beiden VSEs allerdings gut tolerierbar, wenn aus fachdidaktischen Gesichtspunkten der Bandbreite vor der Höhe der internen Konsistenz Vorrang gegeben wird.

5.2 Diskussion zu den Ergebnissen der Evaluation

Anhand der Evaluation durch die Studierenden sollte in dieser Arbeit zudem untersucht werden, wie die Studierenden den Praxisbezug der Videoprüfung einschätzen. Die Studierenden gaben in der Evaluation an, dass durch die Videos ein guter Praxisbezug hergestellt werden konnte. Sie gaben auch an, dass durch die Videos ein höherer Praxisbezug als bei rein schriftlichen Prüfungen realisiert wurde. Die Hypothese 5 konnte somit bestätigt werden.

Dieses Ergebnis deckt sich mit positiven Rückmeldungen in anderen Untersuchungen (Trost et al., 57, Ludwig, 64).

Eine mögliche Einschränkung könnte sein, ob Studierende vielleicht aufgrund der Neuheit von Videos und damit einhergehendem höheren Interesse zu einer positiven Bewertung gelangten (102). Persike berichtet allerdings von ersten Längsschnittstudien, die eher gegen diese Annahme sprechen, d.h. dass die positiven Effekte und Einschätzungen nach aktuellem Stand tendenziell eher nicht auf die Neuheit des Videoeinsatzes zurückzuführen sind.

Kritisch anzumerken ist, dass Studierende im zweiten Semester in der Regel noch nicht über viel praktische Berufserfahrung verfügen und deren Einschätzung optimalerweise noch durch Expertenurteile ergänzt werden sollte. Die Fälle und Dialoge wurden zwar einem Expertenurteil unterzogen, rückblickend wäre allerdings eine weitere Befragung bezüglich der fertiggestellten VSEs unter gleichzeitiger Vorlage einer rein schriftlichen Prüfung als Vergleichsmöglichkeit noch wünschenswert gewesen.

Des Weiteren gilt es noch zu bedenken, dass auch die Studierenden keine schriftliche Prüfung im Vergleich zur VSE im Bereich ärztliche kommunikative Kompetenz absolviert haben. Ein direkter Vergleich zwischen VSE und schriftlicher Prüfung sollte also ebenfalls noch untersucht werden.

5.3 Diskussion zu den Ergebnissen in Bezug auf Familienkommunikation, Alter und Geschlecht

Für den weiteren Teil dieser Arbeit, die Untersuchung des Zusammenhangs zwischen den Prüfungsleistungen und der Qualität der Familienkommunikation, des Alters und des Geschlechts, sollen die Ergebnisse im Folgenden diskutiert werden. Der Fragebogen zur Familienkonstruktion zeigte eine sehr zufriedenstellende interne Konsistenz (Einordnung Cronbachs $\alpha > .9$ exzellent, $> .8$ gut, 100). Eine einfaktorielle Lösung erscheint zudem inhaltlich plausibel. Die beiden Dimensionen von Fitzpatrick und Ritchie wurden, wie unter 3.5 dargestellt, in der Konzeption und Formulierung der Items nicht umgesetzt. Die Items des Fragebogens für diese Arbeit beziehen sich alle theoretisch auf die Dimension „Gesprächsorientierung“. Diese eine inhaltliche Ausrichtung scheint sich in den statistischen Ergebnissen widerzuspiegeln.

Die Annahme, dass eine von Offenheit und intensivem Austausch geprägte Familienkommunikation bzw. dass weitere Facetten der Familienkommunikation (wie respektvoller Umgangston, gemeinsame Entscheidungsfindung in Anlehnung an „The Revised Family Communication Pattern Instrument“) mit den Prüfungsleistungen zusammenhängen (Hypothese 6a und 6b), hat sich nicht gezeigt.

Dieses Ergebnis ist aufgrund der vorliegenden Ergebnisse vor allem von Fischbeck (73), aber auch weiterer empirischer Befunde (vgl. 2.10) und theoretischer Überlegungen überraschend. In Bezug auf die VSE mit ihren Unterschieden zur OSCE wäre das Ergebnis

noch eher nachvollziehbar (keine direkt beobachtbaren Verhaltensweisen, höhere Bandbreite der Themen), in Bezug auf die OSCE hingegen zunächst weniger.

Weitere Untersuchungen zu dieser Fragestellung sind sicherlich notwendig, um zu eruieren, ob der Zusammenhang bei Fischbeck oder das Ergebnis in dieser Arbeit repliziert werden können.

Für die beiden videobasierten Prüfungen VSE (im SS18 und WS18/19) zeigte sich zudem ein signifikanter Einfluss des Geschlechts auf die Prüfungsleistungen (weibliche Studierende erzielten bessere Prüfungsleistungen als männliche Studierende). Zusätzlich ergab sich ein signifikanter Effekt für das Alter auf die Leistung in der VSE im WS18/19 (ältere Studierende schnitten schlechter ab als jüngere Studierende). Für die OSCE im WS18/19 und die VSE im SS18 zeigte sich dies hingegen nicht. Für die OSCE im WS18/19 zeigte sich zudem kein Effekt des Geschlechts auf die Prüfungsleistung. Somit konnte die Annahme, dass das Geschlecht einen relevanten Einfluss auf die Prüfungsleistungen hat, nur für zwei der drei Prüfungen bestätigt werden. Dass das Alter keinen Einfluss auf die Prüfungsleistung hat, zeigte sich in dieser Arbeit hingegen für eine Prüfung nicht, für die beiden anderen Prüfungen allerdings schon (Hypothese 6c).

Die Ergebnisse sind somit heterogen und inkonsistent. Das Ergebnis zum Geschlecht in Bezug auf die VSE im SS18 und im WS18/19 deckt sich mit zahlreichen Vorbefunden (73, 86-89), wonach weibliche Studierende in Prüfungen besser abschneiden als männliche Studierende. Für die OSCE im WS18/19 konnte dieser Befund allerdings nicht repliziert werden. Bezüglich des Alters deckt sich das Ergebnis für die VSE im SS18 und die OSCE im WS18/19 mit den berichteten Vorbefunden, für die VSE im WS18/19 hingegen nicht.

Zunächst ist es mit Blick auf die Familienkommunikation denkbar, dass der Bereich ärztlicher kommunikativer Kompetenz hoch spezifisch ist und damit die in der Familie gelernten Beziehungsschemata keine gute Passung für die Arzt-Patient-Kommunikation darstellen. Für verschiedene Arzt-Patienten-Interaktionen, die im Kursus gelehrt und geprüft werden, gibt es beispielsweise Modelle und enge Vorgaben (wie beim SPIKES-Protokoll, 30, oder dem NURSE-Modell, 29), die speziell für diese Art der Beziehung entwickelt wurden und die ggf. eher als Teil eines eigenen, neu gelernten Beziehungsschemas angesehen werden könnten.

Zudem sind neben der Familienkommunikation, dem Geschlecht und dem Alter viele weitere Determinanten für die ärztliche kommunikative Kompetenz denkbar und teilweise auch bereits untersucht worden. In Studien zum generellen Studienerfolg im Bereich Medizin

wurden Variablen wie Staatsbürgerschaft, Berufsausbildungserfahrungen, Lernstil und Persönlichkeitseigenschaften als Determinanten ermittelt (86, 103, 104).

Relevant sind zusätzlich Inhalte, die in der weiteren Dissertation zum vorliegenden Projekt untersucht werden, wie Lernprozess und Studierverhalten, Lebensbedingungen, Studienbedingungen oder Eingangsvoraussetzungen. Um in einer Prüfung, wie auch in der VSE und der OSCE gut abschneiden zu können, muss der Studierende der Annahme nach zudem beispielsweise funktionale Lernstrategien für die Inhalte, Strategien während der Prüfung zur Einteilung von Aufgaben, zur Bewältigung von Stress und Druck, Einordnungs- und Schlussfolgerungsfähigkeiten und viele Fertigkeiten mehr vorweisen können.

Es ist somit denkbar, dass andere Aspekte als Familienkommunikation entscheidend für die Prüfungsergebnisse waren.

Bezüglich des Alters und des Geschlechts erscheint eine Interpretation aufgrund der uneinheitlichen Ergebnisse schwierig. Der in der Literatur berichtete Befund, dass weibliche Studierende besser abschneiden als männliche Studierende, muss sicherlich als durchschnittliches Ergebnis angesehen werden, dass nicht in Bezug auf jede Stichprobe und in jedem Kontext Gültigkeit besitzt. Hier greift vermutlich erneut, dass verschiedene Determinanten für die Prüfungsleistung relevant sind und andere Aspekte in den Stichproben die entscheidenden Einflussgrößen für das Prüfungsergebnis darstellen könnten. Auffällig ist, dass die weiblichen Studierenden in den schriftlichen videobasierten Prüfungen besser abschnitten als männliche Studierende, nicht aber in der praktischen Prüfungsform OSCE. Eventuell existiert eine grundsätzliche, diesbezügliche Tendenz, die untersucht werden müsste. Möglicherweise bedingen in einer praktischen Prüfung aber auch noch andere Variablen den Prüfungserfolg, die den Einfluss des Geschlechts reduzieren oder sogar irrelevant werden lassen. Es erscheint auch plausibel, dass das Alter teilweise eine Rolle spielen kann, aber nicht immer spielen muss. Die Richtung des Effekts bei der videobasierten Prüfung im WS18/19 lässt vermuten, dass jüngere Studierende unter bestimmten Umständen einen Vorteil gegenüber älteren Studierenden haben. Eventuell könnte beispielsweise eine höhere kognitive Leistungsfähigkeit oder eine bessere Erinnerungsleistung diesen Effekt beeinflussen, gleichzeitig könnten aber auch eine höhere Lebenserfahrung oder ein höherer Lerneinsatz den Einfluss des Alters kompensieren (105). Zusätzlich gilt es zu bedenken, dass die untersuchten Stichproben homogen bezüglich der Altersverteilung sind und zu wenige ältere Studierende in die Untersuchungen eingeschlossen wurden, um hier verlässlichere Ergebnisse erzielen zu können.

Zukünftige Arbeiten sollten idealerweise die angesprochenen Aspekte weiter und eingehender untersuchen.

In Bezug auf die Methodik soll zudem noch angemerkt werden, dass sich die Antwortskala der Items in dieser Arbeit und in der Untersuchung von Fischbeck in Bezug auf die Familienkommunikation unterscheiden. Während bei Fischbeck eine fünfstufige Likert-Skala zum Einsatz kam, wurde für diese Arbeit eine vierstufige Likert-Skala gewählt. Inwieweit dies einen Einfluss auf die Ergebnisse hatte, ist schwer abzuschätzen.

Das Auslassen einer mittleren Kategorie könnte zu einem Messfehler geführt haben, bei dem Studierende, die tatsächlich eine neutrale Ausprägung aufweisen, nicht erfasst wurden (106, 107).

Zudem gibt es einige Arbeiten, die aufzeigen, dass eine fünf- bis siebenstufige Skala optimal hinsichtlich Reliabilität und Validität zu sein scheint (Übersicht über diese Arbeiten in 108). Die hier gewählte vierstufige Likert-Skala könnte somit unter diesen Gesichtspunkten weniger reliabel und valide gemessen haben, eventuell entstand dadurch auch ein Einfluss auf den untersuchten Zusammenhang.

Um die genannten Aspekte als mögliche Einflussgrößen auszuschließen, sollte in zukünftigen Untersuchungen die Wahl der Likert-Skala nochmal überdacht werden.

Abschließend soll noch die Überlegung angestellt werden, welche Bedeutung ein nicht vorhandener Einfluss der Familienkommunikation auf die Prüfungsleistung hinsichtlich der Chancengleichheit der Studierenden hätte. Würde sich in zahlreichen Untersuchungen zeigen, dass kein relevanter Zusammenhang besteht, würde dies bedeuten, dass Studierende, die in Familien mit einer schlechten Kommunikationsqualität aufgewachsen sind, nicht durch diesen unverschuldeten Nachteil schlechter in den Prüfungen abschneiden müssen als Studierende, die in einer Familie mit einer besseren Kommunikationsqualität aufgewachsen sind. Der Erkenntnis von Kurtz et al. (21) folgend, dass kommunikative Kompetenzen erlernbar sind, würde dies dafürsprechen, dass alle Studierenden im Kursus zumindest hinsichtlich des Themas Qualität der Familienkommunikation in gleichem Maße die Chance haben, die Kompetenzen zur ärztlichen kommunikativen Gesprächsführung zu erwerben und den Kursus erfolgreich zu absolvieren. Eine Kompensation fehlender erworbener Kompetenzen aus der Familie wäre somit nicht notwendig.

5.4 Bewertung und Limitationen in Bezug auf Videos und das Prüfungsformat

Die Produktion der Videos gelang reibungslos in Zusammenarbeit mit dem Medienzentrum und unter Einsatz von Schauspielern, die die Rollen verkörperten. Die erzielte Qualität kann durch die Studierendenbewertung als gut angenommen werden. Durch die Förderung des

Projekts durch das BMBF waren die finanziellen Mittel für die Umsetzung vorhanden. Der Umfang des Videomaterials ist ausreichend, um für weitere Prüfungen videobasierte Prüfungsfragen erstellen zu können. Für den Einsatz videobasierter Prüfungen abseits von geförderten Projekten stellt sich allerdings die Frage der Finanzierbarkeit von Videoproduktionen (vgl. auch Trost et al., 57), zumal eine hohe Qualität, die in der Regel mit Kosten verbunden ist, sichergestellt werden sollte. Hierfür könnte in Bezugnahme zu Kopp et al. (48) eine nationale Datenbank eine Idee sein, um Szenarien und auch Videos zu teilen.

Für das Format der Videoprüfung sind zudem andere Herangehensweisen und Antwortformen denkbar. Für diese Arbeit wurde sich hauptsächlich aufgrund einer hohen Ökonomie und Objektivität für Antworten in Single-Choice-Form entschieden, allerdings hat diese Form natürlich auch Nachteile. Ansätze wie bei Baribeau et al. (23) oder Humphris und Kaney (62), bei denen Videoausschnitte kommentiert werden müssen bzw. offene Antwortformen zum Einsatz kamen, erfordern ein höheres Maß an Eigenleistung und teilweise auch an Transferleistung durch die Studierenden. Gleichzeitig sind auch solche Prüfungen, wie bereits dargelegt, theoretisch genauso wie Single-Choice-Formate auf der Ebene von Handlungswissen einzuordnen (vgl. Fischbeck, 68 oder Ludwig, 64) und können auch nicht die Ebene des direkt beobachtbaren Verhaltens wie bei einer OSCE adressieren (nach Miller, 46, Krathwohl, 53).

Bezüglich des Einsatzes von Videos wurde in dieser Arbeit des Weiteren bisher vorwiegend auf die Vorteile eingegangen. Allerdings sollen, damit ein vollumfängliches Bild und eine umfassende Bewertung möglich sind, auch die Nachteile des Einsatzes von Videos gesammelt beleuchtet und diskutiert werden.

Bereits angesprochen wurden die Kosten für die Produktion und der Einwand nach Ludwig (64) in Bezug auf die Theorie der kognitiven Belastung beim Lernen (65), dass Videos für den eigentlich relevanten Sachverhalt zu viele unwesentliche Informationen enthalten und damit den Lernerfolg behindern könnten. Da sich diese Überlegung allerdings auf den Lernerfolg bezieht, ist unklar, inwieweit dieser Aspekt für die Prüfung eine wichtige Rolle spielt. Von Bedeutung erscheint hierbei die Frage, ob zu viele Informationen in der Prüfungssituation vorhanden sind und ob diese Studierende eventuell von den wesentlichen Inhalten in Bezug auf das abzuprüfende Lernziel und die Prüfungsfrage ablenken. Hierzu wären weitere Untersuchungen wünschenswert. Gleichzeitig argumentiert Ludwig, dass auch andere Prüfungsformate deutlich mehr Informationen enthalten als textbasierte Formate, insbesondere die reale Interaktion mit einem Patienten (und damit auch eine OSCE, die mit Simulationspatienten ebenfalls eine komplexe Prüfungssituation mit vielen Informationen und auch möglichen Störquellen schafft). Zudem betont Kok, dass über eine gute Passung von

verbalen und visuellen Inhalten das Maß an Belastung gesteuert werden kann (66). Für die vorliegende Arbeit wird davon ausgegangen, dass die Passung hoch und der Anteil an irrelevanter Information gering gewesen ist. Die Videos zeigten jeweils einen Gesprächsausschnitt, bei dem das Setting jeweils passend zum Inhalt des Gesprächs gewählt wurde und der Fokus lag permanent auf dem Gespräch und den Sprechenden. Ein hohes Maß an zusätzlichen, irrelevanten Informationen liegt so nach eigener Einschätzung nicht vor.

Der Punkt, dass durch die Art der Kameraeinstellung, die Wahl des Bildausschnitts oder Schnitte generell eine nicht gewollte Lenkung oder Beeinflussung stattfinden könnte, wurde im Rahmen dieses Projekts berücksichtigt und durch ein Split-Screen-Verfahren zu reduzieren versucht (vgl. Kapitel 3). Jedoch stellt auch dieses Thema einen Bereich dar, der bei der Entscheidung für oder gegen den Einsatz von Videos Beachtung finden und abgewogen werden sollte.

Die Abhängigkeit von Technologie stellt einen weiteren zu berücksichtigenden Aspekt bei der Verwendung von Videos dar. An den Hochschulen müssen die technischen Voraussetzungen gegeben sein, gleichzeitig besteht während der laufenden Prüfung das Risiko eines Ausfalls oder einer eingeschränkten Nutzbarkeit der Prüfungssoftware und der Videos (109). Wie dargestellt, ergaben sich nach guter Vorbereitung in Zusammenarbeit mit dem Zentrum für Datenverarbeitung der Johannes Gutenberg-Universität Mainz nahezu keine Schwierigkeiten bei der Umsetzung der Videoprüfungen im Rahmen dieses Projekts. Allerdings sollte bei der Planung und Umsetzung einer technisch anspruchsvolleren Prüfung als der rein textbasierten Prüfungsform auf die Gegebenheiten und die technischen Möglichkeiten vor Ort geachtet werden. Eine Überlastung durch die hohe Anzahl an Videozugriffen sollte beispielsweise vermieden werden. Grundsätzlich kann angenommen werden, je technischer ein Ablauf wird, desto mehr Fehlerquellen erscheinen denkbar. Allerdings sollten die Voraussetzungen für die Umsetzung einer Videoprüfung an den meisten Hochschulen vorliegen, sodass dieser Punkt nicht zu schwer ins Gewicht fallen dürfte.

Die Darstellung möglicher Nachteile bei der Verwendung von Videos erhebt nicht den Anspruch auf Vollständigkeit, ggf. sind weitere Aspekte zu bedenken, um insgesamt abschließend eine sachgerechte Abwägung zwischen den Vor- und Nachteilen der Verwendung von Videos vornehmen und eine Entscheidung hinsichtlich des Einsatzes von Videos treffen zu können.

5.5 Abschließende Betrachtung der Prüfungsformen

In einer abschließenden Betrachtung kann festgehalten werden, dass die VSEs und die OSCE erfolgreich durchgeführt werden konnten. Die Entwicklung und Erprobung der VSE hat einige Ressourcen gekostet (Entwicklung und Umsetzung des Konzepts, Videoproduktion, Einbindung der Videos in Software). Theoretisch birgt die VSE durch diese Vorarbeit aber nun wie dargestellt die Möglichkeit, in zukünftigen Semestern mit weniger Aufwand als bei der ersten Entwicklung mit der VSE prüfen zu können (hohe Anzahl an produzierten Videos, ausreichend Material, um Prüfungssitems für zukünftige Klausuren entwickeln zu können). Zudem könnte eine nationale Datenbank zum Teilen von Materialien ein Ansatz sein.

Die Prüfung OSCE war an den Prüfungstagen selbst sehr aufwändig, der Ablauf verlief allerdings reibungslos.

Die dargestellten und diskutierten Ergebnisse führen zu dem Schluss, dass die VSE in ihrer vorgestellten Form und die OSCE in weiten Teilen nicht dieselben Inhalte messen und unterschiedlich zum Einsatz kommen sollten. In Bezug auf Miller (46) und Krathwohl (53) kann davon ausgegangen werden, dass die OSCE bei dem Wunsch nach der Messung der ärztlichen kommunikativen Kompetenz mit höherem Praxisbezug zum Einsatz kommen sollte und die VSE in ihrer vorgestellten Form die Ebene darunter, die kognitive Ebene, abfragt. Ob mit den verschiedenen, vorgestellten Modifikationen ein höherer Zusammenhang zwischen VSE und OSCE erreicht werden kann, müsste noch untersucht werden.

Entsprechend der Annahme, dass die VSE durch ihren Einsatz von Videos einen höheren Praxisbezug erzielen sollte als rein schriftliche Prüfungen, wäre der VSE aber auch schon in ihrer aktuellen Form hinsichtlich der Höhe des Praxisbezuges vor einer rein schriftlichen Prüfung Vorrang einzuräumen. Der direkte Vergleich zwischen VSE und einer rein schriftlichen Prüfung steht aber noch aus.

Die Ergebnisse dieser Arbeit scheinen zudem die Theorie zu untermauern, nach der die ärztliche kommunikative Kompetenz ein komplexes und vielschichtiges Feld ist. Es werden vermutlich verschiedenste Bereiche in einer Prüfung gemessen und erfasst, was bei der Konstruktion zukünftiger Prüfungen/der Zusammenstellung von Prüfungssitems, wie in der Diskussion dargestellt, bedacht und je nach Ziel (wie z.B. Bandbreite oder Messung einzelner Schwerpunkte, dann mit bestenfalls höherer interner Konsistenz) eine bewusste Entscheidung für ein Format getroffen werden sollte.

Es wäre wünschenswert, wenn das vorgestellte Prüfungskonzept oder Abwandlungen davon auch an anderen Hochschulen erprobt werden würde, um zuverlässigere und allgemeingültigere Aussagen zum Einsatz von Videoprüfungen im Feld der ärztlichen kommunikativen Kompetenzen im deutschsprachigen Raum treffen bzw. Empfehlungen dazu abgeben zu können.

6. Zusammenfassung

In der vorliegenden Arbeit wurde eine videobasierte Prüfung zur Messung der ärztlichen kommunikativen Kompetenz bei Medizinstudierenden im Fach Medizinische Psychologie und Medizinische Soziologie entwickelt und erprobt sowie hinsichtlich Ergebnis und Ökonomie mit einer bereits bewährten praktischen Prüfungsform OSCE in Beziehung gesetzt.

Es zeigte sich ein geringer, signifikanter Zusammenhang zwischen den Ergebnissen beider Prüfungsformen ($r = .21$, $p \leq .01$). Dies spricht dafür, dass die entwickelte videobasierte Prüfung (VSE, Video-based Single-Choice Examination) in den Prüfungsinhalten in der Mehrheit von den Inhalten der OSCE abweicht und diese nicht unbedingt gleichwertig einzusetzen sind. Die Annahme, dass die VSE Fakten- und Handlungswissen in zwei entsprechenden Dimensionen empirisch abbildet, bestätigte sich nicht.

In Erwägung des primären Prüfungsziels sollte die OSCE entsprechend theoretischen Überlegungen bei einem Schwerpunkt auf die direkt beobachtbare Handlungsebene zum Einsatz kommen, da sie vermutlich dem realen Arzt-Patienten-Kontakt am nächsten kommt. Die VSE hingegen kann diesen hohen Praxisbezug nicht in gleichem Maße leisten, ggf. aber durch den Einsatz von Videos eine stärkere Annäherung an die Praxis erzielen als rein schriftliche Prüfungen. Diese Annahme wurde von Studierenden in der Evaluation so rückgemeldet, sollte aber durch Expertenurteile und weiterführende Untersuchungen überprüft werden (direkter Vergleich zwischen VSE und schriftlicher Prüfung).

Des Weiteren sollten bei der Wahl der Prüfungsform die Aspekte Testgüte und Ressourcenaufwand Beachtung finden. In dieser Arbeit fielen die internen Konsistenzen beider Prüfungen gering aus (Cronbachs $\alpha = .55$ VSE im SS18, Cronbachs $\alpha = .62$ VSE im WS18/19 und Cronbachs $\alpha = .60$ OSCE im WS18/19). Die Objektivität der VSE dürfte augenscheinlich höher ausfallen als diejenige der OSCE. Dadurch, dass die VSE noch entwickelt werden musste und das OSCE-Konzept schon vorlag, erforderte die VSE auch einige Ressourcen (vor allem für die Videoproduktion). Für zukünftige Klausuren kann allerdings auf die entwickelten Inhalte zurückgegriffen werden, sodass sich der Ressourcenaufwand reduzieren und geringer als bei einer OSCE ausfallen sollte.

Aufgrund der ersten Erfahrungen mit der entwickelten videobasierten Prüfung lässt sich festhalten, dass einige Limitationen festzustellen sind und die Notwendigkeit besteht, bestimmte Punkte zu verbessern (Schwierigkeit der Items, weitere Überlegungen zur internen Konsistenz und entsprechende Umsetzung, direkter Vergleich mit einer rein

schriftlichen Prüfung). Gleichzeitig erscheint die videobasierte Prüfung aber auch sehr vielversprechend und kann eine wichtige Ergänzung im Pool an möglichen Prüfungsformen im Fach Medizinische Psychologie und Medizinische Soziologie darstellen. Die VSE kann als eine moderne, der fortschreitenden Digitalisierung Rechnung tragende Prüfung eingeschätzt werden, die im Vergleich zur OSCE insgesamt Ressourcen schont. Das Feedback der Studierenden zur Videoprüfung fiel insgesamt positiv aus und sie sprachen sich mehrheitlich für eine Beibehaltung dieser Prüfungsform aus.

Grundsätzlich sollte bei der Wahl der geeigneten Prüfungsform aus dem Pool an Konzepten auf die Anforderungen der jeweiligen Prüfung geachtet und dann je nach Bedarf und Abwägung der vorhandenen Ressourcen entschieden werden.

Zusätzlich wären zudem Erfahrungen mit dem Konzept an anderen Hochschulen sowie die Untersuchung von Modifikationen wünschenswert.

Darüber hinaus zeigte sich in der vorliegenden Arbeit kein signifikanter Zusammenhang zwischen der erlebten Kommunikationsqualität in der Ursprungsfamilie und den Prüfungsleistungen der Studierenden ($r = .06$, $p = .48$ VSE im SS18, $r = .01$, $p = .92$ VSE im WS18/19 und $r = .10$, $p = .31$ OSCE im WS18/19). Dieses Ergebnis deckt sich nicht mit einer vorherigen Untersuchung. Hinsichtlich des neben der Familienkommunikation zusätzlich betrachteten Einflusses von Alter und Geschlecht auf die Prüfungsleistungen ergaben sich zudem heterogene Befunde, wobei in zwei von drei Prüfungen weibliche Studierende besser abschnitten als männliche Studierende und in einer Prüfung jüngere Studierende bessere Ergebnisse erzielten als ältere Studierende. Sicherlich determinieren insgesamt zahlreiche Aspekte, Eigenschaften und Kompetenzen die Prüfungsleistung in Bezug auf ärztliche kommunikative Kompetenz, die in Abhängigkeit von Kontext und Stichprobe ggf. unterschiedlich zum Tragen kommen. Die insgesamt inkonsistenten Befunde zu möglichen Determinanten der Prüfungsleistungen sollten in weiteren Studien untersucht werden, um verlässlichere Aussagen und Schlussfolgerungen treffen zu können.

7. Literaturverzeichnis

1. Bundesministerium der Justiz. Approbationsordnung für Ärzte [Internet]. 2002, zuletzt geändert 2023 [zitiert am 18.07.2024]. URL: https://www.gesetze-im-internet.de/_appro_2002/BJNR240500002.html
2. Bundesministerium für Gesundheit. Beschlusstext zum "Masterplan Medizinstudium 2020" [Internet]. 2017 [zitiert am 18.07.2024]. URL: https://www.bmbf.de/SharedDocs/Downloads/files/2017-03-31_masterplan-beschlusstext.pdf?__blob=publicationFile&v=2
3. The Medical School Objectives Writing Group. Learning Objectives for Medical Student Education-Guidelines for Medical Schools: Report I of the Medical School Objectives Project. *Acad Med.* 1999;74(1):13-8.
4. UK Foundation Programme Office. The Foundation Programme Curriculum 2016 [Internet]. 2016 [zitiert am 24.07.2024]. URL: https://www.gmc-uk.org/-/media/documents/fp-curriculum-2016-v2_pdf-101343872.pdf
5. Rossmann C. Gesundheitskommunikation: Eine Einführung aus kommunikationswissenschaftlicher Perspektive. In: Rossmann C, Hastall MR, Hrsg. *Handbuch der Gesundheitskommunikation*. Wiesbaden: Springer Fachmedien; 2019: 3-14.
6. Maletzke G. *Kommunikationswissenschaft im Überblick. Grundlagen, Probleme, Perspektiven*. Opladen: Westdeutscher Verlag; 1998.
7. Röhner J, Schütz A. *Psychologie der Kommunikation. Basiswissen Psychologie*. 2. Aufl. Wiesbaden: Springer Fachmedien; 2016.
8. Krauss RM, Fussell SR. Social psychological models of interpersonal communication. In: Higgins ET, Kruglanski A, eds. *Social psychology. Handbook of basic principles*. New York: Guilford Press; 1996: 655-701.
9. Shannon CE, Weaver W. *The mathematical theory of communication*. Urbana, Illinois: University of Illinois Press; 1964.
10. Schulz von Thun F. *Miteinander reden 1-4*. Hamburg: Rowohlt; 2014.
11. Grice HP. Logic and conversation. In: Cole P, Morgan J, eds. *Syntax and semantics*. Volume 3. New York: Academic Press; 1975: 41-58.
12. Rogers CR. *Die klientenzentrierte Gesprächspsychotherapie. Client-Centered Therapy*. Frankfurt a. M.: Fischer Taschenbuch; 1983.
13. Watzlawick P, Beavin JH, Jackson DD. *Menschliche Kommunikation*. Bern: Huber, Hogrefe; 2017.
14. Rothenfluh F, Schulz PJ. Arzt-Patient-Kommunikation. In: Rossmann C, Hastall MR, Hrsg. *Handbuch der Gesundheitskommunikation*. Wiesbaden: Springer Fachmedien; 2019:

57-67.

15. Kempf D. Untersuchung der Gesprächszeit mit Patienten und Angehörigen unter Zugrundelegung der Arbeitszeitverteilung von Krankenhausärzten [Dissertation]. Freiburg: Fachbereich Medizin der Albert-Ludwigs-Universität; 2007.
16. Geisler LS. Das ärztliche Gespräch – eine vernachlässigte Aufgabe. Vortrag am 16.06.2003 in der Rheinischen Friedrich-Wilhelms-Universität, Bonn. Vortragsreihe „Ärztliche Ethik“ im Sommersemester 2003 im Rahmen des Studium Universale [Internet]. 2003 [zitiert am 18.07.2024]. URL: <http://www.linus-geisler.de/vortraege/0306kommunikation.html>
17. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet*. 2001;357(9258):757-62.
18. Harris SR, Templeton E. Who's listening? Experiences of women with breast cancer in communicating with physicians. *Breast J*. 2001;7(6):444-9.
19. Kerr J, Engel J, Schlesinger-Raab A, Sauer H, Hölzel D. Communication, quality of life and age: results of a 5-year prospective study in breast cancer patients. *Ann Oncol*. 2003;14(3):421-7.
20. Maguire P, Pitceathly C. Key communication skills and how to acquire them. *BMJ* 2002;325(7366):697-700.
21. Kurtz S, Silverman J, Draper J. Teaching and learning communication skills in medicine. 2nd. ed. Oxon: Radcliffe Publishing Ltd; 2005.
22. Aspegren K. Beme guide no. 2: Teaching and learning communication skills in medicine- a review with quality grading of articles. *Med Teach*. 1999;21(6):563–70.
23. Baribeau DA, Mukovozov I, Sabljic T, Eva KW, deLottinville CB. Using an objective structured video exam to identify differential understanding of aspects of communication skills. *Med Teach*. 2012;34(4):e242-50.
24. Schweickhardt A, Fritzsche K. Kursbuch ärztliche Kommunikation. Grundlagen und Fallbeispiele aus Klinik und Praxis. Köln: Deutscher Ärzte-Verlag; 2007.
25. Hamann J, Loh A, Kasper J, Neuner B, Spies C, Kissling W, et al. Partizipative Entscheidungsfindung. Implikationen des Modells des "Shared Decision Making" für Psychiatrie und Neurologie. *Nervenarzt*. 2006;77(9):1071-8.
26. Härtl A, Bachmann C, Blum K, Höfer S, Peters T, Preusche I, et al. Desire and reality-- teaching and assessing communicative competencies in undergraduate medical education in German-speaking Europe-a survey. *GMS Z Med Ausbild*. 2015;32(5):Doc56.
27. Kiessling C, Dieterich A, Fabry G, Hölzer H, Langewitz W, Mühlinghaus I, et al. Communication and social competencies in medical education in German-speaking countries: The Basel Consensus Statement. Results of a Delphi Survey. *Patient Educ Couns*. 2010;81(2):259-66.

28. Hahn EG, Fischer MR. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT). *GMS Z Med Ausbild.* 2009;26(3):Doc35.
29. Back AL, Arnold RM, Baile WF, Fryer-Edwards KA, Alexander SC, Barley GE, et al. Efficacy of communication skills training for giving bad news and discussing transitions to palliative care. *Arch Intern Med.* 2007;167(5):453–60.
30. 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(4):302–11.
31. Laidlaw A, Salisbury H, Doherty EM, Wiskin C. National survey of clinical communication assessment in medical education in the united kingdom (UK). *BMC Med Educ.* 2014;14:10.
32. Chong L, Taylor S, Haywood M, Adelstein B-A, Shulruf B. Examiner seniority and experience are associated with bias when scoring communication, but not examination, skills in objective structured clinical examinations in australia. *J Educ Eval Health Prof.* 2018;15:17.
33. Daniels VJ, Harley D. The effect on reliability and sensitivity to level of training of combining analytic and holistic rating scales for assessing communication skills in an internal medicine resident OSCE. *Patient Educ Couns.* 2017;100(7):1382-6.
34. Dong T, LaRochelle JS, Durning SJ, Saguil A, Swygert K, Artino AR. Longitudinal effects of medical students' communication skills on future performance. *Mil Med.* 2015;180(4 Suppl):24-30.
35. Fischbeck S, Mauch M, Leschnik E, Beutel ME, Laubach W. Überprüfung ärztlicher kommunikativer Kompetenz mittels einer OSCE bei Studierenden der Medizin im ersten Studienjahr. *Psychother Psychosom Med Psychol.* 2011;61:465–71.
36. Fischbeck S. Medizinische Psychologie und Medizinische Soziologie: Skriptum zum Kursus Teil II - Patientenorientierung in der Medizin, Arzt-Patient-Kommunikation. [Unveröffentlichtes Skriptum]. Mainz: Fachbereich Psychosomatische Medizin und Psychotherapie; 2017.
37. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ.* 1975;1(5955):447-51.
38. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ.* 1979;13(1):41-54.
39. Chenot J-F, Ehrhardt M. Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. *Z Allg Med.* 2003;79(9):437–42.
40. Harden RM. Revisiting "Assessment of clinical competence using an objective structured clinical examination (OSCE)". *Med Educ.* 2016;50(4):376–9.
41. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication

- skills in the OSCE format: reliability and generalizability. *Med Educ.* 1996;30(1):38-43.
42. Jünger J, Nikendei C. OSCE Prüfungsvorbereitung Innere Medizin. 1. Aufl. Stuttgart: Thieme; 2005.
43. Blume J-H, Perleth M, Busse R. Was kostet die Objective Structured Clinical Examination? *Med Ausbildung.* 2000;17:12-7.
44. Mossbrugger H, Kelava A. Testtheorie und Fragebogenkonstruktion. 2. Aufl. Berlin, Heidelberg: Springer-Verlag; 2012.
45. Krebs D, Menold N. Gütekriterien quantitativer Sozialforschung. In: Baur N, Blasius J, Hrsg. Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS; 2019: 489-504.
46. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):63-7.
47. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic attributes in selection. *Med Educ.* 2012;46(9):850–68.
48. Kopp V, Möltner A, Fischer MR. Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbildung.* 2006;23(3):Doc50.
49. Schubert S, Ortwein H, Dumitsch A, Schwantes U, Wilhelm O, Kiessling C. A situational judgement test of professional behaviour: development and validation. *Med Teach.* 2008;30(5):528–33.
50. Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Med Educ.* 2012;46(4):399–408.
51. Witheridge A, Ferns G, Scott-Smith W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. *Int J Med Educ.* 2019;10:191-2.
52. Bloom BS. Taxonomy of educational objectives. The classification of educational goals. Handbook 1, Cognitive Domain. London: Longmans, Green & Co. (Ltd.); 1956.
53. Krathwohl DR. A Revision of Bloom's Taxonomy. An Overview. *Theory Pract.* 2002;41(4):212-8.
54. Michel LP, Goertz L, Radomski S, Fritsch T, Baschour L. Digitales Prüfen und Bewerten im Hochschulbereich. Arbeitspapier Nr. 1. Berlin: Hochschulforum Digitalisierung; 2015.
55. Ehlers JP, Windt K-H, Tipold A. Workshop 12: Neue Möglichkeiten durch elektronische Prüfungen. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bochum, 23. 25.09.2010. Düsseldorf: German Medical Science GMS Publishing House; 2010.
56. Mayer RE. The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educ Psychol Rev.* 2024;36(8).

57. Trost S, Dieterle J, Herrmann M, Signerski-Krieger J. Videodarstellungen in psychiatrischen Prüfungsfragen. Eine Pilotstudie. *Nervenarzt*. 2020;91:635-41.
58. Rottmann J, Stratmann J, Kerres M. Handlungsorientiertes Prüfen in der beruflichen Aus- und Weiterbildung: Eine Herausforderung für computergestützte Testverfahren. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*. 2006;Occasional Papers:1-24.
59. Bandura, Albert. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, New Jersey: Prentice-Hall; 1986.
60. Findeisen S, Horn S, Seifried J. Lernen durch Videos – Empirische Befunde zur Gestaltung von Erklärvideos. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*. 2019;Occasional Papers:16-36.
61. Schmitz FM, Schnabel KP, Bauer D, Bachmann C, Woermann U, Guttormsen S. The learning effects of different presentations of worked examples on medical students' breaking bad-news skills: A randomized and blinded field trial. *Patient Educ Couns*. 2018;101(8):1439–51.
62. Humphris GM, Kaney S. The objective structured video exam for assessment of communication skills. *Med Educ*. 2000;34(11):939–45.
63. Hulsman RL, Mollema ED, Oort FJ, Hoos AM, Haes JCJM de. Using standardized video cases for assessment of medical communication skills: reliability of an objective structured video examination by computer. *Patient Educ Couns*. 2006;60(1):24–31.
64. Ludwig S. Effektivität videogestützter Key-Feature-Prüfungen beim Erwerb differentialdiagnostischer und -therapeutischer Kompetenzen im Medizinstudium [Dissertation]. Göttingen: Fachbereich Medizin der Georg-August-Universität; 2021.
65. Chandler P, Sweller J. Cognitive Load Theory and the Format of Instruction. *Cogn Instr*. 1991;8(4):293-332.
66. Kok A. Video-Tutorial versus textbasierte Anleitung. Eine vergleichende Untersuchung zu Interaktion und Lernerfolg am Beispiel von Einführungen in die Literaturverwaltungssoftware Citavi. In: Petras V, Hrsg. *Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft*. Heft 475. Berlin: Humboldt-Universität; 2020.
67. Mousavi SY, Low R, Sweller J. Reducing cognitive load by mixing auditory and visual presentation modes. *J Educ Psychol*. 1995;87(2):319-34.
68. Fischbeck S, Mauch M, Unterrainer J. Vom Kompetenzwissen zum praktischen Können: Arzt-Patient-Kommunikation prüfen mit Key-Feature-Klausuren und OSCE [Internet]. 2013 [zitiert am 25.08.2024]. URL: https://www.unimedizin-mainz.de/fileadmin/kliniken/verhalten/Dokumente/Fischbecketal_Postter_Dresden_DGMP180_913_Key_Feature.pdf

69. Biggs J. Enhancing teaching through constructive alignment. *High Educ.* 1996;32(3):347-64.
70. Wildt J, Wildt B. Lernprozessorientiertes Prüfen im „Constructive Alignment“. In: Berendt B, Voss H-P, Wildt J, Hrsg. *Neues Handbuch Hochschullehre, Teil H: Prüfungen und Leistungskontrollen. Weiterentwicklung des Prüfungssystems in der Konsequenz des Bologna-Prozesses.* Berlin: Raabe; 2011.
71. Parsons T. *Das System moderner Gesellschaften.* Weinheim, München: Beltz Juventa; 2009.
72. Schmidt U, Becker N, Lübbe H, Hümpfner M, Schmidt F, Yugay N, et al. Modellbasierte Lehrevaluation: Konzept und empirische Ergebnisse. In: Daniel H-D, Hümmert-Harris S, Heger M, Hornbostel S, Krempkow R, Mitterauer L et al., Hrsg. *Qualität der Wissenschaft.* Bielefeld: Universitätsverlag Webler; 2014.
73. Fischbeck S. Persönlichkeit, Einstellung und sozialer Hintergrund als Determinanten ärztlich kommunikativer Kompetenz bei Studierenden der Medizin in der Vorklinik [Unveröffentlichte Masterarbeit]. Mainz: Fachbereich Psychosomatische Medizin und Psychotherapie; 2014.
74. Epstein NB, Ryan CE, Bishop DS, Miller IW, Keitner GI. The McMaster model: a view of healthy family functioning. In: Walsh F, Hrsg. *Normal Family Processes: Growing Diversity and Complexity.* New York City: The Guilford Press; 2003: 581–607.
75. Koerner AF, Fitzpatrick MA. Toward a theory of family communication. *Communication Theory.* 2002;12(1):70–91.
76. Fletcher GJO. Cognition in close relationships. *New Zealand Journal of Psychology.* 1993;22(2):69-81.
77. Fletcher GJO, Thomas G. Close relationship lay theories: Their structure and function. In: Fletcher GJO, Fitness J, Eds. *Knowledge structures in close relationships: A social psychological approach.* Hillsdale, NJ: Lawrence Erlbaum; 1996.
78. Koesten J. Family communication patterns, sex of subject, and communication competence. *Commun Monogr.* 2004;71(2):226-44.
79. Chaffee SH, McLeod JM, Atkin CK. Parental influences on adolescent media use. *American Behavioral Scientist.* 1971;14(3):323-40.
80. Stone VA, Chaffee SH. Family communication patterns and source-message orientation. *Journalism Quarterly.* 1970;47(2):239-46.
81. Ritchie LD, Fitzpatrick MA. Family communication patterns: Measuring intrapersonal perceptions of interpersonal relationships. *Communication Research.* 1990;17(4):523–44.
82. Fitzpatrick MA, Ritchie LD. Communication schemata within the family: Multiple perspectives on family interaction. *Hum Commun Res.* 1994;20(3):275–301.
83. McLeod JM, Chaffee SR. The construction of social reality. In: JT Tedeschi, ed. *The*

social influence processes. Chicago, New York: Aldine Atherton; 1972.

84. Schrodt P, Witt PL, Messersmith AS. A Meta-Analytical Review of Family Communication Patterns and their Associations with Information Processing, Behavioral, and Psychosocial Outcomes. *Commun Monogr.* 2008;75(3):248–69.

85. Hojat M, Callahan CA, Gonnella JS. Students' personality and ratings of clinical competence in medical school clerkships: a longitudinal study. *Psychol Health Med.* 2004;9(2):247–52.

86. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ.* 2002;324(7343):952-7.

87. Erdel B. Welche Determinanten beeinflussen den Studienerfolg? Eine empirische Analyse zum Studienerfolg der ersten Kohorte der Bachelorstudenten in der Assessmentphase am Fachbereich Wirtschaftswissenschaften der Friedrich-Alexander-Universität Erlangen-Nürnberg [Internet]. 2010 [zitiert am 13.08.2024]. URL: https://www.ssoar.info/ssoar/bitstream/handle/document/22022/ssoar-2010-erdel-welche_determinanten_beeinflussen_den_studienerfolg.pdf?sequence=1&isAllowed=y&Inkname=ssoar-2010-erdel-welche_determinanten_beeinflussen_den_studienerfolg.pdf

88. Arrenberg J, Kowalski S. Lernen Frauen und Männer unterschiedlich? Eine Studie über das Lernverhalten von Studierenden [Internet]. 2007 [zitiert am 17.08.2024]. URL: <https://fh-koeln.arrenberg.com/pdf/gender-artikel-formatiert.pdf>

89. Saravani S, Dehghan Haghighi J, Abed-Saeedi Z, Homayouni Zand R. An examination of effective factors on academic success of students of Zehedan Medical University. *Future of medical education journal.* 2013;3(3):35-40.

90. Kanfer FH, Reinecker H, Schmelzer D. *Selbstmanagement-Therapie.* 5. Aufl. Berlin, Heidelberg: Springer; 2012.

91. Hoppe-Seyler T, Gartmeier M, Möller G, Bauer J, Wiesbeck A, Karsten G. Entwicklung von Lehrfilmen zur Gesprächsführung zwischen Realitätsnähe und systematischer didaktischer Gestaltung. *ZFHE.* 2014;9(3):127–35.

92. Koerner FA, Fitzpatrick MA. Understanding family communication patterns and family functioning: The roles of conversation orientation and conformity orientation. *Ann Int Commun Assoc.* 2002;26(1):36-65.

93. Edwards AWF. The Measure of Association in a 2 × 2 Table. *Journal of the Royal Statistical Society. Series A (General).* 1963;126(1):109–14.

94. Leonhart R. *Lehrbuch Statistik.* 2. Aufl. Bern: Huber; 2009.

95. Wermuth N. How can graphical Markov models aid causal inference? In: *DAGStat 2019. Statistics under one umbrella. Conference Guide.* München, 18.-22. März 2019.

Göttingen: Deutsche Arbeitsgemeinschaft Statistik; 2019.

96. Rost DH, Schermer FJ. Strategien der Prüfungsangstverarbeitung. *Z Diff D P.* 1986;7(3):127–43.
97. Fischer A-S. Kommunikative Kompetenz im Medizinstudium: Relevanz und Erwerb von Empathie in der medizinischen Ausbildung [Dissertation]. Ulm: Fachbereich Psychosomatische Medizin und Psychotherapie des Universitätsklinikum; 2019.
98. Pander T, Kiessling C, Görlitz A, Fischer MR. Eine qualitative Erhebung typischer Situationen aus dem klinisch-praktischen Alltag zur ärztlichen Empathie. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Graz, 26.-28.09.2013. Düsseldorf: German Medical Science GMS Publishing House; 2013.
99. Herzig S, Koerfer A. Ärztliche Schlüsselkompetenzen. In: Koerfer A, Albus C, Hrsg. Kommunikative Kompetenz in der Medizin. Göttingen: Verlag für Gesprächsforschung; 2018: 229-61.
100. Blanz M. Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen. 2. Aufl. Stuttgart: Kohlhammer; 2021.
101. Schecker H. Überprüfung der Konsistenz von Itemgruppen mit Cronbachs alpha. chapter online-Zusatzmaterial. In: Krüger D, Parchmann I, Schecker H, Hrsg. Methoden in der naturwissenschaftsdidaktischen Forschung. Berlin: Springer Spektrum; 2014.
102. Persike M. Videos in der Lehre: Wirkungen und Nebenwirkungen. In: Niegemann H, Weinberger A, Hrsg. Lernen mit Bildungstechnologien. Berlin, Heidelberg: Springer Reference Psychologie; 2019.
103. Doherty EM, Nugent E. Personality factors and medical training: a review of the literature. *Med Educ.* 2011;45(2):132-40.
104. Adam J, Bore M, McKendree J, Munro D, Powis D. Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Med Educ.* 2012;12:69.
105. Kruse A. Bildung im Alter. In: Tippelt R, von Hippel A, Hrsg. Handbuch Erwachsenenbildung/Weiterbildung. 5. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften; 2011: 827-840.
106. Stier W. Empirische Forschungsmethoden. 2. Aufl. Berlin, Heidelberg: Springer; 1999.
107. Cox EP. The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research.* 1980;17(4):407-22.
108. Faulbaum F, Prüfer P, Rexroth M. Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität. Wiesbaden: VS Verlag für Sozialwissenschaften; 2009.
109. Bisang D, Bäuerlein K, Wyss C. Ein Video-Portfolio als summative Prüfung am Ende der Lehrpersonenausbildung. *DiMawe - Die Materialwerkstatt.* 2022;4(3):40-48.

8. Anhang

8.1 Fragebogen Familienkommunikation

Im Folgenden werden Sie darum gebeten, Ihren Grad der Zustimmung zu verschiedenen Aussagen bezüglich der von Ihnen wahrgenommenen Familienkommunikation anzugeben. Bitte beziehen Sie sich dabei auf Ihre Ursprungsfamilie, in der Sie während Ihrer Kindheit und Jugend aufgewachsen sind. Sollten Sie nicht in einem klassischen Familienverbund oder in wechselnden Familien/unterschiedlichen Familienzusammenstellungen aufgewachsen sein, beziehen Sie sich bitte auf den Familienverbund in dem Sie vorwiegend aufgewachsen sind bzw. beziehen Sie sich bitte auf die Kommunikation, die in Ihrem Umfeld während des Aufwachsens vorwiegend vorherrschte.

In meiner Ursprungsfamilie...	Trifft nicht zu	Trifft etwas zu	Trifft eher zu	Trifft genau zu
1. ... wurde intensiv miteinander gesprochen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. ... konnte ich offen über meine Anliegen und Probleme sprechen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. ... wurden Gefühle frei zum Ausdruck gebracht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. ... wurde ich dazu ermutigt meine Gefühle frei zum Ausdruck zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. ... wurde jedem zugehört.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. ... wurden Entscheidungen gemeinsam getroffen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. ... wurden meine Äußerungen ernst genommen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. ... galt es als wichtig, sich in die anderen Familienmitglieder einzufühlen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. ... war der Umgangston respektvoll und wertschätzend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. ... kam es bei Gesprächen selten zu Streit und Auseinandersetzungen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. ... wurden Konflikte offen und möglichst konstruktiv ausgetragen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. ... wurde Kritik angemessen geäußert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Danksagung

Lebenslauf