

Aus der Klinik und Poliklinik für Kinder- und Jugendpsychiatrie,
Psychosomatik und Psychotherapie der Ludwig-Maximilians-Universität

München

Direktor: Prof. Dr. med. Gerd Schulte-Körne

**Die Beurteilung von Arztgesprächen
beim Überbringen schlechter
Nachrichten
- *Eine Replikationsstudie***

Dissertation zum Erwerb des Doktorgrades der Medizin an der
Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München

vorgelegt von Clemens Frühstück aus Starnberg

2010

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatter: Prof. Dr. med. Reiner Frank

Mitberichterstatter: Prof. Dr. med. Norbert Nedopil

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 28.10.2010

AGBS – Das Manual	(5)
<u>1 Einführung – Das ärztliche Gespräch</u>	(6-15)
1.1 Überbringen einer schlechten Nachricht – Hintergrund und Forschung	(9-10)
1.2 Anforderungen für Ärzte – Entwicklungen in Aus- und Weiterbildung	(11-15)
<u>2 Instrumente zur Gesprächsbewertung</u>	(16-23)
2.1 Identifizierte Instrumente	(17-19)
2.2 Auswahl einer Übertragung des Breaking Bad News Assessment Schedule (Miller et al. 1999), der AufklärungsgesprächBewertungsSkala (Wand et al. 2008)	(20-23)
<u>3 Zielsetzung der Arbeit</u>	(24-25)
<u>4 Methoden</u>	(26-34)
4.1 Rater und Schulung	(26)
4.2 Datengrundlage – Stichprobe 1 und 2	(27-28)
4.3 Deskriptive und statistische Darstellung	(28-33)
4.3.1 Trennschärfe: Item-Skala-Korrelation	(29-30)
4.3.2 Testvalidität: Cronbach's Alpha	(30-31)
4.3.3 Interraterreliabilität: Intraklassenkorrelation	(31-32)
4.3.4 Re-Testreliabilität	(33)
<u>5 Ergebnisse – Deskription und statistische Analyse der Bewertungen von Arztgesprächen</u>	(34-50)
5.1 Validität und Interraterreliabilität der Hauptauswertung	(43-45)
5.2 Validität, Interraterreliabilität und Re-Testreliabilität des Pre-/Re-Test	(46-48)
5.3 Validitäten und Interraterreliabilitäten insgesamt	(49-50)

<u>6</u>	<u>Diskussion – Ziele und Nutzen der Untersuchung</u>	(51-71)
6.1	Vergleich der Untersuchungsbedingungen	(52-53)
6.2	Vergleich der Ergebnisse mit externen Daten	(54-62)
6.2.1	Pilotstudie – Pre-Test	(54-56)
6.2.2	Expertenrating	(56-57)
6.2.3	Hauptauswertung	(57-62)
6.2.4	Abschätzung von Trainingseffekten im Re-Test	(62-63)
6.3	Zusammenhang zwischen Intuition und Bewertung	(64)
6.4	Emotionen	(65-68)
6.5	Nonverbale Kommunikation	(68-69)
6.6	Standardisierte Raterschulung	(69-71)
6.7	Weiterentwicklung und Anwendung	(71)
<u>7</u>	<u>Zusammenfassung</u>	(72-75)
7.1	Einleitung	(72)
7.2	Zielsetzung	(72-73)
7.3	Methodik	(73)
7.4	Ergebnisse	(73-74)
7.5	Diskussion	(74-75)
<u>8</u>	<u>Literaturverzeichnis</u>	(76-78)
<u>9</u>	<u>Abbildungsverzeichnis</u>	(79)
<u>10</u>	<u>Tabellenverzeichnis</u>	(80)
<u>11</u>	<u>Anhang</u>	(81-86)
<u>12</u>	<u>Danksagung</u>	(87)
<u>13</u>	<u>Lebenslauf</u>	(88)

AGBS – Das Manual

Abbildung 1
Manual der AufklärungsGesprächsBewertungsSkala (Kurzfassung)

Themenblock	Frage	
A (Gesprächseinleitung – Gestaltung der Szene)	1	Wie gut ist es dem Arzt gelungen, eine geeignete Umgebung für das Gespräch zu schaffen?
	2	Hat der Arzt den Patienten angemessen begrüßt und sich vorgestellt?
	3	Hat sich der Arzt zu Beginn des Gespräches nach dem momentanen Befinden und den Lebensumständen des Patienten erkundigt?
B (Die schlechte Nachricht überbringen)	4	Hat der Arzt erfragt, wie viel der Patient bereits weiß, bevor er die Nachricht überbracht hat?
	5	Hat der Arzt die schlechte Nachricht vorsichtig eingeleitet („Warnschuss“), bevor er sie ausgesprochen hat, und sich anschließend an den Bedürfnissen des Patienten orientiert?
	6	Hat der Arzt beim Überbringen der Nachricht eine verständliche Sprache verwendet und dem Patienten die Entscheidung überlassen, wie detailliert über den Sachverhalt gesprochen wird?
	7	Hat der Arzt die Nachricht in einem angemessenen Tempo und an die Aufnahmefähigkeit des Patienten angepasst überbracht?
	8	Hat der Arzt nach dem Überbringen der Nachricht eine angemessene Pause gelassen?
C (Bedenken des Patienten herausfinden)	9	Hat der Arzt den Patienten eingeladen Fragen zu stellen?
	10	Hat der Arzt versucht, die Sorgen/Bedenken des Patienten zu erfassen (siehe Abbildung 16)?
	11	Hat der Arzt erfasst, welche Sorgen/Bedenken im Zusammenhang mit der Erkrankung und Behandlung (siehe Abbildung 16) für den Patienten am wichtigsten sind?
D (Informationsübermittlung)	12	Waren die vom Arzt übermittelten Informationen den vom Patienten genannten Fragen bzw. Bedenken angepasst?
	13	Hat der Arzt die Informationen strukturiert vermittelt und sich vergewissert, dass der Patient die Informationen verstanden hat?
	14	Hat der Arzt es geschafft, das Augenmerk auf positive Aspekte zu lenken?
	15	War der Inhalt des Gesprächs faktisch korrekt (siehe Abbildung 16)?
E (Allgemeine Überlegungen)	16	Wie viele der Schlüssel-Bereiche (Behandlung, Prognose, Gefühle & Emotionen, Familie & Partnerschaft, Auswirkungen auf soziale Stellung) wurden besprochen?
	17	Wurden psychosoziale Aspekte (siehe Abbildung 16), die der Patient während des Gespräches verbal signalisierte, erörtert?
	18	Hat der Arzt es geschafft, dem Patienten während des Gespräches hilfreich zur Seite zu stehen?
	19	Hat der Arzt sich während des Gesprächs einer angemessenen Körpersprache bedient?
	20	Hat es der Arzt während des Gespräches vermieden unbeholfen zu wirken?
	21	Hat der Arzt – bezogen auf das ganze Gespräch – ein angemessenes Gesprächstempo eingehalten?
	22	Hat der Arzt das Gespräch rechtzeitig und angemessen beendet?

1 Einführung – Das ärztliche Gespräch

Das Gespräch zwischen Arzt und Patient stellt eine der wichtigsten Interventionen des Arztes in der Patientenversorgung dar. Die aktuelle Ausrichtung des Gesundheitswesens bringt einen zunehmend ökonomischen Aspekt in der Behandlung mit sich. Dies legt die Notwendigkeit von Evaluation dieser ärztlichen Intervention und deren Einbeziehung in Aus- und Weiterbildung nahe.

Video- und die im digitalen Zeitalter damit verbundene Computertechnik ist besonders geeignet, um Prozesse im medizinischen Kontext generell, und die Interaktion zwischen Personen im Speziellen, festzuhalten. Verschiedene Arbeitssituationen wurden in einer Krankenhausambulanz und im Operationssaal gefilmt und die Aus- und Bewertbarkeit begutachtet (Heath et al. 2007). Videoaufzeichnungen stellten dabei eine möglichst objektive Forschungsgrundlage dar, anwendbar als Ansatz für medizinisch relevante Sozialwissenschaft sowie als wertvolle Ressource für die Lehre.

Die kommunikative Kompetenz des Arztes hat einen Einfluss darauf, ob ein Gespräch eine Hilfe oder eine Belastung für den Patienten darstellt. So wurde für Aufklärungsgespräche nachgewiesen, dass ein hauptsächlich an den Bedürfnissen des Patienten orientierter Gesprächsstil in allen untersuchten Kategorien besser bewertet wurde, als jeweils einer mit dem Schwerpunkt auf Erkrankung oder Gefühlen (Schmidt-Mast et al. 2005). Denn mit dem Fokus auf Emotionen kam häufig die eigentliche Informationsübermittlung zu kurz – und bei reiner Konzentration auf die Übermittlung der Informationen entsprechend die der Emotionen.

Ein unerwarteter Effekt ist, zusätzlich zu dem Profit der Patienten an kommunikativen Kompetenzen des Arztes, die positive Auswirkung für Ärzte selbst. Die Ergebnisse empirischer Untersuchungen zeigen, dass eine gute Beurteilung des

eigenen Gesprächsverhaltens auch positive Auswirkungen auf die Gesundheit der Ärzte hat (Ramirez et al. 1996). So hatten in England 882 befragte Ärzte jeden Alters aus vier medizinischen Disziplinen, mit unterschiedlichem Patientenkontakt eine einheitliche Prävalenz von 27 Prozent für eine psychiatrische Erkrankung. Hierbei zeigte sich eine direkt proportionale Korrelation der Gesundheit mit der allgemeinen Berufszufriedenheit. Diese setzte sich zu dreiviertel der erhobenen Ursachen aus der guten Auseinandersetzung mit Patienten, deren Angehörigen und Kollegen zusammen. Im Bereich der Gesprächsführung fühlten sich 45 Prozent der Befragten jedoch nicht adäquat aus- beziehungsweise fortgebildet. Somit kommen die Autoren zu dem Schluss, dass gute Kommunikationsfähigkeit einen protektiven Effekt auf die psychische Gesundheit von Ärzten hat.

Eine gelegentlich unterschätzte Qualität der Kommunikation ist die nonverbale Interaktion, also die Bedeutung von Gesten, Ausdruck und Mimik.

Ein Zitat des bedeutenden Didaktikers Hilbert Meyer erläutert die Wichtigkeit von Nonverbalia: „Die Körpersprache ist nicht nur eine wünschenswerte, zur Not aber entbehrliche Zutat zur Verbalsprache, sondern eine umfassende leibliche Grundlage.“ (Jank und Meyer 1991)

Zunehmend gibt es auch wissenschaftliche Untersuchungen zur Bedeutung von nichtsprachlichen Interaktionen im medizinischen Bereich. In einer Studie hierzu wurden Videoaufnahmen von Gesprächen zwischen Arzt und Patient auf verbale und vor allem nonverbale Kommunikationsmuster bezüglich Sympathie, Führbarkeit und Vorhersage untersucht. In jeder der drei inhaltlichen Kategorien korrelierte der Anteil an sprachlicher und nicht sprachlicher Interaktion des Arztes signifikant ($r=0.28$, $p<0.05$) mit den kommunikativen Eigenschaften des Patienten. Hierdurch wurde die Hypothese bestätigt, dass sich verbale und nonverbale Kommunikation in der untersuchten Situation bedingen (Thierer et al. 2001). Hieraus lässt sich, aufgrund

der bereits dargelegten Relevanz von Kommunikationskompetenz allgemein, die Bedeutung von nicht sprachlich getragener Interaktion ableiten.

Es findet sich ferner häufig die Meinung, dass der Arztberuf selbst die Fähigkeit zur Gesprächsführung mit dem Patienten verleiht. Seit den 1990er-Jahren verbreitet sich hingegen ausgehend vom angelsächsischen Raum stetig die Anerkennung der Bedeutung von Kommunikationstraining im Rahmen der ärztlichen Aus- (Schildmann et al. 2006) und Weiterbildung (Weber et al. 2003). Dies wird nicht zuletzt durch die Zunahme von wissenschaftlichen Untersuchungen (Verzehnfachung der Publikationen ab ca. 1960), sowie Lehrveranstaltungen im medizinischen Ausbildungscurriculum verdeutlicht (Frank et al. 2010; Kopecky-Wenzel et al. 2009). Wird im vorklinischen Ausbildungsabschnitt des Medizinstudiums noch ein allgemeiner und theorielastiger Ansatz zur Schulung von Fertigkeiten in der Gesprächsführung vermittelt, so sieht beispielsweise die deutsche Studienordnung für den klinischen Bereich ein eher problem- und praxisorientiertes Lehrangebot vor. Hier liegt der Fokus auf dem Überbringen schlechter Nachrichten mit Blick auf spezielle Problematiken in den einzelnen Disziplinen, wie exemplarisch der Onkologie oder auch der Kinder- und Jugendpsychiatrie.

1.1 Überbringen einer schlechten Nachricht – Hintergrund und Forschung

Bis in die 1960er-Jahre galt allgemein die Aufklärung von Patienten über schwerwiegende Diagnosen als „kontraindiziert“. Seit dieser Zeit kam es jedoch weltweit und mehrheitlich zu einer Änderung der Einstellung unter Ärzten. Lehnten damals in einer Umfrage noch neun von zehn Ärzten die Aufklärung von Patienten über eine Krebsdiagnose ab (Oken 1961), so stimmten Ende der 1970er-Jahre 90 Prozent der befragten Ärzte einer Aufklärung über lebensbedrohliche Erkrankungen zu (Novack et al. 1979).

Die Begründung, die früher gegen eine Aufklärung angeführt wurde, basierte auf der paternalistischen Haltung von Ärzten (Madder 1997), dass das Wissen der Patienten um die Krankheit und ihre Schwere die Heilung und damit die notwendigen Ressourcen für eine Therapie beeinträchtigen könnte. Weiterhin wurde argumentiert, dass es für Patienten mit einer infausten Prognose besser sei, die verbleibende Zeit unbelasteter verbringen zu können.

Die Ergebnisse mehrerer empirischer Studien widersprechen diesen Ansichten jedoch klar, indem 87 Prozent beziehungsweise 79 Prozent der 2331 beziehungsweise der 250 befragten Krebspatienten angaben, so viel Information wie möglich erhalten zu wollen. Nahezu alle Studienteilnehmer wollten außerdem wissen, ob sie Krebs haben (Jenkins et al. 2001; Meredith et al. 1996).

Im ärztlichen Alltag ist das Überbringen von schlechten Nachrichten eine häufige Aufgabe. So gab in etwa die Hälfte der 500 bei einer empirischen Untersuchung befragten Onkologen an, wenigstens zehn Mal pro Monat Patienten eine schlechte Nachricht zu überbringen (Baile et al. 2000). Die meisten der gegenwärtig in der Patientenversorgung in Deutschland tätigen Ärzte haben allerdings im Studium oder in der Weiterbildung zum Facharzt keine Vorbereitung auf diese Aufgabe erfahren.

Auch wenn sich die Abkehr von einer paternalistischen Haltung theoretisch innerhalb der letzten Jahrzehnte vollzogen hat, gibt es Belege dafür, dass eine Umsetzung dieser geänderten Vorstellungen um einiges mehr an Zeit benötigt. Eine wissenschaftliche Untersuchung konnte beispielsweise zeigen, dass die Sorgen der Patienten mit unter fünf Prozent Gesprächsanteil in der Kommunikation mit Ärzten sehr selten zur Sprache kommen (Ford et al. 1982). Außerdem wurde nachgewiesen, dass schwierige aber entscheidend wichtige Themen, wie eine Krebsdiagnose, von bis zu 60 Prozent der behandelnden Ärzte einer Disziplin überhaupt nicht mitgeteilt wurden (Buckman 1996).

Die Aufklärung des Patienten über eine lebensbedrohliche Erkrankung oder die Unterrichtung von Angehörigen über den Tod eines Patienten werden in der Literatur häufig als Beispiele für Gesprächssituationen genannt, in denen schlechte Nachrichten überbracht werden. Dies stellt aber überwiegend die ärztliche Sicht einer schwierigen Situation, beziehungsweise einer schlechten Nachricht, dar. Es sollte daher berücksichtigt werden, dass das Erleben von schwierigen Situationen und die Bewertung einer Information sowohl durch den Arzt als auch durch den Patienten erfolgt und Diskrepanzen aufweisen kann. So ist die Information über eine nicht unmittelbar lebensgefährdende Erkrankung, deren Therapie allerdings eine grundlegende Änderung des Lebensstils erfordert, aus der Perspektive von Patienten ebenfalls eine schlechte Nachricht.

1.2 Anforderungen für Ärzte – Entwicklungen in Aus- und Weiterbildung

Im Rahmen mehrerer Untersuchungen zur Befragung von Patienten, deren Eltern, der behandelnden Ärzte und Jugendhilfe-Beteiligten, befasst sich eine Promotionsarbeit (Finsterer 2008) mit 96 jugendlichen Patienten, die zwischen 1999 und 2000 zur ambulanten Beratung oder Behandlung in der Poliklinik der Klinik für Kinder- und Jugendpsychiatrie, Psychosomatik und Psychotherapie der Ludwig-Maximilians- Universität in München gewesen waren. Der Fragebogen zur Zufriedenheit mit der Behandlung (FBB) und der Verhaltens- Fragebogen nach Achenbach für Jugendliche (YSR) oder (YASR) wurden im Jahr 2002 erhoben.

Die Gesamtzufriedenheit in dieser Studie ist mit einem Mittelwert von 2,7 auf einer Skala von 0 (unzufrieden) bis 4 (zufrieden) und einer Standardabweichung von 1,0 hoch. Die Gesamtzufriedenheit ist in allen Studien hoch und vergleichbar.

Dabei ist die Zufriedenheit mit dem Behandlungserfolg jedoch relativ gering. Das Gesamtergebnis ist somit auf die sehr hohe Zufriedenheit mit der Beziehung zu den Behandelnden zurückzuführen.

Die Leistung, die also von einem Arzt in der Interaktion mit einem Patienten gefordert wird, und unter dem Gesichtspunkt der Patientenzufriedenheit in einigen Studien untersucht wurde, hat sich demzufolge im Laufe einer langen Zeit verändert und verändert sich weiter. Die Erfordernisse sind zusätzlich zu den somatisch-medizinischen Fähigkeiten um die sogenannten Soft Skills der Gesprächsführung zu einer fachlichen Kompetenz erweitert worden. Dies ist der Konsens der gesamten für diese Arbeit ausgewerteten Literatur. Die **Leistung der Ärzte** (siehe 2) gilt es demnach zu erfassen.

Zu den Anforderungen an Ärzte in einem Gespräch kommt hinzu, dass der Arzt sowohl mit den Emotionen des Patienten als auch mit seinen eigenen umgehen muss. Er sollte versuchen, ein gewisses Maß an Kontrolle zu erlangen, um gegebenenfalls den Patienten und sich selbst durch gezielte Interventionen zu entlasten, wenn Emotionen zu überwältigen drohen. Es bedarf eines Eingehens durch Ansprache oder einer Geste. Es gilt außerdem, häufige Fehler, wie zum Beispiel das Ausreden oder Übergehen von Emotionen oder psychosozialen Themen, zu vermeiden. Als Beispiele (siehe Abbildung 2) und daraus abzuleiten auch **Anforderungen für ärztliche Fähigkeiten** wurde Folgendes gezeigt:

Abbildung 2
Beispiel psychosozialer Themen

Attitudes	Knowledge
<ul style="list-style-type: none">• Openness• Flexibility• Non-judgement• Altruism• Acceptance	<ul style="list-style-type: none">• Function of each emotional state• Links between emotion, cognition and behaviour• Physiology of emotional states• Knowledge of strategies to identify and regulate emotion• Understanding cultural beliefs that influence emotion
Emotion skills	
Other-directed	Self-directed
<ul style="list-style-type: none">• Making empathic statements• Reading facial expressions• Soothing an uneasy patient• Communicating emotions to others• De-escalating a tense situation• Instilling hope• Eliciting patient emotions• Shared emotional processing• Supporting a colleague	<ul style="list-style-type: none">• Positive self-talk• Interoceptive awareness• Somatic quieting strategies• Awareness of own emotions• Impulse control• Cognitive reframing to self• Emotional suppression• Self-reflection on emotions• Self-forgiveness

(Satterfield und Hughes 2007)

Mehrere Untersuchungen ergaben weiterhin, dass psychosoziale Themen als wichtiger Grund für einen Arztbesuch zu sehen sind. **30-50 Prozent aller Arztbesuche haben psychosoziale Hintergründe.** Ärzte tendieren dazu, diese zu übersehen, wenn sie nicht direkt vom Patienten angesprochen werden. Sogar wenn sie direkte Fragen zu psychosozialen Themen stellen, passiert dies in ungefähr 25 Prozent der Gespräche (Satterfield und Hughes 2007). Ärzte greifen Emotionen und psychosoziale Inhalte im Gespräch nicht immer adäquat auf. Dies streicht noch einmal die Forderung nach einer Verbesserung heraus.

Der **Nutzen von Lehr- und Fortbildungsveranstaltungen** zum Überbringen schlechter Nachrichten wird häufig in der Annahme infrage gestellt, dass die Fähigkeit schwierige Gesprächssituationen zu bewältigen, nicht erlernt werden könne. Dieser Meinung stehen empirische Untersuchungen entgegen, die einen positiven Effekt erfahrungsorientierter Lehr- und Fortbildungsveranstaltungen auf das Gesprächsverhalten belegen.

Im Rahmen einer prospektiven Studie aus Großbritannien mit 160 Onkologen auf Facharzniveau wurde festgestellt, dass Berufsausübung und der damit verbundene Erfahrungsgewinn allein sich nicht so sehr auf die Kommunikationsfähigkeiten auswirkt. In einer Bewertung, die sich an dem „Medical Interaction Process System“ (siehe 2.1) orientiert, erzielten 80 Ärzte nach dem Besuch einer dreitägigen Schulung in vier von sieben Kategorien um durchschnittlich 25 Prozent verbesserte Ergebnisse im Vergleich zur Kontrollgruppe (Fallowfield et al. 2002).

Eine ähnliche Arbeit aus den USA evaluierte die Fähigkeiten von 106 Assistenzärzten aus der Onkologie vor und nach einem viertägigen Workshop mittels der Elemente des „Six-Step Protocol for Delivering Bad News“ (siehe 2.1). Hierbei

zeigten nach dem Training 31 Prozent mehr Ärzte ($p < 0.001$) Fähigkeiten im Überbringen schlechter Nachrichten (Back et al. 2007).

Die medizinische Fakultät der Ludwig-Maximilians-Universität München (LMU) und **Harvard Medical International** (HMI) gründeten 1996 eine Allianz zur Förderung der medizinischen Ausbildung (**Alliance for Medical Education**). Unter diesem Einfluss entwickelte sich durch die erneuerte Approbationsordnung für Ärzte (BundesministeriumFürGesundheit 2002) die Reform des Medizinstudiums an der medizinischen Fakultät München, des Weiteren resultierte hieraus 2003 die Einführung von „MeCuM“, des neuen Medizinischen Curriculums München (Medizinische Fakultät der LMU München 2009). In diesem Rahmen entstand bereits eine Vielzahl neuer, moderner Lehrveranstaltungen, sowohl für Studenten, als auch für Dozenten.

In dem neunzigminütigen **Seminar „Überbringen schlechter Nachrichten“** wird Medizinstudenten die Möglichkeit gegeben, in einem Rollenspiel mit Kommilitonen eine schwierige Gesprächssituation zwischen Arzt und Patient zu erproben (Kopecky-Wenzel et al. 2007; Kopecky-Wenzel et al. 2009). Es handelt sich hierbei um Situationen zwischen einem Arzt und zwei Gesprächspartnern, also drei Personen insgesamt.

Die „Alliance for Medical Education“ führt seit 1996 im Seminarzentrum Frauenchiemsee zweimal jährlich in Form von Trainingskursen für Dozenten die reguläre, inzwischen standardisierte Ausbildung zum Tutor durch. Der Lehrkörper der Fakultät bekommt somit Gelegenheit zu einer konzentrierten und zielorientierten **Weiterbildung für Hochschullehrer**. In den fünftägigen Intensivkursen erhalten die

Dozenten unter Beteiligung der Harvard Medical School eine professionelle Ausbildung für ihre Lehrtätigkeit an der medizinischen Fakultät der LMU München (Muntau und Maier 2007).

Im Rahmen dieser Veranstaltung besuchen die Teilnehmer ein videodokumentiertes, rollenspielbasiertes Kommunikationsseminar, in dem – wie auch im studentischen Seminar – eine Situation mit drei Gesprächspartnern geprobt wird.

Insbesondere Ausbildungs- und Lehrsituationen für Gesprächsführung legen die **Notwendigkeit zuverlässiger Bewertungswerkzeuge für verschiedene Gesprächssituationen** nahe. Überdies erscheint nicht nur die Möglichkeit zur möglichst zeitnahen Bewertung, optimaler Weise parallel in Echtzeit, sondern auch die Anwendbarkeit zur Reflektion und Gruppendiskussion im Ausbildungsumfeld als erstrebenswert. Weiterhin kann in Zukunft auch die Handhabung solcher Instrumente zur Kommunikationsbeurteilung im Rahmen einer evidenzbasierten Medizin und Qualitätssicherung eine Rolle spielen.

Es wurden bereits einige Bestrebungen angestrengt, Instrumente zur Gesprächsbewertung zu entwickeln. Allerdings fehlt es zumeist an Validierung, im Sinne einer Überprüfung der Bewertungsergebnisse durch Replikation, beziehungsweise auf Anwendbarkeit in unterschiedlichen Situationen.

In der dieser Arbeit hauptsächlich zugrundeliegenden Beobachtungsstudie (Wand et al. 2008) wurde das Gesprächsverhalten mithilfe einer Beurteilungsskala bewertet.

Nach persönlicher Auskunft der Autoren im Juni 2008 zeigte sich hier vergleichsweise zur Kontrollgruppe eine Verbesserung nach Besuch der Lehrveranstaltung.

2 Instrumente zur Gesprächsbewertung

Der Kommunikationsprozess der Arzt-Patient-Interaktion ist ein hoch komplexer Vorgang. So existiert bislang keine Methodik der Analyse, die alle Dimensionen der Kommunikation umfasst.

Anhand eingehender, systematischer Literaturrecherche unter Zuhilfenahme der Suchportale PUBMED und MEDLINE wurden wissenschaftliche Beiträge im Kontext ärztlicher Kommunikationskompetenzen untersucht. Unter Verwendung der Schlüsselwörter „Doctor-patient communication“, „physician-patient relation“, „communication skills training“ und „breaking bad news“ ließen sich 1144 wissenschaftliche Beiträge ausfindig machen (Stand Mai 2008).

Im Folgenden soll nun die Anwendbarkeit eines als günstig festzustellenden Bewertungssystems auf ausgewählte Videoaufzeichnungen untersucht werden. Hierbei ist besonders wichtig, dass gemäß dem Konzept, im Medizinstudium die Fähigkeiten der Studenten zu schulen und zu überprüfen, der Schwerpunkt der Beurteilung auf der Leistung der Rolle des Arztes liegt. Diese wird durch den Anteil an gelungener Kommunikation durch eigene Gesprächsführungskompetenz dargestellt.

Die Intention von Bewertungsinstrumenten im Allgemeinen und in der Anwendung in der vorliegenden Arbeit im Speziellen, liegt in der praktikablen und präzisen Erfassung von Kommunikationsleistung. Daher wurden im Rahmen dieser Arbeit die dadurch positiv bedingten Stärken, Exaktheit, Vergleichbarkeit und Verlässlichkeit, für die Auswahl des Instrumentes bewertet. Dies lässt sich in der statistischen Interpretation der Ergebnisse ersehen.

2.1 Identifizierte Instrumente

Als das bekannteste und meistgenannte Ratingsystem (engl.: to rate = bewerten) für Kommunikation im ärztlichen Umfeld – und somit gewissermaßen als Referenz – darf das **„Roter Interaction Analysis System“**, kurz „RIAS“ (Roter und Larson 2002) gelten. Es stellt eine sehr ausführliche Möglichkeit dar, eine Gesprächssituation zu untersuchen und auszuwerten; allerdings bleibt es dem Untersucher überlassen, eine Bewertung zu erstellen. Es existiert daher eine Vielzahl von Adaptationen für verschiedene Anwendungen und mehr als hundert Publikationen mit einem Bezug zu diesem System. Auch deshalb ist für dieses etablierte Instrument bislang kaum eine exakt vergleichende Reliabilitätstestung im Sinne einer Replikation erfolgt.

Für die Analyse von Arzt-Patient-Interaktionen während einer Konsultation wurde das **„Medical Interaction Process System“**, kurz „MIPS“ (Ford et al. 2000) entwickelt, welches erlaubt, formale und inhaltliche Elemente in der Arzt-Patient-Interaktion darzustellen. Die Struktur der Skala zielt in erster Linie auf die Informationsübermittlung ab, hierbei werden Verhaltensweisen wiederum quantifiziert aber nicht explizit bewertet. Auch hier fehlt es bislang an genau wiederholenden, vergleichenden Untersuchungen, die primär auf die Validierung des Instrumentes abzielen. Es lassen sich einige Anwendungen dieses Instrumentes identifizieren (siehe 1.2), wie zum Beispiel eine Studie, bei der 48 Onkologen an einer dreitägigen Assistentenschulung teilnahmen und danach vergleichsweise ein signifikant besseres Ergebnis zeigten (Jenkins und Fallowfield 2002). So verbesserte sich beispielsweise die Wahrnehmung psychosozialer Themen im Gegensatz zur Kontrollgruppe ($p=0.002$) ähnlich wie in den anderen untersuchten Bereichen auch. Die Interrater-Reliabilität wurde mittels ICC mit 0,69 bestimmt (siehe 4.3.2). In den Veröffentlichungen, die sich mit diesem Instrument beschäftigen, findet sich im

Unterschied zu den meisten übrigen Publikationen auch die Überprüfung der Stabilität der Ratingergebnisse über jeweils drei Monate. Diese ist mit einem Korrelationskoeffizienten (siehe 4.3.3.) von 0.7, beziehungsweise 0.77 auf einem guten Niveau.

Das „SPIKES“-Protokoll, „**A Six-Step Protocol for Delivering Bad News**“ (Baile et al. 2000), ist bislang nur in englischer Sprache verfügbar. Es bietet einen gut strukturierten Leitfaden für die Gesprächsführung zum Überbringen einer schlechten Nachricht zwischen Arzt und Patient. Eine Arbeit, welche sich mit diesem Instrument beschäftigt, wurde bereits unter 1.2 beschrieben (Back et al. 2007). Außerdem wurde im Expertengespräch eine Verwendung des Protokolls in der Lehre identifiziert und erörtert. Es mangelt aber, wie in einigen anderen Systemen auch, an der unkomplizierten, exakt vergleichbaren Aus- und Bewertbarkeit. Daher fehlt für dieses Instrument ebenfalls die Anwendbarkeit für eine Replikationsstudie.

Zur Bewertung von Anamnesegesprächen in der Lehre wurde die „**Maastricht history-taking and advice scoring list of global items**“, kurz „MAAS-Global“ (Thiel et al. 2000) entwickelt. Für dieses Instrument ist eine hohe Testgüte beschrieben (Hobma et al. 2006; Thiel et al. 2000), teilweise jedoch, ohne dass genauere Daten hierzu einsehbar gewesen wären. In einer weiteren Anwendung dieses Instrumentes (Langewitz et al. 1998) ergaben sich signifikante Verbesserungen des Gesprächsverhaltens im Vergleich zur Kontrollgruppe. So zeigten sich bei Untersuchungen von weiterzubildenden Internisten in der Schweiz vor und nach einer Schulung beispielsweise eine durchschnittlich um 41,4 Prozent verbesserte Wahrnehmung der Sorgen und Bedenken der Patienten während Anamnesegesprächen. Ähnliche Ergebnisse wurden für die übrigen Themenbereiche

gezeigt. Dies ist im Vergleich zu anderen Untersuchungen, die einen positiven Effekt von Gesprächstraining zeigen, gleichzeitig ein Nachweis der Güte dieses Instrumentes, ohne dass die Validierung spezifisches Ziel der Arbeit gewesen wäre. Von den Autoren (Thiel et al. 2000) wurde bereits zu bedenken gegeben, dass sich die Skala allerdings nicht ohne Weiteres auf andere Gesprächssituationen als die Anamneseerhebung übertragen lässt und nicht ausschließlich die Leistung des Arztes bewertet werden kann. Aus diesen Gründen kam eine Anwendung von „MAAS-Global“ in der vorliegenden Studie nicht in Betracht.

Der „**Calgary-Cambridge Guide**“ (Kurtz et al. 2003), insbesondere einzelne Items (Fragen) sind für den geplanten Einsatz in dieser Arbeit gut geeignet. Die thematische Gruppierung der Items hält allerdings der Validierung bislang nur unzureichend stand (Simmenroth-Nayda et al. 2008). Eine persönliche Nachfrage Ende des Jahres 2009 ergab die Ankündigung einer detaillierten Veröffentlichung der Ergebnisse.

Zusammenfassend bleibt noch einmal herauszustellen, dass es für nahezu alle Gesprächsbewertungswerkzeuge an Überprüfung der Anwendbarkeit in unterschiedlichen Gesprächssituationen mangelt.

Es konnten Untersuchungen identifiziert werden (Cegala und Lenzmeier Broz 2002; Hulsman et al. 1999), welche sich um eine Aufstellung der bisherigen Forschungen bemühen. Versuche einer konkreten Anwendung bleiben, im Gegensatz zur vorliegenden Studie, jedoch aus.

2.2 Auswahl einer Übertragung des Breaking Bad News Assessment Schedule (Miller et al. 1999), der AufklärungsgesprächBewertungsSkala (Wand et al. 2008)

Nach der Identifikation zahlreicher Gesprächsbewertungssysteme (siehe 2.1) fiel die Wahl für die vorliegende Studie auf eine Übersetzung beziehungsweise Adaption des „Breaking Bad News Assessment Schedule“, kurz **BAS** (Miller et al. 1999). Das BAS ist ein Fragebogen mit 23 Items in fünf Themenbereichen, welche mit jeweils ein bis fünf Punkten zu bewerten sind. So ergibt sich in der Summe ein Ergebnis zwischen 23 und 115 Punkten. Es ist ein sehr kurz gehaltenes Manual, um speziell Gesprächssituationen zu bewerten, in denen eine Krebsdiagnose übermittelt wird.

Die **Anforderungen an das Bewertungssystem** sind ein vertretbarer Zeitaufwand, ein Maß zwischen Ausführlichkeit und Vollständigkeit einerseits und der Verwendbarkeit während einer Lehrinheit sowie zur Vor- und Nachbereitung derselben andererseits.

Demzufolge wurde die „Aufklärungsgesprächbewertungsskala“, kurz **AGBS** (Wand et al. 2008) ausgewählt.

Hierbei handelt es sich um ein Auswertungs- und Bewertungsinstrument von Gesprächssituationen im ärztlichen Alltag. Der Fokus bei der Beurteilung von Gesprächen liegt auf der Leistung der Ärzte (siehe 1.2 und 2), da dies im medizinischen Bereich einen Kernpunkt des Interesses darstellt (Buckman 2002; Egger 2005; Fallowfield 1993; Hobma et al. 2006).

Da die **Auswertungszeit** eines Gesprächs mit 15-20 Minuten um circa ein Drittel bis ein Viertel kürzer als bei vergleichbaren Systeme veranschlagt wurde, sollte auch die Voraussetzung des geringen Zeitaufwandes erfüllt sein.

Die einzelnen Fragen der AGBS werden in Form einer **Checkliste mit 22 Items** in fünf Themenblöcken (siehe Abbildung 1), ähnlich einer Likert-Skala, in fünf Notenschritten von „sehr gut“, „gut“ und „mittelmäßig“ bis „schlecht“ sowie „sehr schlecht“ bewertet. So ergibt sich in der Gesamtbewertung eines Gespräches ebenfalls eine **Note** zwischen eins und fünf **im Gegensatz zum Punktwert** der BAS.

Durch die Auswertung der Leistung des Arztes während des Gespräches in Form von Schulnoten bietet sich die Möglichkeit, eine einfache, **vergleichbare Bewertung mehrerer Ärzte** zu treffen.

Die **Raterschulung** ist ein allgemein etabliertes Vorgehen, welches in allen Veröffentlichungsarbeiten von Bewertungssystemen erwähnt wird. Es mangelt hierbei aber zumeist an einer strukturierten und standardisierten Anweisung seitens der Autoren. Dies erscheint jedoch für die Bedeutung der Reliabilität als Gütekriterium sinnvoll.

Dies wurde nicht zuletzt auch bei dem Zusammentreffen mit der Erstautorin der Originalarbeit, Frau Wand, im Rahmen des Workshops „Methoden der Schulung in Gesprächsführung“ am 21.07.2008, intensiv diskutiert und bot zusätzliche Perspektiven zur Weiterentwicklung.

Die Autoren des **BAS** (Miller et al. 1999) gaben an, für die Schulung der **drei Beurteilenden**, eines Krankenpflegers aus der Onkologie, eines Onkologen und eines Psychiaters, etwa **drei Stunden** verwendet zu haben.

Die Arbeitsgruppe der Veröffentlichungsarbeit der **AGBS** (Wand et al. 2008) veranstaltete zu diesem Zweck eine kompakte Schulung. Diese wurde für die **drei Rater**, eine Assistenzärztin und zwei Medizinstudenten, an zwei Tagen und in insgesamt etwa **sieben Stunden** durch die Autoren abgehalten.

Dieser geringe Aufwand an Raterschulung kommt dem gesamten Zeitaufwand und damit dem Anwendungsspektrum zu Gute.

Gemäß dem zugrundeliegenden **BAS** (Miller et al. 1999), basiert die **AGBS** auf der Zielsetzung, ein Instrument zur Auswertung speziell von Aufklärungsgesprächen über bösartige Tumorerkrankungen zu entwickeln. Dementsprechend bildeten auch Videoaufnahmen von Rollenspielen aus einem studentischen Seminar „Wahrheit am Krankenbett – Das Aufklärungsgespräch mit schwerkranken Patienten“ die Datengrundlage beziehungsweise die untersuchte **Stichprobe**. In diesem Seminar überbringen Medizinstudenten, standardisiert geschulten Laiendarstellern als Patienten die schlechte Nachricht über vier ausgewählte Erkrankungen (siehe Abbildung 16). Die Autoren um Frau Wand verwendeten **54 Videoaufnahmen** aus dem studentischen Seminar mit **standardisierten Patienten** und eng beschränkter Erkrankungsvielfalt aus der **Onkologie**. Weiterhin wurde teilweise selektiert und bestimmte Inhalte, wie beispielsweise die Brustkrebserkrankung, herausgenommen.

Die veröffentlichten Ergebnisse in Bezug auf die **Testgüte** (Wand et al. 2008), welche die AGBS als praktikables Instrument zur Beurteilung wichtiger Gesprächsfertigkeiten ausweisen, ließen schließlich endgültig die Entscheidung für dieses System im Rahmen dieser Arbeit fallen.

Die AGBS kann also in Lehrveranstaltungen zum **Überbringen schlechter Nachrichten** eingesetzt werden.

Die Kommunikation zwischen verschiedenen Gesprächspartnern bedingt einen Informationsaustausch über vielerlei „Kommunikationskanäle“, welche sowohl biologische als auch soziokulturelle Wurzeln haben können. Ein sehr spezialisierter und durch die AGBS gut erfassbarer „Kanal“ ist die Sprache. Weniger präzise stellt sich die Situation für die nonverbale Kommunikation dar.

Überlegungen, die **Medizin im Rahmen des bio-psycho-sozialen Modells** zu verstehen (Egger 2005; Engel 1977), begründen die Notwendigkeit, der emotionalen Verfassung des Patienten prinzipiell große Bedeutung beizumessen. Diese Art der Betrachtung hilft nicht nur dabei, die Entstehung der Krankheit zu ergründen, sondern auch deren Folgen auf die Psyche des Patienten zu beurteilen. Dies zeigt sich im Äußern von Emotionen und in psychosozialen Themen (siehe 1.1 und 1.2), welche von den Autoren der Originalarbeit (Miller et al. 1999) sehr knapp und scharf umrissen formuliert sind. Diese Punkte werden ebenso durch die AGBS abgefragt (siehe Abbildung 1).

Laut den Autoren weist das Instrument im Bereich der **Erfassung von Emotionen Schwächen** auf (Wand et al. 2008).

3 Zielsetzung der Arbeit

Die vorliegende Arbeit überprüft die **Gesprächsleistung von Ärzten** mittels der „Aufklärungsgesprächsbewertungsskala“, AGBS (Wand et al. 2008) als Übertragung des „Breaking Bad News Assessment Schedule“, BAS (Miller et al. 1999), in Form einer **Replikationsstudie**.

Daher ergibt sich die Fragestellung nach der **Vergleichbarkeit** zu den zwei Veröffentlichungsarbeiten (Miller et al. 1999; Wand et al. 2008). Dies wird durch die Aufstellung beziehungsweise Gegenüberstellung der Studienbedingungen in den drei Untersuchungsreihen erreicht.

Eine allgemeine Standortbestimmung erfolgte im Rahmen einer **Pilotstudie**. Im Folgenden wird einerseits das **Gefüge der Stichproben** aus den Videos analysiert, sowohl im Hinblick auf die Probanden, beziehungsweise **Ärzte**, und deren Ausbildungsstand, als auch mit Blick auf die unterschiedlichen **Gesprächspartner** und deren Anzahl. Andererseits wird die **Zusammensetzung der Rater** berücksichtigt.

Im Rahmen der Schulung und im Zeitraum der Bewertung der Videos wurde im intensiven Gespräch der Rater untereinander auch der Zusammenhang zwischen dem intuitiven Eindruck eines Videos und der danach ermittelten Note diskutiert. Zusätzlich zum primären **Ziel**, die **Gütekriterien** Reliabilität (dt.: Zuverlässigkeit) und Validität (dt.: Gültigkeit) **zu überprüfen**, wurde daher das Studiendesign erweitert. So ließ sich mit dem durchgeführten **Re-Test**-Verfahren, neben der Bedeutung für die Testgüte, auch die Frage nach der **Stabilität der Raterleistung** und dem notwendigen **Umfang der Raterschulung** beantworten.

Zudem wird in der vorliegenden Arbeit die **Erfassung** von und Bewertung des Umganges mit **Emotionen** (siehe 2.2) überprüft.

Die so gewonnenen Erkenntnisse sollen Aufschluss über die **Anwendungs- sowie Fortentwicklungsmöglichkeiten** des untersuchten Instrumentes ergeben.

4 Methoden

Im Rahmen der Anfertigung der vorliegenden Studie erfolgten über einen Zeitraum von sechs Monaten wöchentliche Treffen mit Mitgliedern des Teams der Kinder- und Jugendpsychiatrischen Ambulanz der Ludwig-Maximilian-Universität München unter Leitung von Professor Frank.

Zu Beginn lag der Schwerpunkt dieser Treffen auf der Entwicklung des Themas dieser Arbeit (siehe 3) und im Verlauf auf dem Training der Video-Beobachtung allgemein, so dass sich das Ratertraining anschloss.

4.1 Rater und Schulung

Die Raterschulung nahm einen Zeitraum von insgesamt etwa **zehn Stunden** ein und wurde durch die Rater selbst unter der Leitung von Professor Frank durchgeführt. Dabei wurden, nach der Einarbeitung in das Manual der AGBS (siehe Abbildung 1), Schwierigkeiten bei der Beurteilung analysiert und ein einheitliches Vorgehen für die Bewertung festgelegt.

Die Beurteilung des Gesprächsverhaltens von Medizinstudenten und Ärzten unter Verwendung der AGBS erfolgte durch **zwei Rater**, eine Diplomsozialpädagogin und einen Medizinstudenten, dem Autor. Die Pädagogin befand sich im Rahmen der Psychotherapie-Ausbildung im „Praktikum für Kinder- und Jugendpsychiatrie“ für die Zulassung zur Psychotherapeutin für Kinder.

4.2 Datengrundlage – Stichprobe 1 und 2

In den den Aufzeichnungen für die Replikationsstudie zugrundeliegenden Rollenspielen werden auch die **Patienten** von **Seminarteilnehmern ohne intensive Schulung** dargestellt.

Das Spektrum umfasste nicht ausschließlich **Medizinstudenten**. Es wurden auch **Ärzte mit Berufserfahrung** miteinbezogen. Außerdem werden in den Rollenspielsituationen mehrere unterschiedliche Krankheitssituationen berücksichtigt, zum Beispiel auch Notfälle. Weiterhin handelt es sich um Situationen mit **mehreren Kommunikationspartnern** des Arztes.

Insgesamt **36 Aufzeichnungen von Medizinstudenten** entstanden im Rahmen der Lehrveranstaltungen im **Seminar „Überbringen schlechter Nachrichten“** der medizinischen Fakultät der Ludwig-Maximilians-Universität München (siehe 1.2). Davon stammen 14 Videos aus dem thematischen Kontext onkologischer Erkrankungen und 12 Aufnahmen beinhalten einen notfallmedizinischen Fall.

Außerdem stammen **14 Videoaufnahmen mit Ärzten** beziehungsweise Dozenten aus Schulungen im Rahmen eines **Trainingsseminars für Hochschullehrer** (siehe 1.2.).

Die Datengrundlage bilden somit 40 Aufnahmen, 26 Studentenvideos und 14 Ärzte-/Dozentenvideos standardisierter Gesprächssituationen im Rollenspiel, im Folgenden als „**Stichprobe 1**“ benannt.

Zusätzlich wurden weitere 10 Aufnahmen, ausschließlich Studentenvideos, zu einem Pre-Test und für einen Re-Test ausgewertet und im Folgenden als „**Stichprobe 2**“ benannt.

Die Teilnehmer aller Videoaufzeichnungen haben einer wissenschaftlichen Auswertung durch Dritte zugestimmt.

Die Auswahl beziehungsweise Zuteilung der Aufzeichnungen als Datengrundlage erfolgte durch eine wissenschaftliche Mitarbeiterin, Frau Dr. Maru Kopecky-Wenzel und wurde ihrerseits aufgrund formaler Kriterien (Dauer, Aufnahmequalität) zur besseren Bewertbarkeit getroffen.

4.3 Deskriptive und statistische Darstellung für Bewertungen von Arztgesprächen

Zur Darstellung der Ergebnisse und Bestimmung der Testgüte der AGBS wurden die statistischen Berechnungen vergleichend zur Erstveröffentlichung (Wand et al. 2008) beziehungsweise der des zugrundeliegenden „BAS“ (Miller et al. 1999) durchgeführt. Hierzu wurde das „SPSS“ (Statistical Package for the Social Sciences) in der Version 16.0.2 für Microsoft Windows verwendet.

Die Bewertung der Fragen erfolgte anhand einer Likert-Skala in fünf Notenschritten von eins bis fünf (siehe Abbildung 1.).

Bei einer Messung muss das **Datenniveau** überprüft werden und damit auch, welchen der bei den reellen Zahlen gültigen Beziehungen eine empirische Bedeutung bei den Merkmalen der Objekte zukommt.

Hauptsächlich werden drei **Skalentypen** mit verschiedener Wertigkeit differenziert. Im Gegensatz zur metrischen Skala, mit dem Spezialfall der Intervallskala und der Nominalskala, entspricht die erwähnte Benotung einer **Ordinalskala**.

Hierbei werden Daten dem Rang oder der Wertigkeit nach geordnet. So entsteht eine Rangordnung, die aussagt, dass etwas zutreffender oder weniger zutreffend als etwas anderes ist. Wichtig ist, dass mit der Zuweisung von Noten keine Aussage

über den Abstand der Zensuren untereinander getroffen wird. So kann nicht behauptet werden, dass die Schulnote „zwei“ doppelt so gut ist wie „vier“. Die Ordinalskala bildet eine Relation von Messwerten ab, das bedeutet sie können in einer Rangreihe von „klein“ nach „groß“, oder umgekehrt, geordnet werden. Man nennt daher die Skalenwerte einer Ordinalskala auch Ränge und die Skala auch Rangskala.

In der vorliegenden Studie – wie auch in allen anderen Veröffentlichungen, welche sich dieser Verfahren bedienen – wird aber als **Bedingung der Darstellbarkeit und Berechnung** eine metrische, nämlich eine **Intervallskala** angenommen. Hierdurch wird eine deskriptive Darstellung unter Verwendung von Mittelwert und Streuung ermöglicht. Daraus ergeben sich verschiedene Varianzen.

Die **Varianz** einer oder mehrerer Zielvariablen wird durch den Einfluss einer oder mehrerer Faktoren erklärt. Die **Varianzanalyse** kann auch zur Ermittlung und Differenzierung von systematischen oder zufälligen Fehlern dienen. Unter den datenanalytischen Verfahren verleiht die große Gruppe der Varianzanalyse sinngemäß Aufschlüsse über die hinter den Daten steckenden Gesetzmäßigkeiten.

4.3.1 Trennschärfe: Item-Skala-Korrelation

Die Trennschärfe ist die Korrelation eines einzelnen oder, wie in dem vorliegenden Fall der Themenbereiche, mehrerer Items mit der Gesamtskala.

Daraus ist zu entnehmen, in wieweit das Gesamtergebnis mit der Beantwortung eines Themenblockes korreliert. Zur Bestimmung der Trennschärfe, die einen Zusammenhang der Bewertung jedes einzelnen Raters mit der mittleren Bewertung durch die übrigen Rater wiedergibt, wurde die korrigierte Item-Skala-Korrelation

berechnet. Hierbei steht der Wert 0 für einen fehlenden Zusammenhang der Bewertung des einen Raters mit der Bewertung durch den anderen Rater, und der Wert 1 für ein Maximum der entsprechenden Korrelation. Eine hohe Trennschärfe bedeutet also, dass der Themenbereich zwischen den Probanden im Sinne des Gesamttests zu differenzieren vermag.

Die Formel für die Trennschärfe „ r_{xy} “, die Korrelation zwischen den Werten je Item „ x “ und dem korrigierten Gesamtwert „ y “, lautet unter Verwendung der Kovarianz:

$$r_{xy} = \frac{\text{COV}(x, y)}{s_x \cdot s_y}$$

Die Kovarianz zweier Variablen ist dabei analog zur Standardabweichung die Größe der Abweichung vom jeweiligen Mittelwert untereinander. Sie wird durch den Durchschnitt der Produktterme berechnet, von den Subtraktionen der Einzelwerte mit den jeweiligen Mittelwerten:

$$\text{COV}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

4.3.2 Testvalidität: Cronbach's Alpha

Die für die Validität (dt.: Gültigkeit) eines Testverfahrens wichtige interne Konsistenz gibt an, ob alle Items (Fragen) sinngemäß dasselbe messen. Diese interne Konsistenz wurde mittels Cronbach's Alpha dargestellt. Diese Maßzahl aus der multivariaten Statistik stellt fest, inwieweit eine Gruppe von Test-Items als Messung einer einzelnen latenten Variablen angesehen werden kann und also die Ergebnisse jeder einzelnen Frage miteinander zusammenhängen. Cronbach's Alpha ist das Ergebnis einer Varianzanalyse der Itemdaten hinsichtlich der Varianz zwischen den

Testpersonen und der Varianz zwischen allen Items und kann Werte zwischen 0 und 1 annehmen. Befriedigende Werte für psychometrische Instrumente sind circa 0,7 oder mehr. Cronbach's Alpha sollte verwendet werden, wenn Items substantiell unterschiedliche Bereiche innerhalb eines einzelnen Konstrukts messen.

Es ist anfällig für falsch positive Werte, wenn die Items des Konstrukts so formuliert werden, dass sie sich nur oberflächlich unterscheiden. Denn je breiter und allgemeiner ein Instrument misst, umso mehr Chancen bestehen in der Regel auch breite und entfernte Kriterien vorherzusagen.

Die Formel zur Berechnung eines standardisierten Cronbach's Alpha lautet:

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$$

Wobei N der Anzahl der Items entspricht und \bar{r} der durchschnittlichen Korrelation (siehe 4.3.1).

4.3.3 Interraterreliabilität: Intraklassenkorrelation

Die Reliabilität (dt.: Zuverlässigkeit) allgemein ist ein Maß für die Güte der Methode, die zur Messung einer bestimmten Variablen eingesetzt wird. Sie stellt den Anteil an Varianz im Gegensatz zu Messfehlern dar.

Die Interrater-Reliabilität bezeichnet in der empirischen (Sozial-)Forschung das Ausmaß der Übereinstimmungen der Ergebnisse einer Betrachtung, beziehungsweise Bewertung bei unterschiedlichen Ratern. Dadurch kann angegeben werden, inwieweit diese Ergebnisse vom Beobachter unabhängig sind. Die Analyse erfolgte hier anhand des **Intraklassen-Koeffizienten**, der einen beliebigen Zusammenhang von Messungen beschreibt, ohne dass Abweichungen zwischen den Urteilern eine Rolle spielen. So spielen beispielsweise gleichbleibende gute oder

schlechte Bewertungstendenzen keine Rolle. In der vorliegenden Studie wurde die Interrater-Reliabilität mithilfe des Modells der zweifaktoriellen, unjustierten Intraklassen-Korrelation beziehungsweise des Intraklassen-Koeffizienten (**ICC**) berechnet.

Als **Voraussetzung für die Berechnung** (siehe 4.3) gilt neben der Normalverteilung der Daten, dass keine Interaktion zwischen beurteilten Fällen und Ratern vorliegt. Diese gleicht einer Inhomogenität der Ratermittelwerte oder der Zufallsfehler und ist somit ebenfalls ein Teil der Varianz und kann mittels **Tukey's Additivitätstest** bestimmt werden. Bei nicht erfüllten Vorbedingungen wird die Reliabilität gemeinhin unterschätzt und es resultieren schlechte Werte für den ICC.

Unterschiede in den Ergebnissen der bewerteten Gespräche resultieren zum einen aus der Performance der die Ärzte Spielenden, was sich äquivalent im ICC darstellt. Weiterhin gibt es naturgemäß Unterschiede zwischen den verschiedenen Rollenspielsituationen (siehe 4.1). Zum anderen treten aber auch Varianzen in den Bewertungen ein und derselben Gesprächssituation durch die unterschiedlichen Rater auf.

Der ICC wird aus dem Quotienten der Differenz der Varianzen „zwischen“ und „innerhalb“, sowie der Summe der Varianzen „zwischen“ und „innerhalb“ der untersuchten Fälle ohne Erwartungsverzerrung gebildet (Bias=0 → MS=Var):

$$ICC = \frac{MS_{\text{zwischen}} - MS_{\text{innerhalb}}}{MS_{\text{zwischen}} + MS_{\text{innerhalb}}}$$

4.3.4 Re-Testreliabilität

Um die Zuverlässigkeit der Messmethode zum einen, sowie Trainingseffekte zum anderen beurteilen zu können, wurde eine andere Form der Reliabilität, die Re-Testreliabilität bestimmt. Dies erweitert die Gütekriterien auch im Sinne der Validität um eine externe Konsistenz und wird für Ordinaldaten durch einen Rangkorrelationskoeffizienten berechnet. Dies ist ein nicht-parametrisches Maß für Zusammenhänge. Parameterfreie statistische Methoden sind mathematische Anwendungen, um Hypothesen zu testen. Sie beschreiben, wie gut eine beliebige, monotone Funktion die Bedingung zweier Variablen beschreiben kann, ohne Annahmen über die Verteilung der Variablen zu machen. Sie sind deswegen im Gegensatz zu parametrischen Tests auch anwendbar, wenn die bei vielen statistischen Aussagen notwendigen Verteilungsvoraussetzungen nicht erfüllt sind.

Es gibt mit Spearmans Rangkorrelationskoeffizient und Kendalls Tau zwei Rangkorrelationskoeffizienten als parameterfreie Methoden. Während ersterer von einer Äquidistanz der Skalenränge ausgeht, basiert Tau rein auf ordinaler Information.

Zur Ermittlung der **Übereinstimmungen (Reliabilität) auf Intervallskalenniveau** (siehe 4.3) wird dagegen auf die Varianzanalyse zurückgegriffen. Der besseren Vergleichbarkeit (Miller et al. 1999; Wand et al. 2008) halber, wurde für die Re-Testreliabilität ebenfalls der **ICC** bestimmt.

5 Ergebnisse – Deskription und statistische Analyse der Bewertungen von Arztgesprächen

Die **Bewertung der einzelnen Fragen** der AGBS (siehe Abbildung 1) durch die zwei Rater, und damit die Leistung der Ärzte in der **Stichprobe 1** (siehe 4.2) ist in Tabelle 1 dargestellt.

Tabelle 1

Bewertung der 22 Einzelfragen der AGBS für die Stichprobe 1 (n=40)

(Skala: 0 = „nicht beurteilbar“, 1 = „sehr gut“, 2 = „gut“, 3 = „befriedigend“, 4 = „schlecht“, 5 = „sehr schlecht“)

		Frage01 "Umgebung" (A)	Frage02 "Begrüßung & Vorstellung" (A)	Frage03 "Befinden" (A)			
Rater 1	Videos	40	40	40			
	Mittelwert	2,93	2,88	1,23			
	Median	3,00	3,00	,00			
	Minimum	0	0	0			
	Maximum	3	5	5			
	Standardabweichung	,474	1,202	2,106			
	Varianz	,225	1,446	4,435			
Rater 2	N	40	40	40			
	Mittelwert	2,98	2,80	,48			
	Median	3,00	3,00	,00			
	Minimum	2	0	0			
	Maximum	3	5	5			
	Standardabweichung	,158	1,305	1,377			
	Varianz	,025	1,703	1,897			
		Frage04 "Kenntnisstand" (B)	Frage05 "Warnschuss" (B)	Frage06 "Verständlichkeit der Nachricht" (B)	Frage07 "Tempo" (B)	Frage08 "Pause nach Nachricht" (B)	
Rater 1	Videos	40	40	40	40	40	
	Mittelwert	2,88	3,40	3,13	3,08	2,78	
	Median	3,00	3,00	3,00	3,00	3,00	
	Minimum	0	1	0	0	0	
	Maximum	5	5	5	4	4	
	Standardabweichung	1,159	1,236	1,067	,764	,800	
	Varianz	1,343	1,528	1,138	,584	,640	
Rater 2	Videos	40	40	40	40	40	
	Mittelwert	2,93	3,25	3,18	3,13	2,68	
	Median	3,00	3,00	3,00	3,00	3,00	
	Minimum	0	0	0	0	0	
	Maximum	5	5	5	5	5	
	Standardabweichung	1,289	1,581	1,152	,992	1,047	
	Varianz	1,661	2,500	1,328	,984	1,097	
		Frage09 "Rückfragen zulassen" (C)	Frage10 "Sorgen/Bedenken" (C)	Frage11 "Wichtigste Sorge" (C)			
Rater 1	Videos	40	40	40			
	Mittelwert	2,73	3,23	3,55			
	Median	3,00	3,00	3,00			
	Minimum	0	0	0			
	Maximum	4	5	5			
	Standardabweichung	,784	,768	1,108			
	Varianz	,615	,589	1,228			
Rater 2	Videos	40	40	40			
	Mittelwert	2,95	3,28	3,38			
	Median	3,00	3,00	3,00			
	Minimum	0	0	0			
	Maximum	4	5	5			
	Standardabweichung	,714	,784	,925			

		Frage12 "Adäquate Information" (D)	Frage13 "Struktur" (D)	Frage14 "Positive Aspekte" (D)	Frage15 "Faktisch korrekt" (D)
Rater 1	Videos	40	40	40	40
	Mittelwert	2,58	3,68	2,68	2,50
	Median	3,00	4,00	2,50	2,00
	Minimum	0	0	0	0
	Maximum	5	5	5	5
	Standardabweichung	1,394	1,207	1,366	1,468
	Varianz	1,943	1,456	1,866	2,154
Rater 2	Videos	40	40	40	40
	Mittelwert	2,30	3,28	2,78	2,53
	Median	2,00	3,00	2,50	2,00
	Minimum	0	0	0	0
	Maximum	5	5	5	5
	Standardabweichung	1,400	1,198	1,330	1,502
	Varianz	1,959	1,435	1,769	2,256

		Frage16 "Schlüsselbereiche" (E)	Frage17 "Psychosoziale Aspekte" (E)	Frage18 "Hilfreiche Unterstützung" (E)	Frage19 "Körpersprache" (E)	Frage20 "Unbeholfenheit" (E)	Frage21 "Tempo insgesamt" (E)	Frage22 "Ende" (E)
Rater 1	Videos	40	40	40	40	40	40	40
	Mittelwert	3,05	3,05	2,68	2,63	2,95	2,25	3,63
	Median	3,00	3,00	2,00	3,00	3,00	2,00	5,00
	Minimum	0	0	0	0	0	0	0
	Maximum	4	4	5	5	5	5	5
	Standardabweichung	,783	1,413	1,289	1,125	1,413	1,104	1,644
	Varianz	,613	1,997	1,661	1,266	1,997	1,218	2,702
Rater 2	Videos	40	40	40	40	40	40	40
	Mittelwert	3,18	3,18	2,70	2,50	2,78	2,18	3,60
	Median	3,00	3,00	2,00	2,00	3,00	2,00	5,00
	Minimum	0	0	0	0	0	0	0
	Maximum	5	5	5	5	5	5	5
	Standardabweichung	,903	1,338	1,400	1,155	1,368	1,217	1,676
	Varianz	,815	1,789	1,959	1,333	1,871	1,481	2,810

Zur besseren Veranschaulichung der Ergebnisse sind die **Mittelwerte und Standardabweichungen** für die **Stichproben 1 und 2** (Pre- und Re-Test) in der Bewertung der beiden Rater für die **einzelnen Fragen** als Balkendiagramm in **Abbildung 3 bis 8** aufgeführt.

Abbildung 3
Mittlere Bewertung der Einzelfragen durch RATER 1 (Stichprobe 1)

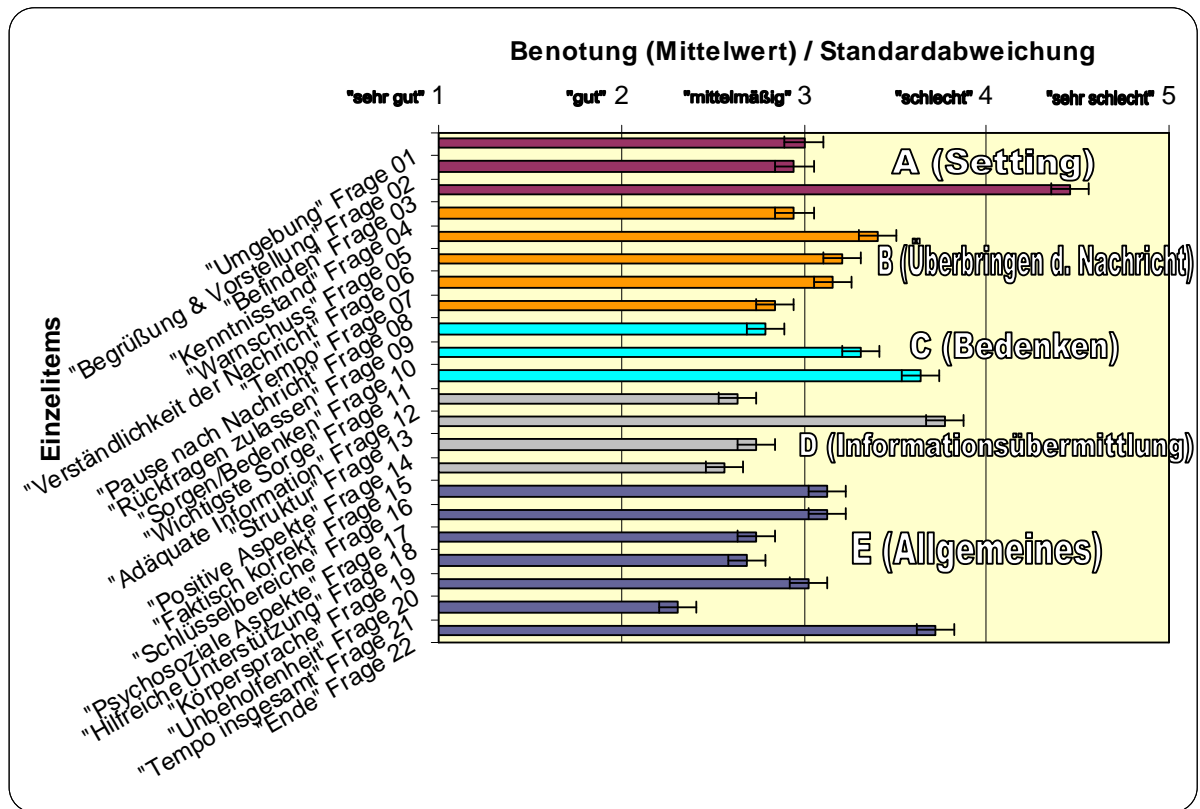


Abbildung 4
Mittlere Bewertung der Einzelfragen durch RATER 2 (Stichprobe 1)

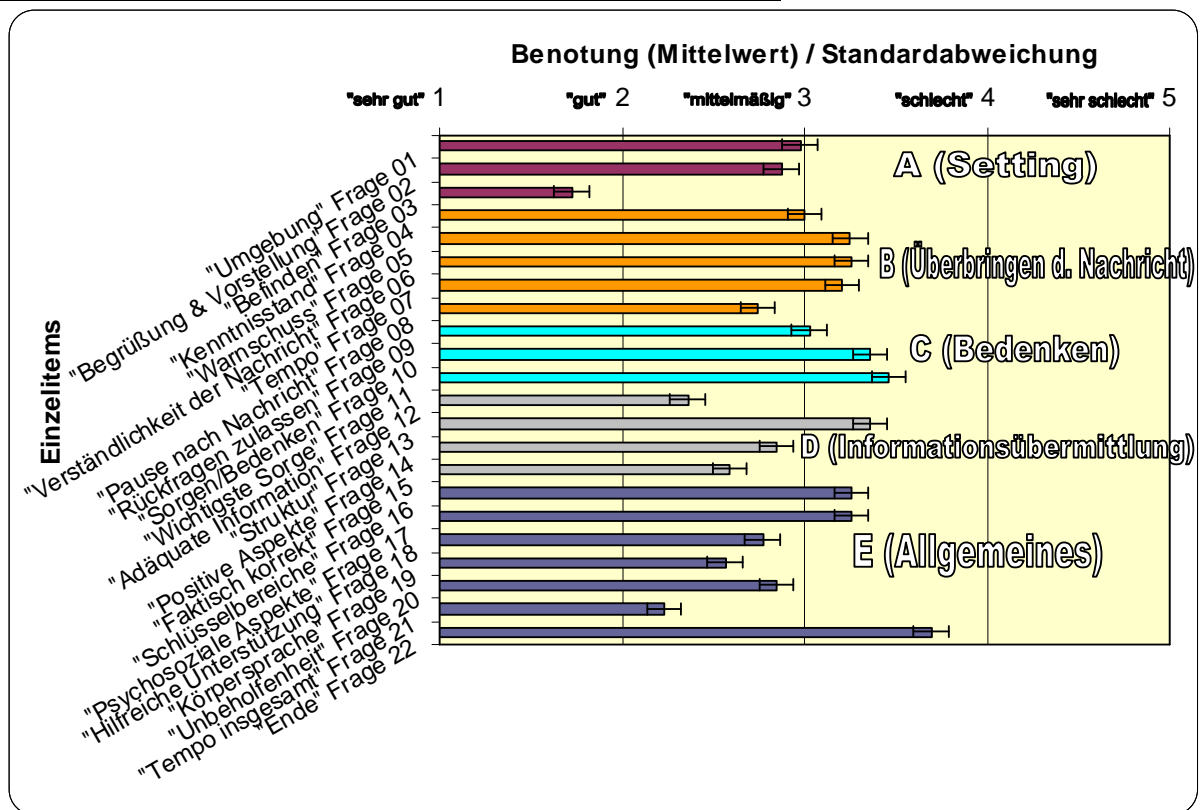


Abbildung 5
Mittlere Bewertung der Einzelfragen durch RATER 1 im Pre-Test (Stichprobe 2)

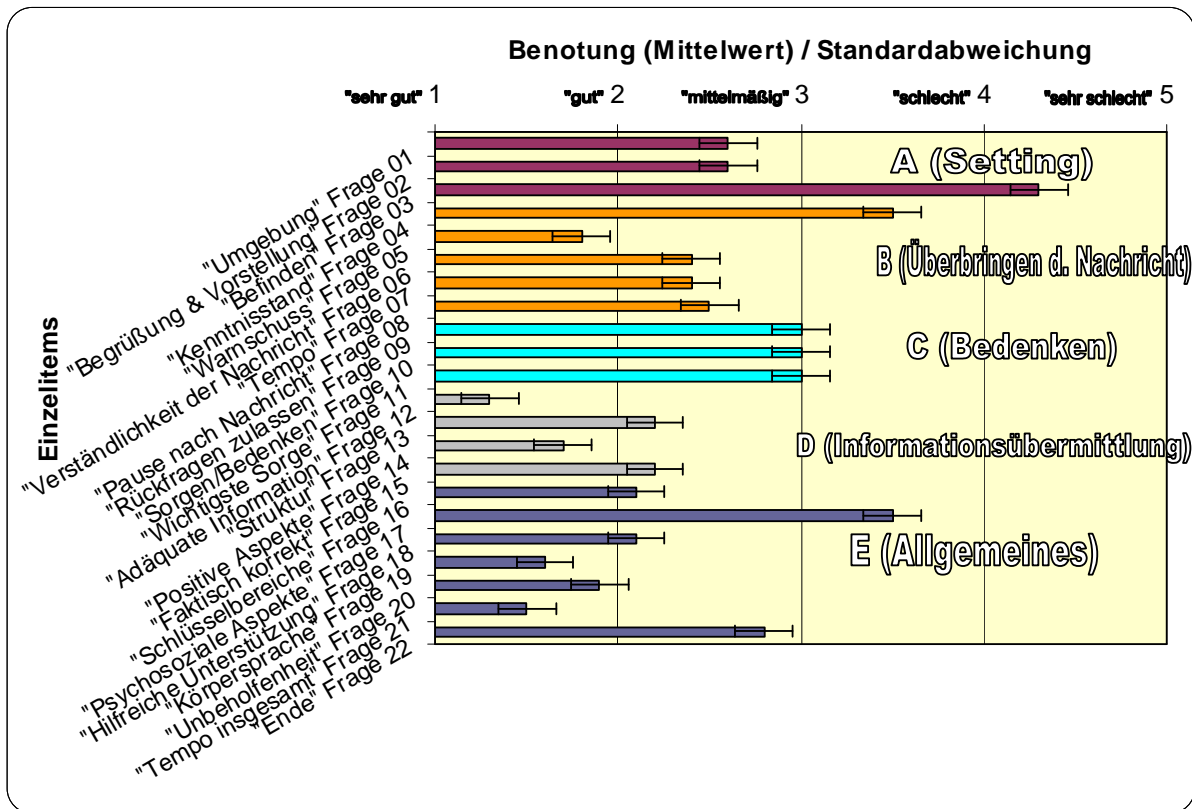


Abbildung 6
Mittlere Bewertung der Einzelfragen durch RATER 2 im Pre-Test (Stichprobe 2)

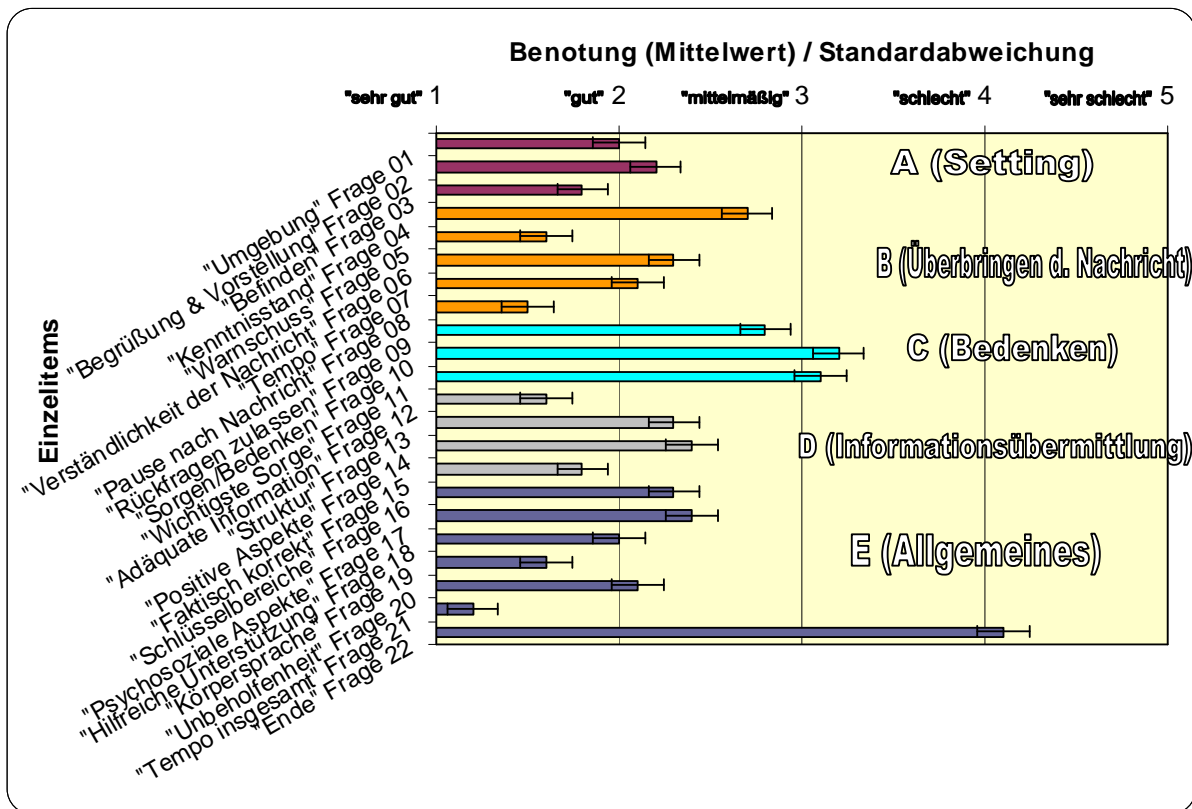


Abbildung 7
Mittlere Bewertung der Einzelfragen durch RATER 1 im Re-Test (Stichprobe 2)

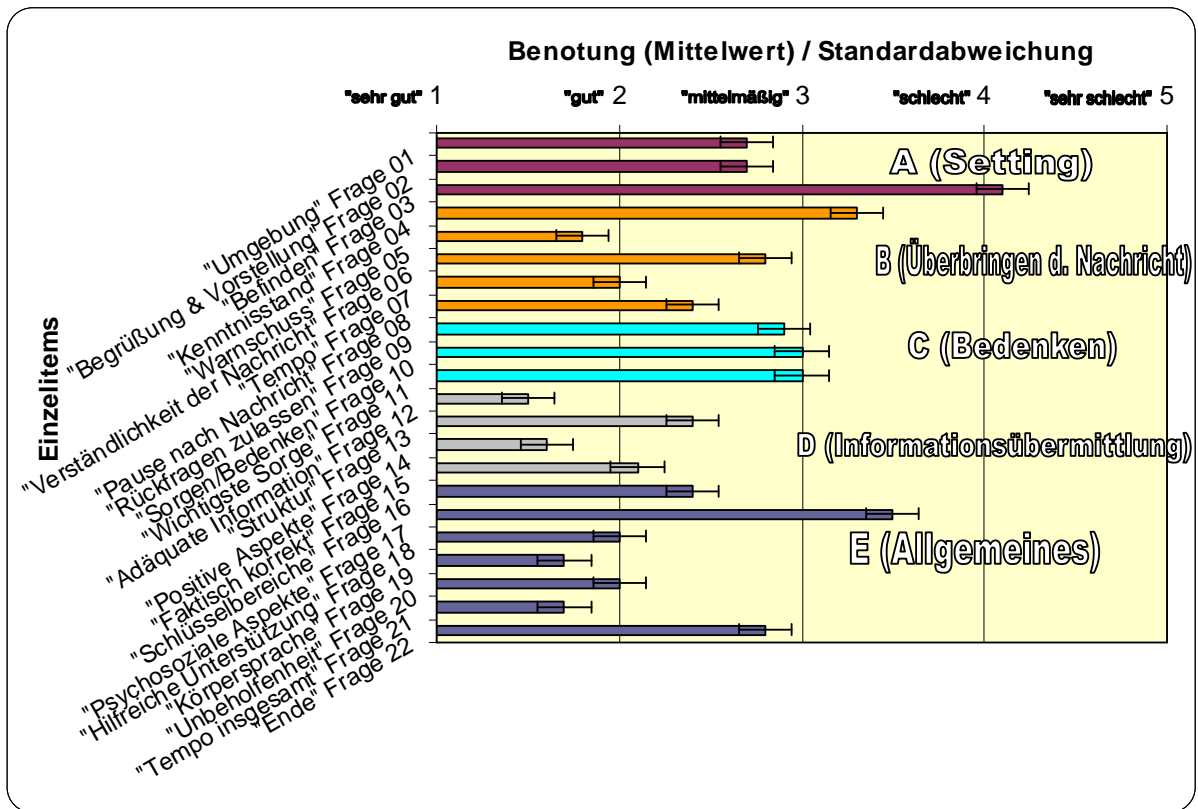
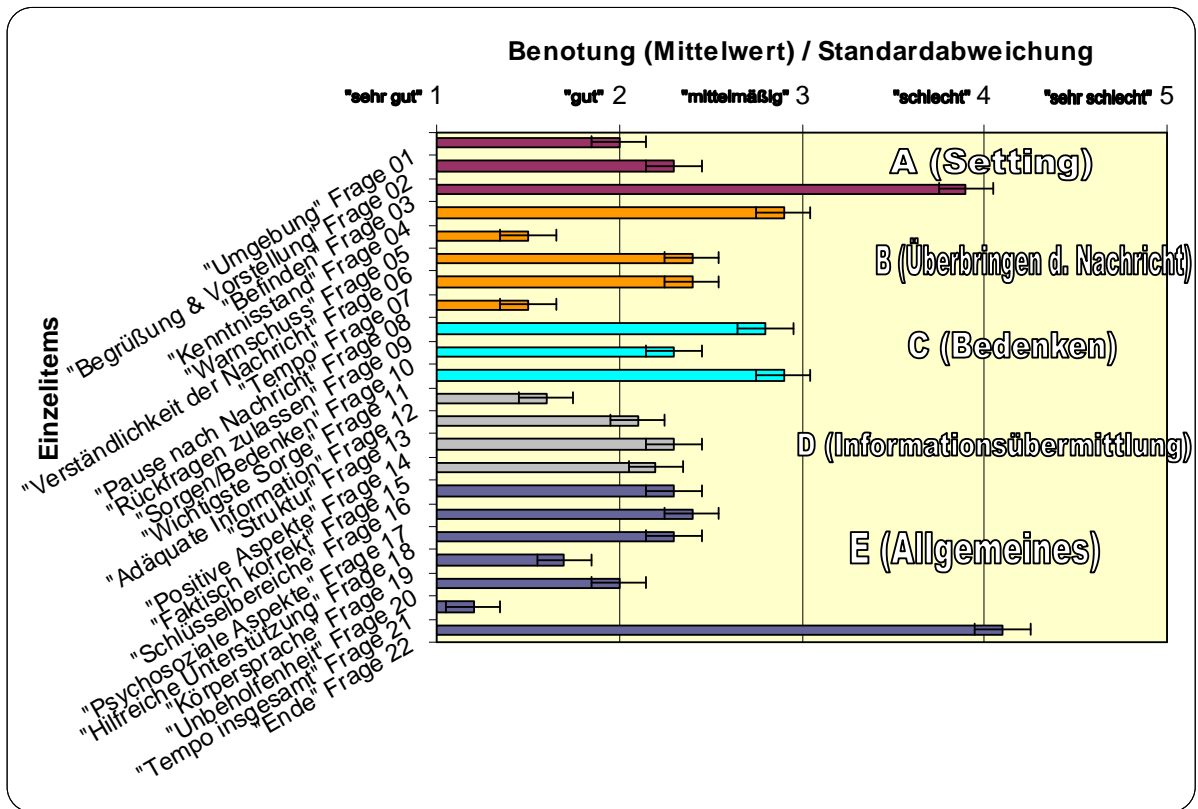


Abbildung 8
Mittlere Bewertung der Einzelfragen durch RATER 2 im Re-Test (Stichprobe 2)



Die **Zusammenfassung** der Bewertungen der Einzelitems als mittlere Bewertung der **Themenblöcke** (siehe Abbildung 1) und der Gesamtbewertung für die **Stichprobe 1** ist in Tabelle 2 aufgeführt.

Der Mittelwert der Gesamt-Bewertung der Gespräche durch die beiden Rater liegt bei 2,82 und 2,88. Dabei entspricht 1,0 der besten und 5,0 der schlechtesten Bewertung. Die geringste Differenz zwischen den mittleren Bewertungen der einzelnen Rater findet sich in den „Themenblöcken B“ (3,03-3,05) und „E“ (2,87-2,89), die größte in „Themenblock A“ (1,63–1,94).

Tabelle 2
Mittlere Bewertung der Themenblöcke beider Rater für die Stichprobe 1 (n=40)

	Mittelwert	Standardabweichung
Gesamtbewertung AGBS		
Rater 1	2,88	1,31
Rater 2	2,82	1,35
Themenblock A – Gestaltung der Szene		
Rater 1	2,34	1,62
Rater 2	2,08	1,58
Themenblock B – Überbringen der Nachricht		
Rater 1	3,05	1,04
Rater 2	3,03	1,24
Themenblock C – Bedenken des Patienten herausfinden		
Rater 1	3,17	0,96
Rater 2	3,20	0,83
Themenblock D – Informationsübermittlung		
Rater 1	2,86	1,43
Rater 2	2,72	1,40
Themenblock E – Allgemeine Überlegungen		
Rater 1	2,89	1,33
Rater 2	2,87	1,37

Die **Zusammenfassung** der Bewertungen der Einzelitems als mittlere Bewertung der **Themenblöcke** (siehe Abbildung 1) und der Gesamtbewertung für die **Stichprobe 2** ist in Tabelle 3 aufgeführt.

Der Mittelwert der Gesamt-Bewertung der Gespräche durch die beiden Rater liegt bei 2,45 und 2,23 im Pre-Test und 2,47 und 2,32 im Re-Test. Dabei entspricht wieder 1,0 der besten und 5,0 der schlechtesten Bewertung. Die geringste Differenz zwischen den mittleren Bewertungen der einzelnen Rater findet sich im „Themenblock E“ (2,21-2,24 bzw. 2,29-2,30) übereinstimmend für Pre-Test und Re-Test sowie in „Themenblock C“ (3,0-3,03) für den Pre-Test allein (siehe Abbildung 1). Die größte Differenz findet sich in „Themenblock A“ (2,0–3,17 bzw. 2,73-3,17) ebenfalls übereinstimmend für Pre-Test und Re-Test (siehe Tabelle 3).

Tabelle 3
Mittlere Bewertung der Themenblöcke
 beider Rater für die Stichprobe 2 (n=10)

	Mittelwert		Standardabweichung	
	Pre-Test	Re-Test	Pre-Test	Re-Test
Gesamtbewertung AGBS				
Rater 1	2,45	2,47	1,06	1,02
Rater 2	2,23	2,32	1,20	1,09
Themenblock A – Gestaltung der Szene				
Rater 1	3,17	3,17	1,26	1,15
Rater 2	2,00	2,73	1,05	1,11
Themenblock B – Überbringen der Nachricht				
Rater 1	2,52	2,46	0,99	0,99
Rater 2	2,04	2,14	1,17	1,04
Themenblock C – Bedenken des Patienten				
Rater 1	3,00	2,97	0,00	0,32
Rater 2	3,03	2,67	0,61	0,55
Themenblock D – Informationsübermittlung				
Rater 1	1,85	1,90	0,74	0,67
Rater 2	2,03	2,05	1,33	1,08
Themenblock E – Allgemeine Überlegungen				
Rater 1	2,21	2,30	1,12	1,09
Rater 2	2,24	2,29	1,29	1,23

Die **Zusammenfassung** der Mittleren Bewertung der **Themenblöcke** (siehe Abbildung 1) und der Gesamtbewertung für die **beide Stichproben** ist in Tabelle 4.aufgeführt.

Der Mittelwert der Gesamt-Bewertungen der Gespräche durch die beiden Rater in Pre-, Re-Test (Stichprobe 2) und Hauptauswertung (Stichprobe 1) liegt auf ähnlichem Niveau zwischen 2,23 und 2,88. Dabei entspricht wieder 1,0 der besten und 5,0 der schlechtesten Bewertung. Die Gesamtbewertungen durch den Rater 2 sind hierbei immer besser als durch den Rater 1. Die geringsten Differenzen zwischen den mittleren Bewertungen der einzelnen Rater finden sich übereinstimmend für Pre- (0,03), Re-Test (0,01) und Hauptauswertung (0,02) im „Themenblock E“ (siehe Abbildung 1). Die größten Differenzen finden sich in „Themenblock A“ ebenfalls übereinstimmend für Pre- (1,17), Re-Test (0,44) und Hauptauswertung (0,26) (siehe Tabelle 4.).

Tabelle 4
Mittlere Bewertung der Themenblöcke beider Rater und Stichproben

	Mittelwert			Standardabweichung			Videos		
	Pre-Test (Stich- probe 2)	Haupt- auswertung (Stich- probe 1)	Re-Test (Stich- probe 2)	Pre-Test (Stich- probe 2)	Haupt- auswertung (Stich- probe 1)	Re-Test (Stich- probe 2)	Pre-Test (Stich- probe 2)	Haupt- auswertung (Stich- probe 1)	Re-Test (Stich- probe 2)
Gesamtbewertung AGBS									
Rater 1	2,45	2,88	2,47	1,06	1,31	1,02	10	40	10
Rater 2	2,23	2,82	2,32	1,20	1,35	1,09	10	40	10
Themenblock A – Gestaltung der Szene									
Rater 1	3,17	2,34	3,17	1,26	1,62	1,15	10	40	10
Rater 2	2,00	2,08	2,73	1,05	1,58	1,11	10	40	10
Themenblock B – Überbringen der Nachricht									
Rater 1	2,52	3,05	2,46	0,99	1,04	0,99	10	40	10
Rater 2	2,04	3,03	2,14	1,17	1,24	1,04	10	40	10
Themenblock C – Bedenken des Patienten herausfinden									
Rater 1	3,00	3,17	2,97	0,00	0,96	0,32	10	40	10
Rater 2	3,03	3,20	2,67	0,61	0,83	0,55	10	40	10
Themenblock D – Informationsübermittlung									
Rater 1	1,85	2,86	1,90	0,74	1,43	0,67	10	40	10
Rater 2	2,03	2,72	2,05	1,33	1,40	1,08	10	40	10
Themenblock E – Allgemeine Überlegungen									
Rater 1	2,21	2,89	2,30	1,12	1,33	1,09	10	40	10
Rater 2	2,24	2,87	2,29	1,29	1,37	1,23	10	40	10

Aus den vorangegangenen Darstellungen lässt sich nun die Verteilung der Bewertungen ansehen und somit die **Varianz** der **Rater** untereinander sowie der einzelnen Rater in den verschiedenen **Durchläufen**. Diese Betrachtung bildet die **Grundlage** für die weiteren **statistischen Analysen**.

5.1 Validität und Interraterreliabilität der Hauptauswertung

Die **interne Konsistenz** für die mittlere Gesamtbewertung auf der AGBS beträgt für die Stichprobe 1 unter Verwendung von Cronbach's Alpha 0,93 (siehe Tabelle 5.).

Tabelle 5
Testvalidität für die Stichprobe 1 (n=40)

Cronbach's Alpha, Gesamtbewertung AGBS	Anzahl der Items
0,93	22

Hierbei und für die Homogenität steht der Aspekt im Vordergrund, wie gleichmäßig und wiederholbar die einzelnen Items oder Themenblöcke des Erhebungsinstruments zum Gesamtergebnis beitragen.

Die korrigierte Item-Skala-Korrelation als Maß für die **Trennschärfe** beträgt für die Gesamtbewertung der Stichprobe 1 0,993 und 0,996. Die geringste Trennschärfe zeigt sich für „Themenblock A“ mit 0,002 und 0,11 für die einzelnen Rater. Die höchsten Werte sind mit 0,84-0,88 in „Themenblock E“ zu finden (siehe Tabelle 6).

Tabelle 6
Trennschärfe in den Themenblöcken für die Stichprobe 1 (n=40)

	Korr. Item-Skala-Korrelation
Gesamtbewertung AGBS	
Rater 1	0,993
Rater 2	0,996
Themenblock A – Gestaltung der Szene	
Rater 1	0,11
Rater 2	0,002
Themenblock B – Überbringen der Nachricht	
Rater 1	0,68
Rater 2	0,74
Themenblock C – Bedenken des Patienten herausfinden	
Rater 1	0,81
Rater 2	0,84
Themenblock D – Informationsübermittlung	
Rater 1	0,80
Rater 2	0,84
Themenblock E – Allgemeine Überlegungen	
Rater 1	0,84
Rater 2	0,88

Der Intraklassen-Koeffizient (siehe Tabelle 7) zur Bestimmung der **Interrater-Reliabilität** nimmt für die Gesamtbewertung der Stichprobe 1 einen Wert von 0,93 ein (Minimum = 0, Maximum = 1). Die Werte für die einzelnen Themenblöcke (siehe Abbildung 1.) liegen hier zwischen 0,11 („Themenblock D“) und 0,90 („Themenblock E“).

Tabelle 7
Interraterreliabilität sowie Signifikanz-Niveaus für die **Inhomogenität der Mittelwerte** und Interaktion in den Themenblöcken für die Stichprobe 1 (n=40)

	Inhomogenität der Ratermittelwerte (p-Wert)	Interaktion zw. Rater und Proband (p-Wert)	ICC
Gesamtbewertung AGBS	0,10	0,38	0,93
Themenblock A – Gestaltung der Szene	0,00	0,04	0,11
Themenblock B – Überbringen der Nachricht	0,01	0,53	0,85
Themenblock C – Bedenken des Patienten	0,00	0,59	0,81
Themenblock D – Informationsübermittlung	0,08	0,65	0,88
Themenblock E – Allgemeine Überlegungen	0,00	0,55	0,90

Eine signifikante Inhomogenität der Ratermittelwerte zeigt sich für die mittlere Gesamtbewertung sowie für alle Themenblöcke einzeln.

Eine signifikante Interaktion zwischen Rater und Proband (siehe 5.1) ist nur für den „Themenblock A“ (siehe Abbildung 1) nachzuweisen.

Somit zeigt die Tabelle 5 auch die erfüllten Voraussetzungen (siehe 4.3.3) für die Gültigkeit des ICC.

Die weiteren Ergebnisse der **Varianzanalyse** für die Stichprobe 1 sind in Tabelle 8 dargestellt. Die Varianz der Gesamtbewertung ist zu 53,8 Prozent auf die Leistung der Probanden zurückzuführen. In den einzelnen Themenblöcken (siehe Abbildung 1) bewegt sich dieser Wert zwischen 45,5 Prozent („Themenblock C“) und 69,6 Prozent („Themenblock A“). Der auf einen unsystematischen Fehler zurückzuführende Varianzanteil (Interaktion zwischen Rater und Proband) liegt für beide Rater bei 45,7 Prozent.

Tabelle 8
Varianz der Bewertungen in den Themenblöcken durch Leistung d. Probanden, Inhomogenität d. Ratermittelwerte sowie Interaktion für d. Stichprobe 1 (n=40)

	Leistung d. Probanden	Inhomogenität der Ratermittelwerte	Interaktion zw. Rater und Proband
Gesamtbewertung AGBS	53,75%	69,75%	49,90%
Themenblock A – Gestaltung der Szene	45,50%	55,31%	52,91%
Themenblock B – Überbringen der Nachricht	0,60%	2,90%	0,20%
Themenblock C – Bedenken des Patienten	0,30%	1,20%	0,20%
Themenblock D – Informationsübermittlung	45,65 %	27,35 %	49,90 %
Themenblock E – Allgemeine Überlegungen	54,20 %	43,49 %	46,89 %

5.2 Validität, Interraterreliabilität und Re-Testreliabilität des Pre-/Re-Tests

Die **interne Konsistenz** für die mittlere Gesamtbewertung auf der AGBS beträgt in der Stichprobe 2 unter Verwendung von Cronbach's Alpha 0,89 für den Pre-Test beziehungsweise 0,84 für den Re-Test (siehe Tabelle 9.).

Tabelle 9
Testvalidität für die Stichprobe 2 (n=10)

Cronbach's Alpha, Gesamtbewertung AGBS		Anzahl der Items
Pre-Test	Re-Test	
0,89	0,84	22

Analog zu 5.1 steht hierbei und für die Homogenität der Aspekt im Vordergrund, wie gleichmäßig und wiederholbar die einzelnen Items oder Themenblöcke des Erhebungsinstruments zum Gesamtergebnis beitragen.

Wendet man diese Formel für Pre- und Re-Test der Stichprobe 2 jeweils eines Raters an (siehe 4.3.4), ergibt sich für den Rater 1 ein Ergebnis von 0,80 und für den Rater 2 eines von 0,91 (siehe Tabelle 10).

Tabelle 10
Validität des Pre-/Re-Tests
für die Stichprobe 2 (n=10)

Cronbach's Alpha, Gesamtbewertung AGBS		Anzahl der Items
Rater 1	Rater 2	
0,80	0,91	22

Der Intraklassen-Koeffizient zur Bestimmung der **Interrater-Reliabilität** für die Stichprobe 2 nimmt für die Gesamtbewertung der beiden Rater untereinander im Pre-Test einen Wert von 0,51 ein (Minimum = 0, Maximum = 1). Die Werte für die einzelnen Themenblöcke (siehe Abbildung 1) liegen hier zwischen 0,21 („Themenblock E“) und 0,81 („Themenblock B“). Im Vergleich dazu ergeben sich im Re-Test Werte von 0,81 für die Gesamtbewertung und 0,50 für „Themenblock E“ bis 0,91 für die „Themenblöcke B und C“ (siehe Tabelle 11).

Tabelle 11
Interraterreliabilität in den Themenblöcken
für die Stichprobe 2 (n=10)

	ICC	
	Pre-Test	Re-Test
Gesamtbewertung AGBS	,51	,81
Themenblock A – Gestaltung der Szene	,38	,78
Themenblock B – Überbringen der Nachricht	,81	,91
Themenblock C – Bedenken des Patienten	(keine Varianz)	,91
Themenblock D – Informationsübermittlung	,34	,60
Themenblock E – Allgemeine Überlegungen	,21	,50

Im Vergleich zur Interrater-Reliabilität von Hauptauswertung, Pre-Test und Re-Test wurde zusätzlich eine **Re-Testreliabilität**, sinngemäß als Intrarater-Reliabilität, bestimmt (siehe 4.3.3). Dies stellt ein Maß für die Bewertungsstabilität der Rater dar. Hierfür ergibt sich für die Gesamtbewertung von Rater 1 der ICC von 0,93 (Minimum = 0, Maximum = 1). Die Werte für die einzelnen Themenblöcke (siehe Abbildung 1) liegen hier zwischen 0,91 („Themenblock B“) und 0,95 („Themenblock A“). Im Vergleich dazu ergeben für Rater 2 Werte von 0,95 für die Gesamtbewertung und 0,96 für „Themenblock C“ bis 0,99 für „Themenblock B“ (siehe Tabelle 12).

Tabelle 12
Re-Testreliabilität in den Themenblöcken
für die Stichprobe 2 (n=10)

	ICC	
	Rater 1	Rater 2
Gesamtbewertung AGBS	,93	,95
Themenblock A – Gestaltung der Szene	,95	1,75
Themenblock B – Überbringen der Nachricht	,91	,99
Themenblock C – Bedenken des Patienten	(keine Varianz)	,96
Themenblock D – Informationsübermittlung	,93	,98
Themenblock E – Allgemeine Überlegungen	,93	,98

5.3 Validitäten und Interraterreliabilitäten insgesamt

Die **interne Konsistenz** für die mittlere Gesamtbewertung auf der AGBS beträgt unter Verwendung von Cronbach's Alpha 0,89 für den Pre-Test beziehungsweise 0,84 für den Re-Test und 0,93 für die Hauptauswertung (siehe Tabelle 13).

Tabelle 13
Testvalidität für drei Durchläufe und zwei Stichproben

Cronbach's Alpha, Gesamtbewertung AGBS			Anzahl der Items
Pre-Test (Stichprobe 2, n=10)	Hauptauswertung (Stichprobe 1, n=40)	Re-Test (Stichprobe 2, n=10)	
0,89	0,93	0,84	22

Hierfür ergibt sich also eine hohe Übereinstimmung von 91 beziehungsweise 95 Prozent. Analog zu 5.1 und 5.2 steht hierbei und für die Homogenität der Aspekt im Vordergrund, wie gleichmäßig und wiederholbar die einzelnen Items oder Themenblöcke des Erhebungsinstruments zum Gesamtergebnis beitragen.

Der Intraklassen-Koeffizient zur Bestimmung der **Interrater-Reliabilität** nimmt für die Gesamtbewertung in Pre-, Re-Test und Hauptauswertung einen Wert von 0,51, 0,81 und 0,93 ein (Minimum = 0, Maximum = 1). Tabelle 14 zeigt die Werte hier zwischen 0,21 beziehungsweise 0,50 („Themenblock E“) und 0,81 beziehungsweise 0,91 („Themenblock B“) für Pre- und Re-Test in den gleichen Themenblöcken (siehe Abbildung 1).

Tabelle 14
Interraterreliabilität in den Themenblöcken
 für drei Durchläufe und zwei Stichproben

	ICC		
	Pre-Test <i>(Stichprobe 2, n=10)</i>	Haupt-auswertung <i>(Stichprobe 1, n=40)</i>	Re-Test <i>(Stichprobe 2, n=10)</i>
Gesamtbewertung AGBS	,51	0,93	,81
Themenblock A – Gestaltung der Szene	,38	0,11	,78
Themenblock B – Überbringen der Nachricht	,81	0,85	,91
Themenblock C – Bedenken des Patienten	(keine Varianz)	0,81	,91
Themenblock D – Informationsübermittlung	,34	0,88	,60
Themenblock E – Allgemeine Überlegungen	,21	0,90	,50

Dem steht ein Intraklassen-Koeffizient für die Gesamtbewertung in der Hauptauswertung mit einem Wert von 0,93 gegenüber. Die Werte für die einzelnen Themenblöcke (siehe Abbildung 1) liegen hier ebenfalls insgesamt anders verteilt zwischen 0,11 („Themenblock D“) und 0,90 („Themenblock E“).

6 Diskussion – Ziele und Nutzen der Untersuchung

Neben der Überprüfung der Aufklärungsgesprächbewertungsskala (AGBS) tragen die Ergebnisse der vorliegenden Replikationsstudie zur Erweiterung der bisherigen Forschung bei und zeigen somit auch ein Vorgehen für folgende Projekte.

Gemäß den Anforderungen die im täglichen Berufsleben von Ärzten an deren Kommunikationsfähigkeiten gestellt werden (siehe 1.2), ergibt sich der Bedarf, die Leistung des Arztes (siehe 2 sowie 2.1 und 2.2) zu erfassen, zu bewerten und zu vergleichen. Der Zweck der AGBS ist nun die Erfassung und Bewertung der Gesprächsführungskompetenz von Ärzten.

Im Folgenden wird beschrieben, wie die vorliegende Untersuchung belegt, dass sich die AGBS insgesamt sehr gut eignet und weiterentwickelt werden kann.

Hierzu zählt in der Konsequenz auch die Verwendung der Erkenntnisse in der Ausbildung von Studenten und Ärzten.

In der Auseinandersetzung mit dem Thema dieser Arbeit stellte sich heraus, dass es noch einer gewissen Vernetzung oder Standardisierung im Vorgehen allgemein bedarf. Durch vergleichende Untersuchungen (siehe 2.1) wurde diese Auffassung sinngemäß bestätigt. Einen weiteren Beleg hierfür stellt das Ergebnis einer Studie dar, die anhand von 26 Veröffentlichungen von Instrumenten zur ärztlichen Kommunikationsauswertung zwischen 1990 und 2002 systematische Vergleiche anstellte (Cegala und Lenzmeier Broz 2002). Die Autoren dieser Übersichtsarbeit führten die Studiendesigns und die Form der Bewertung auf – sowie die Stichproben und die Art der Interventionen beziehungsweise Schulungen und untersuchten Fähigkeiten. Dadurch kamen sie zu dem Ergebnis, dass es einer besseren Abstimmung zwischen der Lehre von Kommunikationsfertigkeiten und der Bewertung

derselben bedarf. Weiterhin sollte mehr Wert auf fundierte, standardisierte Interventionen mit dem Fokus auf das Timing bei der Gesprächsführung gelegt werden. Und zuletzt erscheint den Autoren die Durchführung und Dokumentation eines standardisierten Ratertrainings, sinngemäß als Gütekriterium, sinnvoll.

Dadurch wird noch einmal die Notwendigkeit betont, allgemein die Untersuchungsbedingungen zu berücksichtigen beziehungsweise zu vergleichen, wie es in der vorliegenden Studie betrieben wird.

Diese Feststellungen zeigen somit die Beweggründe und Ziele der vorliegenden Arbeit noch einmal auf. Es wird hiermit eine wissenschaftlich konsequente Fortführung in Form der Überprüfung durch die Replikationsstudie betrieben. Ferner wird die Standardisierung des Ratertrainings, auch als wichtige Voraussetzung für die Testgüte (siehe 2.2 und 4.3), angestrebt. Letztlich konnte bereits die Anwendung des Instrumentes in Projekten weiterführender Forschung ebenso wie in solchen der Lehre vorangebracht werden.

6.1 Vergleich der Untersuchungsbedingungen

Da es sich bei der „Aufklärungsgesprächbewertungsskala“ (AGBS) (Wand et al. 2008) um eine Übertragung des „Breaking Bad News Assessment Schedule“ (BAS) (Miller et al. 1999) handelt, ergibt sich die Frage nach der **Vergleichbarkeit** der vorliegenden Arbeit zu den beiden Veröffentlichungsarbeiten (siehe Abbildung 9).

Es wird einerseits das **Gefüge der Stichproben** aus den Videos analysiert, sowohl im Hinblick auf die Probanden, beziehungsweise **Ärzte**, und deren Ausbildungsstand, als auch mit Blick auf die unterschiedlichen **Gesprächspartner** und deren Anzahl. Andererseits wird die **Zusammensetzung der Rater** berücksichtigt.

Abbildung 9

Studienbedingungen in der Untersuchungsreihe zur Gesprächsbewertung

Bedingungen	„Breaking Bad News Assessment Schedule“ (Miller et al. 1999)	„Aufklärungsgespräch-bewertungsskala“ (Wand et al. 2008)	Vorliegende Replikationsstudie
Stichproben			
- Ärzte	X		X
- Studenten	X	X	X
- Standardisierte Patienten	X	X	
- Mehrere Gesprächspartner			X
Rater			
- Mediziner	X	X	X
- Nicht Mediziner			X

Innerhalb des Kontextes „**Überbringen schlechter Nachrichten**“ im ärztlichen Alltag unterscheidet sich die zu untersuchende beziehungsweise auszuwertende Stichprobe der vorliegenden Arbeit von der in der Originalarbeit (Wand et al. 2008) verwendeten (siehe 2.2).

Wie bereits in der Veröffentlichungsarbeit des „BAS“ (Miller et al. 1999) praktiziert, wurden in diesem Fall auch **Ärzte mit Berufserfahrung** miteinbezogen. Weiterhin handelt es sich um Situationen mit **mehreren Kommunikationspartnern** des Arztes, im Vergleich zu Gesprächen mit nur einem Patienten in den beiden zugrundeliegenden Veröffentlichungen.

Diese höhere **Varianz in der Stichprobe** ermöglicht einen Vergleich der Ergebnisse für einen größeren Anwendungsbereich und verschiedene Trainingssituationen.

6.2 Vergleich der Ergebnisse mit externen Daten

Im Folgenden wird die Zusammenschau der Ergebnisse aller im Rahmen der vorliegenden Arbeit erfolgten Bewertungsdurchgänge dargelegt (siehe 5.3). Des Weiteren erfolgt der Vergleich mit den bislang veröffentlichten Daten, im Sinne einer Verortung dieser Untersuchung im wissenschaftlichen Kontext.

6.2.1 Pilotstudie – Pre-Test

Am 21.07.2008 fand der Workshop „Methoden der Schulung in Gesprächsführung“ statt, bei dem neben weiteren Experten zum Thema auch die Erstautorin der Veröffentlichungsarbeit der AGBS, Frau Wand, teilnahm. Dort konnten nach einer Präsentation der Ergebnisse durch Frau Wand auch die Ergebnisse einer ersten Auswertung der Stichprobe 2 (siehe 4.2) vom Autor dieser Arbeit präsentiert und verglichen werden (siehe Tabelle 15 bis 17).

In einer persönlichen Mitteilung der Autorin war zu erfahren, dass die Ergebnisse einer Evaluationsstudie (siehe 1.2) mithilfe der AGBS, noch gesondert ausgewertet und veröffentlicht werden.

Tabelle 15

Bewertung der Themenblöcke für den **Pre-Test „München“** (Stichprobe 2) vergleichsweise zur Erstveröffentlichung **„Erlangen“** (Wand et al. 2008)

	Mittelwert		Standardabweichung		Videos	
	Erlangen	München	Erlangen	München	Erlangen	München
Gesamtbewertung AGBS						
Rater 1	2,06	2,45	,36	1,06	58	10
Rater 2	2,14	2,23	,39	1,20	58	10
Rater 3	1,98		,39		58	
Themenblock A – Gestaltung der Szene						
Rater 1	1,78	3,17	,71	1,26	63	10
Rater 2	1,77	2,00	,71	1,05	63	10
Rater 3	1,98		1,09		63	
Themenblock B – Überbringen der Nachricht						
Rater 1	2,56	2,52	,52	0,99	72	10
Rater 2	2,73	2,04	,63	1,17	72	10
Rater 3	2,66		,65		72	
Themenblock C – Bedenken des Patienten						
Rater 1	2,38	3,00	,31	0,00	72	10
Rater 2	2,40	3,03	,41	0,61	72	10
Rater 3	2,41		,33		72	
Themenblock D – Informationsübermittlung						
Rater 1	1,90	1,85	,55	0,74	74	10
Rater 2	1,74	2,03	,53	1,33	74	10
Rater 3	1,67		,44		74	
Themenblock E – Allgemeine Überlegungen						
Rater 1	1,87	2,21	,51	1,12	73	10
Rater 2	1,94	2,24	,49	1,29	73	10
Rater 3	1,63		,50		73	

Tabelle 16

Testvalidität für den **Pre-Test „München“** (Stichprobe 2, n=10) vergleichsweise zur Erstveröffentlichung **„Erlangen“** (Wand et al. 2008)

Cronbach's Alpha, Gesamtbewertung AGBS		Anzahl der Items
Erlangen 0,88	München 0,89	22

Tabelle 17

Interraterreliabilität in den Themenblöcken

für den **Pre-Test „München“** (Stichprobe 2, n=10) vergleichsweise zur Erstveröffentlichung **„Erlangen“** (Wand et al. 2008)

	ICC	
	Erlangen	München
Gesamtbewertung AGBS	,86	,51
Themenblock A – Gestaltung der Szene	,94	,38
Themenblock B – Überbringen der Nachricht	,81	,81
Themenblock C – Bedenken des Patienten	,63	(keine Varianz)
Themenblock D – Informationsübermittlung	,61	,34
Themenblock E – Allgemeine Überlegungen	,76	,21

Dass die Ergebnisse der Gütekriterien für den **Pre-Test unter dem Niveau der übrigen Auswertungen** liegen (siehe 5.3 und Tabelle 17), beziehungsweise nicht in allen Qualitäten gut vergleichbar sind, lässt sich durch den frühen Zeitpunkt und die noch nicht abgeschlossene Raterschulung erklären.

Das sehr gute **Ergebnis für die Testvalidität** (siehe Tabelle 16) belegt allerdings, dass bereits zu diesem **frühen Zeitpunkt eine hohe Testgüte** vorliegt, und bestärkte letztlich zur Durchführung der nachfolgenden Untersuchungen.

6.2.2 Expertenrating

Auf dem Workshop „Methoden der Schulung in Gesprächsführung“ am 21.07.2008 zeigte Frau Silke Neuderth von der medizinischen Fakultät der Universität Würzburg in ihrem Vortrag zum Einsatz von Schauspielpatienten die durchweg positive Einschätzung im Hinblick auf den Kompetenzgewinn. So fühlten sich beispielsweise 90 Prozent der befragten Studenten nach Absolvierung des klinischen Fachsemesters mit dem Inhalt der Kommunikationsschulung im Rahmen der Onkologie besser dazu in der Lage, eine schlechte Nachricht zu überbringen.

Bezüglich der Systematik zum Beobachtungstraining erläuterte die Erstautorin der AGBS, Frau Wand, das Ratertraining im Rahmen ihrer Arbeit, wobei eine Ähnlichkeit mit dem Vorgehen für diese Arbeit festgestellt wurde. In beiden Fällen lag neben der inhaltlichen Beurteilung der Fokus darauf, ein einheitliches Vorgehen unter den Ratern abzustimmen (siehe 3).

Frau Neuderth konnte hierzu ihre Erkenntnisse im Bezug auf das Beurteilungstraining und die Schulung der standardisierten Schauspielpatienten beitragen. Ein knapper Leitfaden, wie das dort verwendete „SPIKES“-Protokoll (Baile et al. 2000), welches bereits unter 1.2 und 2.1 erwähnt wurde, ist günstiger zu bewerten als eine ausführlichere, langwierigere Methode. Bei einer zu detaillierten

Vorbereitung wurde sich seitens der Studenten zu sehr auf Einzelheiten und leicht erlernbare Fakten konzentriert. Dies führte aber zu insgesamt eher schlechteren Ergebnissen in der Bewertung durch die Dozenten.

Alle genannten Überlegungen und die Ergebnisse für die Konstanz der Raterleistung durch das Re-Testverfahren führten schließlich zu einem Entwurf für das zukünftige Ratertraining.

6.2.3 Hauptauswertung

Wie bereits für die **Pilotstudie** und damit die Stichprobe 2 erfolgt, lassen sich auch die Ergebnisse für die Hauptauswertung, die **Stichprobe 1**, mit denen der Publikation des **BAS** (Miller et al. 1999) beziehungsweise der **AGBS** (Wand et al. 2008) vergleichen.

Für die Ergebnisse mittels **BAS** in Punkten (siehe 2.2) wurde gemäß Tabelle 18 eine lineare **Transformation** in Notenschritte (AGBS) durchgeführt, um eine Vergleichbarkeit zu erhalten.

Tabelle 18
Transformation Score (BAS) \approx Noten (AGBS)

Prozentanteil	\approx	Score/Punkte	\approx	Note
1-20	\approx	23-41	\approx	1
21-40	\approx	42-60	\approx	2
41-60	\approx	61-78	\approx	3
61-80	\approx	79-97	\approx	4
81-100	\approx	98-115	\approx	5

Die Tabellen 19 und 20 listen deskriptiv die mittleren Bewertungen für die Stichprobe 1 neben denen der Erstveröffentlichungen auf.

Tabelle 19

Gesamtbewertung für die Stichprobe 1 „München“ vergleichsweise zu den Erstveröffentlichungen „Oxford“ (Miller et al. 1999) und „Erlangen“ (Wand et al. 2008)

	Mittelwert			Standardabweichung			Videos		
	Oxford	Erlangen	München	Oxford	Erlangen	München	Oxford	Erlangen	München
Gesamtbewertung AGBS									
Rater 1	2,80	2,06	2,88	k. A.	0,36	1,31	20	58	40
Rater 2	2,35	2,14	2,82	k. A.	0,39	1,35	20	58	40
Rater 3	2,25	1,98		k. A.	0,39		20	58	

Da die Daten für die Standardabweichungen (siehe Tabelle 19) und die einzelnen Themenblöcke (siehe Tabelle 20) des BAS nicht veröffentlicht sind, und auch auf Anfrage nicht zu beschaffen waren, können diese nicht mit aufgeführt werden.

Tabelle 20

Bewertung in den Themenblöcken für die Stichprobe 1 „München“ im Vergleich zu der Erstveröffentlichung „Erlangen“ (Wand et al. 2008)

	Mittelwert		Standardabweichung		Videos	
	Erlangen	München	Erlangen	München	Erlangen	München
Gesamtbewertung AGBS						
Rater 1	2,06	2,88	0,36	1,31	58	40
Rater 2	2,14	2,82	0,39	1,35	58	40
Rater 3	1,98		0,39		58	
Themenblock A – Gestaltung der Szene						
Rater 1	1,78	2,34	0,71	1,62	63	40
Rater 2	1,77	2,08	0,71	1,58	63	40
Rater 3	1,98		1,09		63	
Themenblock B – Überbringen der Nachricht						
Rater 1	2,56	3,05	0,52	1,04	72	40
Rater 2	2,73	3,03	0,63	1,24	72	40
Rater 3	2,66		0,65		72	
Themenblock C – Bedenken des Patienten						
Rater 1	2,38	3,17	0,31	0,96	72	40
Rater 2	2,40	3,20	0,41	0,83	72	40
Rater 3	2,41		0,33		72	
Themenblock D – Informationsübermittlung						
Rater 1	1,90	2,86	0,55	1,43	74	40
Rater 2	1,74	2,72	0,53	1,40	74	40
Rater 3	1,67		0,44		74	
Themenblock E – Allgemeine Überlegungen						
Rater 1	1,87	2,89	0,51	1,33	73	40
Rater 2	1,94	2,87	0,49	1,37	73	40
Rater 3	1,63		0,50		73	

Es zeigt sich ein **gleichmäßiges Bild für die Mittelwerte** der Benotung über alle Entwicklungsstufen hinweg, also die Veröffentlichung des **BAS** („Oxford“), der **AGBS** („Erlangen“) und die Hauptauswertung der vorliegenden **Replikation** („München“). Es fällt allerdings eine größere Gemeinsamkeit der etwas schlechteren Benotungen zwischen „Oxford“ und „München“ gegenüber „Erlangen“ auf.

Für die folgenden **Interpretationen** wird eine relativ konstante absolute Gesprächsleistung der Probanden in den drei Stichproben angenommen. Diese Überlegung beinhaltet, dass die tatsächlichen Unterschiede in der Gesprächsleistung geringer sind, als in der Bewertung mittels der Instrumente widergespiegelt.

Der Trend insgesamt positiverer Bewertungen in der Erstveröffentlichung der AGBS, „Erlangen“ (siehe Tabellen 19 und 20) kann zum einen auf die genaue Anpassung des Instrumentes mit Hilfe der Appendizes (siehe Abbildung 16) auf die untersuchte Stichprobe zurückgeführt werden. Zum anderen kann die Ähnlichkeit der Ergebnisse von „Oxford“ und „München“ (siehe Tabelle 19) mit der höheren Übereinstimmung in der Zusammensetzung der Probanden in Verbindung gebracht werden. So stimmt hier die Varianz der Fähigkeiten der Ärzte in der untersuchten Stichprobe eher überein (siehe 4.1).

Wegen diesen vergleichsweisen Uneinheitlichkeiten in der Stichprobe (siehe 4.1) ergeben sich unter anderem auch etwas inhomogenere Ratermittelwerte mit unterschiedlichen Verteilungen und deutlich weniger Interaktion im Vergleich der Auswertungen „Erlangen“ und „München“ (siehe Tabellen 20 und 21).

Dies trifft vor allem für den „Themenblock E“ („Allgemeine Überlegungen“, siehe Abbildung 1) zu. Dies bietet einen Hinweis auf die Unschärfe in den Themenbereichen „Emotionen“ und „Psychosoziales“.

Tabelle 21

Varianz der Bewertung in den Themenblöcken für die Stichprobe 1 (n=40) „München“ vergleichsweise zur Erstveröffentlichung „Erlangen“ (Wand et al. 2008)

	Anteil d. Leistung d. Probanden		Inhomogenität der Ratermittelwerte		Interaktion zw. Rater und Proband	
	Erlangen	München	Erlangen	München	Erlangen	München
Gesamtbewertung AGBS	77,61 %	53,75%	3,14 %	0,60%	19,26 %	45,65 %
Themenblock A – Gestaltung der Szene	88,45 %	69,75%	1,22 %	2,90%	10,33 %	27,35 %
Themenblock B – Überbringen der Nachricht	72,41 %	49,90%	1,24 %	0,20%	26,35 %	49,90 %
Themenblock C – Bedenken des Patienten	57,34 %	45,50%	0,15 %	0,30%	42,51 %	54,2 %
Themenblock D – Informationsübermittlung	55,45 %	55,31%	3,46 %	1,20%	41,09 %	43,49 %
Themenblock E – Allgemeine Überlegungen	66,17 %	52,91%	6,65 %	0,20%	27,18 %	46,89 %

Zur **Beurteilung eines Testverfahrens** wie der AGBS **benötigt man Gütekriterien.**

Zum einen wären dies die Verlässlichkeit beziehungsweise Vergleichbarkeit, die Reliabilität zwischen verschiedenen Beobachtern (siehe 4.3.3); zum anderen die Gültigkeit, die sogenannte Validität (siehe 4.3.1), eines Verfahrens. Diese Kriterien zur Testgüte wurden in der vorliegenden Arbeit gemäß den Veröffentlichungsarbeiten (Miller et al. 1999; Wand et al. 2008) verifiziert.

Einen **Beleg für den Sinn der Anwendung** des Instrumentes liefern die hauptsächlich sehr guten und im Vergleich sogar besseren Werte für die **Interraterreliabilität** und **Testvalidität** (siehe Tabellen 22 und 23). Hier zeigen sich lediglich **geringe Schwächen** im Bereich der Beurteilung des **Settings** (siehe Tabelle 23.). Die diesbezüglich im Moment festzustellenden Unzulänglichkeiten fallen insgesamt für die Testgüte jedoch nicht ins Gewicht (siehe Tabelle 22 und 23).

Tabelle 22
Testvalidität für die Stichprobe 1 (n=40) „München“
vergleichsweise zu den Erstveröffentlichungen „Oxford“
 (Miller et al. 1999) und „Erlangen“ (Wand et al. 2008)

Cronbach's Alpha, Gesamtbewertung AGBS			Anzahl der Items
Oxford	Erlangen	München	
0,93	0,88	0,93	22

Tabelle 23
Interraterreliabilität in den Themenblöcken
 für den **Pre-Test** „München“ (Stichprobe 2, n=10) **vergleichsweise**
 zur Erstveröffentlichung „Erlangen“ (Wand et al. 2008)

	Inhomogenität der Ratermittelwerte (p-Wert)		Interaktion zw. Rater und Proband (p-Wert)		ICC	
	Erlangen	München	Erlangen	München	Erlangen	München
Gesamtbewertung AGBS	0,0002	0,09988	0,7914	0,3800	0,86	0,93
Themenblock A – Gestaltung der Szene	0,0010	0,00381	<0,0000	0,0400	0,94	0,11
Themenblock B – Überbringen der Nachricht	0,0383	0,00536	0,3617	0,5300	0,81	0,85
Themenblock C – Bedenken des Patienten	0,7748	0,00000	0,5191	0,5900	0,63	0,81
Themenblock D – Informationsübermittlung	0,0027	0,07643	0,0253	0,6500	0,61	0,88
Themenblock E – Allgemeine Überlegungen	<0,0000	0,00003	0,9814	0,5500	0,76	0,90

Die Ergebnisse für die Inhomogenität der Ratermittelwerte (siehe 5.2) zeigen eine gute Vergleichbarkeit zwischen der Veröffentlichung der AGBS, „Erlangen“ und der vorliegenden Replikation, „München“ (siehe Tabelle 23). Die Mittelwerte sind für alle Themenbereiche einzeln und die Gesamtbewertung inhomogen verteilt.

Eine Interaktion zwischen Rater und Proband (siehe 5.1 und 5.2) zeigt sich übereinstimmend für „Erlangen“ und „München“ lediglich für den „Themenblock A“ (Gestaltung der Szene, siehe Abbildung 1).

Bei den bislang durchgeführten Testreihen herrschten insgesamt mehr oder weniger große **Unterschiede** bei den Verfahren im Hinblick auf **Rater** und **Stichproben**. Trotzdem lässt sich **durchgehend** von der Veröffentlichung des BAS (Miller et al. 1999) bis hin zur letzten Auswertung in dieser Arbeit, dem Re-Test, eine **stabile**

Testvalidität auf hohem Niveau konstatieren (siehe Abbildung 10). Das ist auch die wichtigste Gemeinsamkeit zwischen den zwei in der vorliegenden Arbeit untersuchten Stichproben (siehe 4.2).

Abbildung 10
Testvalidität im Verlauf



Darum lässt sich an dieser Stelle feststellen, dass das Instrument über alle Entwicklungsstufen und Unterschiede in den Testumgebungen eine gleichbleibend **hohe Güte** erfüllt und deswegen der **Einsatz** dieses Verfahrens **zu empfehlen** ist.

6.2.4 Abschätzung von Trainingseffekten im Re-Test

Um die Aussagekraft im Bezug auf die Testgüte zu erhöhen und die Raterstabilität quantifizieren zu können, wurde das Studiendesign erweitert (siehe 3 und 6.1). Es wurde ein **Re-Test-Verfahren** für die **Stichprobe 2** angewandt. Hierzu wurde nach **sechs Monaten** das Rating von zehn Videoaufnahmen wiederholt, die bereits als Pre-Test vor der Hauptauswertung bearbeitet worden waren.

Zu den wiederum stabilen Mittelwerten zeigte sich eine im Re-Test stark verbesserte Interraterreliabilität fast auf dem Niveau der Hauptauswertung beziehungsweise der Originalarbeit (Wand et al. 2008). Diese Verbesserung ist Ausdruck der in der Zwischenzeit fortgeführten Raterschulung (siehe Abbildung 11).

Abbildung 11

Abschätzung von Trainingseffekten (ICC im Verlauf)



Die **Re-Rate-Reliabilität** oder Intraraterreliabilität (siehe 4.3.4) spiegelt die **Stabilität der Leistung der einzelnen Rater über einen Zeitraum von sechs Monaten** wieder.

Die sehr guten Ergebnisse, sowohl für Cronbach's Alpha von 0,8 und 0,91 als auch für den ICC von jeweils über 0,90 für beide Rater zeigt (siehe 4.3.4), dass die Bewertung von Gesprächen durch die AGBS über einen längeren Zeitraum und unterschiedliche Trainingszustände konstant bleibt. Außerdem wird somit der **Langzeiteffekt des Ratertrainings** belegt. Dies wurde bisher selten untersucht.

6.3 Zusammenhang zwischen Intuition und Bewertung

Die Rater beobachteten übereinstimmend, dass ein positives oder negatives „Bauchgefühl“ mit dem Ergebnis der Bewertung durch die AGBS in den meisten Fällen korrelierte (siehe 3). Hierbei stellte der Zensurenschritt der Benotung zwischen „zwei“ und „drei“ (siehe Abbildung 1) den Wendepunkt zwischen „positiv“ und „negativ“ dar.

Wenn die Intuition und die Bewertung nicht die gleiche Tendenz hatten, handelte es sich entweder um eine überschätzte Gegenübertragung des den Arzt Spielenden, so dass eine gefühlte Antipathie sich nicht auch in einer schlechteren Bewertung widerspiegelte. Oder es schlug sich eine sympathische Haltung dem Arzt gegenüber nicht in einer positiven Bewertung nieder.

In einigen wenigen Videoaufzeichnungen fand sich also eine, von der Benotung relativ unabhängige, zwischen den Ratern aber fast immer übereinstimmende, positive gleichwie negative Einstellung gegenüber den empathischen Fähigkeiten der Ärzte.

6.4 Emotionen

Im Verständnis des **biopsychosozialen Krankheitsmodells** (siehe 2.2) spielen Emotionen allgemein bei der Interaktion mit medizinischem Personal und speziell bei Aufklärungsgesprächen durch Ärzte (siehe 1.1) eine wichtige Rolle.

Es ergibt sich die Anforderung an Gesprächsauswertungssysteme (siehe 2) Emotionen und psychosoziale Themen zu bewerten. Dem kommen die momentan verfügbaren Instrumente nur unzureichend nach. Sei es durch ein Fehlen an Erfassungselementen, fehlender Bewert- oder Vergleichbarkeit (siehe 2.1) oder minderer Testgüte für diesen Bereich, wie es in der vorliegenden Arbeit bedingt nachvollzogen werden konnte.

Eine **Ausnahme** bildet hier das bereits erwähnte (siehe 2.1) „RIAS“-System (Roter und Larson 2002). Exemplarisch konnten in einer Analyse von 167 Tonaufnahmen realer Arzt-Patienten-Gespräche aus der Pädiatrie anhand der Kriterien dieses Systems entmutigende Äußerungen der Ärzte in Bezug auf psychosoziale Themen identifiziert werden (Wissow et al. 2005). Hierbei zeigten sich diese Ablehnungen in 77 Prozent der Fälle für das spezifische Thema der Kindesmisshandlung und zu 34 Prozent für andere, themenverwandte Bereiche.

Aufgrund des **hohen Aufwandes** konnte dieses System hier jedoch nicht angewandt werden (siehe 2.1)

Die in der vorliegenden Studie angewendete AGBS weist laut den Autoren der Veröffentlichungsarbeit ebenfalls im Bereich der **Erfassung von Emotionen Schwächen** auf. Der thematisch zugehörige „Themenblock E“ weist allerdings auch eine Inhomogenität in der Intention der Fragen auf. Dies spiegelt sich auch in der etwas vagen Benennung des Bereiches mit „Allgemeine Überlegungen“ wider. So

zielen im engeren Sinne die Fragen Nummer 16 und 17 und im weiteren Sinne Die Fragen 16 bis 20 auf Emotionen während des Gespräches ab (siehe Abbildung 12).

Abbildung 12

Auszug aus dem Manual der AufklärungsGesprächsBewertungsSkala (Kurzfassung)

Themenblock E (Allgemeine Überlegungen)	16	Wie viele der Schlüssel-Bereiche (Behandlung, Prognose, Gefühle & Emotionen, Familie & Partnerschaft, Auswirkungen auf soziale Stellung) wurden besprochen?
	17	Wurden psychosoziale Aspekte (siehe Abbildung 16), die der Patient während des Gespräches verbal signalisierte, erörtert?
	18	Hat der Arzt es geschafft, dem Patienten während des Gespräches hilfreich zur Seite zu stehen?
	19	Hat der Arzt sich während des Gesprächs einer angemessenen Körpersprache bedient?
	20	Hat es der Arzt während des Gespräches vermieden unbeholfen zu wirken?
	21	Hat der Arzt – bezogen auf das ganze Gespräch – ein angemessenes Gesprächstempo eingehalten?
	22	Hat der Arzt das Gespräch rechtzeitig und angemessen beendet?
	16	Wie viele der Schlüssel-Bereiche (Behandlung, Prognose, Gefühle & Emotionen, Familie & Partnerschaft, Auswirkungen auf soziale Stellung) wurden besprochen?

Weiterhin ergibt sich für diesen „Themenblock E“ in der Reliabilitätsprüfung mittels Intraklassenkorrelation ein Wert von 0,76. Dies liegt in der Mitte des insgesamt guten Ergebnisspektrums dieser Variablen (Wand et al. 2008). Für die Validitätsprüfung mittels Cronbach's Alpha liegen von den Verfassern der AGBS keine Daten für die einzelnen Themenblöcke vor (siehe 6.3.2).

In der vorliegenden Untersuchung ergibt sich zwar aufgrund der Ergebnisse der Interraterreliabilität für den Bereich der **Erfassung von Emotionen und psychosozialen Themen** („Themenblock E“) analog der Originalarbeit (Wand et al. 2008) im Pre- und Re-Test die deutlich schwächste Aussagekraft des Instrumentes mit einem ICC von 0,76, beziehungsweise 0,21 und 0,50. In der Hauptauswertung war dieser Zusammenhang mittels der Berechnung des ICC jedoch nicht nachvollziehbar, denn es ließ sich hier ein ICC von 0,90 bestimmen (siehe 5.3 und 6.3).

Dies wird auf die noch etwas mangelbehaftete Strukturierung der Items in die fünf Themenbereiche zurückgeführt.

Die insgesamt besseren aber nicht ganz ausgewogenen Item-Skala-Korrelationen (siehe 4.3.1) als Maß der **Trennschärfe** in der Hauptauswertung im Vergleich zur Originalarbeit (Wand et al. 2008) sind auf die unterschiedlichen Settings in den Gesprächssituationen zurückzuführen (siehe 6.2).

Tabelle 24
Trennschärfe in den Themenblöcken für die Stichprobe 1 „München“
vergleichsweise zur Erstveröffentlichung „Erlangen“ (Wand et al. 2008)

	Korr. Item-Skala Korrelation		Videos	
	Erlangen	München	Erlangen	München
Gesamtbewertung AGBS				
Rater 1	0,77	0,99	58	40
Rater 2	0,73	1,00	58	40
Rater 3	0,78		58	
Themenblock A – Gestaltung der Szene				
Rater 1	0,89	0,11	63	40
Rater 2	0,93	0,00	63	40
Rater 3	0,93		63	
Themenblock B – Überbringen der Nachricht				
Rater 1	0,79	0,68	72	40
Rater 2	0,57	0,74	72	40
Rater 3	0,69		72	
Themenblock C – Bedenken des Patienten herausfinden				
Rater 1	0,52	0,81	72	40
Rater 2	0,39	0,84	72	40
Rater 3	0,44		72	
Themenblock D – Informationsübermittlung				
Rater 1	0,52	0,80	74	40
Rater 2	0,46	0,84	74	40
Rater 3	0,35		74	
Themenblock E – Allgemeine Überlegungen				
Rater 1	0,66	0,84	72	40
Rater 2	0,63	0,88	72	40
Rater 3	0,62		72	

Um also eine **Weiterentwicklung der AGBS** zu erreichen, und das Manual beziehungsweise die Appendizes (siehe Abbildungen 1 und 15) für die Erfassung der oben erwähnten Bereiche zu optimieren, bieten sich **Änderungen und Ergänzungen** an.

Zum einen sind die Inhalte der Items 21 und 22 mit der Frage nach dem Tempo und der Beendigung des Gespräches im Gegensatz zum restlichen „Themenblock E“ („Allgemeine Überlegungen“ siehe Abbildung 1) leichter objektivierbar. Sie sind daher eher als Ja/Nein-Fragen zu betrachten. Die beiden Items liegen demnach auch in

allen Bewertungsdurchläufen der vorliegenden Arbeit durchweg sehr nahe an den Extremwerten des Bewertungsspektrums eines Raters oder bilden diese meist selbst (siehe Abbildungen 3 bis 8). Sie sind ferner eher zur verbalen Gesprächstechnik zu zählen und passen daher in die „Themenblöcke A“ oder „B“.

Die verbleibenden Fragen des „Themenblockes E“ würden dann einen neuen Bereich „Psychosoziales, Emotionen und Nonverbalia“ ergeben. Hierfür ließen sich die Items 16 und 17 mit den Fragen nach psychosozialen und sogenannten Schlüsselbereichen durch eine Anpassung der Formulierungen und Appendizes für andere Gesprächssituationen adaptieren. Hierzu kann auch die Anweisung für eine standardisierte Raterschulung beitragen. Durch diese Adaptationen für andere Fachbereiche der Medizin kann eine noch bessere und einfachere Anwendbarkeit erreicht werden.

6.5 Nonverbale Kommunikation

Die nicht sprachlich getragene Interaktion als besonderer, wichtiger Teil der Kommunikation (siehe 1) wird durch die identifizierten Gesprächsbewertungssysteme (siehe 2.1) nur unzureichend erfasst (siehe 2.2).

In einer Studie wurden **Videos** von 89 OSCEs („objective structured clinical examination“) mit standardisierten Patienten für Medizinstudenten auf ihre nonverbalen Signale hin untersucht (Ishikawa et al. 2006). Diese Ergebnisse wurden mit denen der **Evaluation der Schauspielpatienten** im Hinblick auf die Qualität der Gesprächsinhalte verglichen. Die Patienten tendierten dazu, bessere Bewertungen zu geben, wenn die Studenten sie direkt ansahen, beim Zuhören ermunternd nickten und sie beim Sprechen und Zuhören gleichermaßen ansahen. Somit wurde durch diese Untersuchung die Bedeutung nonverbaler Kommunikation für den Behandlungserfolg durch Ärzte belegt.

In das Konzept der Gesprächsbewertung durch die AGBS ließe sich die Beurteilung nonverbaler Kommunikation integrieren, indem man entweder die Frage 19 (siehe Abbildung 12) umformulierte, beziehungsweise um eine genaue Beschreibung der Körpersprache ergänzte. Oder man fügt die detaillierte Beschreibung in die Appendizes ein (siehe 6.5). Hierbei sollten Instrumente zur Beurteilung nichtsprachlicher Auseinandersetzung wie das „Relational Communication Scale for Observational Measurement“, „RCS-O“, (Gallagher et al. 2004) zumindest als Ergänzungen berücksichtigt werden.

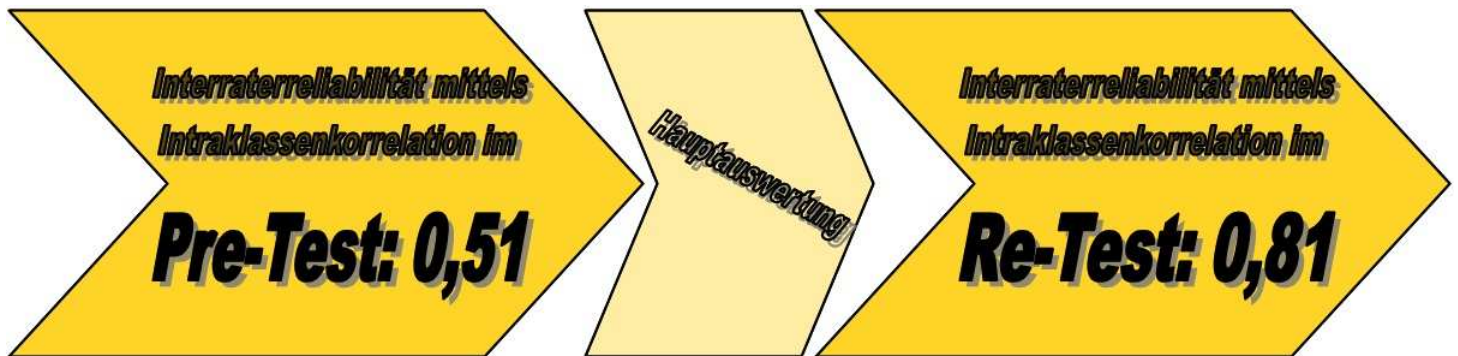
6.6 Standardisierte Raterschulung

Im Verlauf der Arbeit wurde eine Vorgehensweise für eine mögliche standardisierte Schulung der Rater als zusätzliche Perspektive zur Weiterentwicklung diskutiert. Bislang wird in den meisten dieser Arbeit zugrundeliegenden Publikationen wenig detailliert auf das Ratertraining eingegangen, wenngleich auf dessen Stellenwert verwiesen wird. Durch die Abfolge von Pre-Test, Hauptauswertung und Re-Test, lässt sich jedoch eine genaue Aussage zur Stabilität der Ratings über einen längeren Zeitraum treffen.

So zeigt die Güte für den Pre-Test (siehe 6.3.1), genau wie die Expertenmeinung (siehe 6.7), dass bereits **geringer Aufwand** für das Ratertraining **gute Ergebnisse** hervorbringen kann.

Zusätzlich entspricht aber die Verbesserung d. Interraterreliabilität von Pre- zu Re-Test auf ein ausgezeichnetes Niveau auch dem **Erfolg der weiteren Raterschulung** (siehe Abbildung 13). Diese bestand aber im Wesentlichen aus der Beschäftigung mit dem Rating des Pre-Tests selbst und der Auswertung weiterer Videoaufzeichnungen mittels AGBS.

Abbildung 13
ICC im Verlauf



Die Hauptauswertung und der Re-Test erfolgten dann jeweils mit abgeschlossenem Ratertraining. Das gleichbleibend hohe Level der Ergebnisse für die Interraterreliabilität, vor allem dieser beiden Auswertungen im Vergleich, zeigt eine **Konstanz** der Bewertung durch die AGBS **über einen langen Zeitraum** (siehe Abbildung 14). Dies belegt, dass ein etwa **10 Stunden** umfassendes Training auf jeden Fall ausreicht.

Abbildung 14
ICC im Verlauf



Aufgrund dieser Erfahrungen empfiehlt sich das Vorgehen in der **Anwendung der AGBS in der Zukunft**.

Um für eine **einfach und schnell** durchzuführende, aber möglichst standardisierte Raterschulung zu sorgen, wird eine zusätzliche Umarbeitung der Appendizes in eigene **Manuale zum Selbststudium** vorangetrieben. Hierauf sollte dann jeweils

noch eine **kurze Einheit mit Videos** zum probeweisen Rating folgen. Hierzu wäre die Erstellung eines interaktiven Lehrfilmes denkbar, wie es bereits ähnlich in einem anderen Zusammenhang realisiert werden konnte (Frank et al. 2010), was im Folgenden noch weitere Erwähnung finden wird. Es wird sich außerdem zeigen, dass sich hierdurch das Ratertraining auf einen Umfang von etwa **vier bis sechs Stunden** an zwei bis drei Tagen optimieren ließe.

6.7 Weiterentwicklung und Anwendung

Im Rahmen von weiteren Dissertationen wurden bereits **modifizierte Versionen**, die „mAGBS“ verfasst und validiert. Dabei handelt es sich um Anpassungen des Instrumentes für Gesprächssituationen im kinder- und jugendmedizinischen Bereich. Die AGBS wurde für die Kommunikation mit Kindern und Jugendlichen (Englmeier 2010) zum einen, sowie mit den Eltern (Braun 2010) zum anderen, adaptiert und evaluiert.

Anhand der thematischen Gliederung des Manuals der AGBS wurde ein **Evaluationsbogen für einen Lehrfilm über Gesprächsführung** erstellt (Frank et al. 2010).

Bei einem Expertentreffen des Instituts für Medizindidaktik der LMU zum Thema Lehre am 23.11.2009 wurde der Lehrfilm „Ihr Kind hatte einen Unfall“ (Frank 2010) vorgestellt und die Evaluationsmöglichkeiten erörtert. Im Rahmen dieses Projektes, das die Einbindung des Videos in die elektronische Lernplattform der Fakultät umfasste, wurde der erarbeitete Fragenkatalog vorgestellt und diskutiert (siehe Abbildung 15). Die Struktur orientiert sich an den Items und der Anordnung der Themenblöcke der AGBS (siehe Abbildung 1). Wenig später konnte das Projekt online verfügbar gemacht werden, inklusive eines nochmals überarbeiteten Fragenkatalogs (Frank et al. 2010).

7 Zusammenfassung

7.1 Einleitung – Aktueller Forschungsstand

Kommunikationsfähigkeit ist für die Qualität der medizinischen Versorgung wichtig. Daher gibt es eine Reihe von Methoden, die **ärztliche Gesprächsleistung** zu erfassen.

Die identifizierten Auswertungsinstrumente für die Gesprächsauswertung im medizinischen Kontext ergaben eine gute Verwertbarkeit und Zuverlässigkeit. Die Ergebnisse waren dabei unabhängig von den Untersuchungsbedingungen und methodischen Unterschieden immer auf ähnlich hohem Niveau. Dies gilt auch für die Adaptation des „**Breaking Bad News Assessment Schedule**“, kurz **BAS** (Miller et al. 1999), sowie für die im Rahmen der vorliegenden Arbeit verwendete „**Aufklärungsgesprächbewertungsskala**“, kurz **AGBS** (Wand et al. 2008).

7.2 Zielsetzung – Wissenschaftlicher Beitrag dieser Arbeit

In der vorliegenden Arbeit wurde die **Gesprächsleistung von Ärzten** mittels der AGBS in Form einer **Replikationsstudie** untersucht.

Außerdem wurde der Frage nach der **Vergleichbarkeit** durch die Gegenüberstellung der Studienbedingungen und der Methoden in den zwei Veröffentlichungsarbeiten (Miller et al. 1999; Wand et al. 2008) sowie der vorliegenden Arbeit nachgegangen.

Zusätzlich zum primären Ziel, die **Gütekriterien Interrater-Reliabilität** und **Validität** zu überprüfen, wurde das Studiendesign um eine **Re-Test-Reliabilität** erweitert. Neben der Bedeutung für die Testgüte sollte so auch die Frage nach der **Stabilität der Raterleistung** und dem **Umfang der Raterschulung** beantwortet werden.

Weiterhin sollte die Erfassung von Emotionen überprüft werden.

Durch die Erkenntnisse der vorliegenden Arbeit wurden **Anwendungs- sowie Fortentwicklungsmöglichkeiten** des Instrumentes untersucht.

7.3 Methodik

Die AGBS umfasst **22 Items**, die anhand einer **Likertskala von „eins“ bis „fünf“** bewertet werden.

Es wurden insgesamt **50 Videoaufzeichnungen** von Medizinstudenten und Ärzten (Stichprobe 1, n=40 / Stichprobe 2, n=10) mit Rollenspielen von Gesprächssituationen eines Arztes mit einem Elternpaar untersucht. Die Aufnahmen aus dem Kontext **„Überbringen schlechter Nachrichten“** wurden mittels AGBS von **zwei Ratern** unabhängig ausgewertet. Dies erfolgte nach einem **10 stündigen Ratertraining**.

Es wurde die **Testvalidität** mit Hilfe von Cronbach's Alpha und die **Interraterreliabilität** durch den Intraklassenkoeffizienten (ICC) bestimmt.

Ferner wurde auch die **Re-Test- oder Re-Rate-Reliabilität** berechnet.

7.4 Ergebnisse

Der **Mittelwert der Gesamtbewertungen** der Gespräche durch die beiden Rater in Pre-Test, Re-Test (Stichprobe 2, n=10) und Hauptauswertung (Stichprobe 1, n=40) lag auf ähnlichem Niveau zwischen 2,23 und 2,88. Dabei entspricht 1,0 der besten und 5,0 der schlechtesten Bewertung. Die Gesamtbewertungen durch den Rater 2 waren hierbei immer besser als durch den Rater 1. Die geringsten Differenzen zwischen den mittleren Bewertungen der einzelnen Rater fanden sich übereinstimmend für Pre- (0,03), Re-Test (0,01) und Hauptauswertung (0,02) im „Themenblock E“ („Allgemeine Überlegungen“). Die größten Differenzen fanden sich

in „Themenblock A“ („Gesprächsbeginn/-umgebung“) ebenfalls übereinstimmend für Pre-Test (1,17), Re-Test (0,44) und Hauptauswertung (0,26).

Die interne Konsistenz als Maß für die **Testvalidität** für die mittlere Gesamtbewertung auf der AGBS betrug unter Verwendung von **Cronbach's Alpha** 0,89 für den Pre-Test, 0,84 für den Re-Test und **0,93** für die Hauptauswertung.

Der Intraklassen-Koeffizient zur Bestimmung der **Interrater-Reliabilität** nahm für die Gesamtbewertung in Pre-Test, Re-Test (Stichprobe 2) und Hauptauswertung (Stichprobe 1) einen Wert von 0,51, 0,81 und 0,93 ein (Minimum = 0, Maximum = 1).

7.5 Diskussion

Innerhalb des gleichen thematischen Kontexts **unterschied** sich die in der vorliegenden Arbeit untersuchte **Stichprobe** von der in der Originalarbeit (Wand et al. 2008) verwendeten. Es wurden auch **Ärzte mit Berufserfahrung** miteinbezogen, wie es bereits in der Veröffentlichungsarbeit des „BAS“ (Miller et al. 1999) praktiziert wurde.

Weiterhin handelte es sich im Unterschied zu den zugrundeliegenden Veröffentlichungen mit nur einem Kommunikationspartner des Arztes um Situationen mit **mehreren Gesprächsteilnehmern**.

Diese höhere **Varianz in der Stichprobe** ermöglichte einen Vergleich der Ergebnisse für einen größeren Anwendungsbereich, ebenso wie für verschiedene Trainingssituationen.

Alle bisher erfolgten Anwendungen, von der Entwicklung des BAS (Miller et al. 1999) über die Adaptation als AGBS (Wand et al. 2008) bis hin zur vorliegenden Replikationsstudie, konnten zeigen, dass es sich um **ein sehr gutes Instrument** handelt.

So ergab sich über alle Überprüfungsstufen eine durchweg sehr gute Validität mittels Cronbach's α von deutlich über 0,8 und eine stabile Reliabilität mittels ICC von über 0,8.

Die Analyse der Gütekriterien der AGBS ergab analog zur Originalarbeit (Wand et al. 2008), dass die verschiedenen Bereiche der kommunikativen Kompetenzen mithilfe der verwendeten Items unterschiedlich gut erfasst werden können. Ferner konnte gezeigt werden, dass Interraterreliabilität Schulung erfordert und die **Re-Rate- oder Intraraterreliabilität** mit dem durchgeführten Training **über 6 Monate konstant** war.

Diese Unterschiede deuteten daraufhin, dass mithilfe des Instrumentes bislang nicht alle der für die Arzt-Patienten-Kommunikation notwendigen Kompetenzen erfasst werden, so dass **Weiterentwicklungen des Instrumentes** sinnvoll erscheinen.

Ein Praxistest der AGBS konnte im Rahmen studentischer Fortbildung (Frank et al. 2010) durchgeführt werden.

Weiterhin wurden bereits **modifizierte Versionen**, die „mAGBS“ **verfasst und validiert** (Braun 2010; Englmeier 2010).

8 Literaturverzeichnis

- Back A. L., Arnold R. M., Baile W. F., Fryer-Edwards K. A., Alexander S. C., Barley G. E., Gooley T. A. und Tulskey J. A. (2007). Efficacy of Communication Skills Training for Giving Bad News and Discussing Transitions to Palliative Care. *Archives of Internal Medicine*, 167, 453-460.
- Baile W. F., Buckman R., Lenzi R., Glober G., Beale E. A. und Kudelka A. P. (2000). Spikes - A Six-Step Protocol for Delivering Bad News: Application to the Patient with Cancer. *The Oncologist*, 5, 302-311.
- Braun R. (2010). Die modifizierte Aufklärungsgesprächsbewertungsskala für Arzt-Eltern Gespräche. Dissertation, Medizinische Fakultät der Ludwig-Maximilians-Universität München.
- Buckman R. (1996). Talking to patients about cancer. *British Medical Journal*, 313, 699-700.
- Buckman R. (2002). Communications and emotions. *British Medical Journal*, 325, 672.
- BundesministeriumFürGesundheit (2002). Approbationsordnung für Ärzte. Berlin, Bundesministerium für Justiz.
- Cegala D. J. und Lenzmeier Broz S. (2002). Physician communication skills training: a review of theoretical backgrounds, objectives and skills. *MEDICAL EDUCATION*, 36, 1004-1016.
- Egger J. W. (2005). Das Biopsychosoziale Krankheitsmodell. *Psychologische Medizin*, 2, 3-12.
- Engel G. L. (1977). The Need for a New Medical Model: A Challenge for Biomedicine. *Science*, 196, 129-136.
- Englmeier J. (2010). Die modifizierte Aufklärungsgesprächsbewertungsskala für Arzt-Jugendlichen Gespräche. Dissertation, Medizinische Fakultät der Ludwig-Maximilians-Universität München.
- Fallowfield L. (1993). Giving sad and bad news. *THE LANCET*, 341, 476-478.
- Fallowfield L., Jenkins V., Farewell V., Saul J., Duffy A. und Eves R. (2002). Efficacy of a Cancer Research UK communication skills training model for oncologists: a randomised controlled trial. *THE LANCET*, 360, 650-656.
- Finsterer I. (2008). Behandlungs- Zufriedenheit von Patienten einer Kinder- und Jugendpsychiatrischen Poliklinik. Dissertation, Medizinische Fakultät der Ludwig-Maximilians-Universität München.
- Ford S., Fallowfield L. und Lewis S. (1982). Doctor-patient interactions in oncology. *Social Science & Medicine*, 42, 1511-1519.
- Ford S., Hall A., Ratcliffe D. und Fallowfield L. (2000). The Medical Interaction Process System (MIPS): an instrument for analysing interviews of oncologists and patients with cancer. *Social Science & Medicine*, 50, 553-566.
- Frank R. (2010). Präsentation Lehrfilm für Medizinstudenten „Ihr Kind hatte einen Unfall“ München, Ludwig-Maximilians-Universität München.
- Frank R., Finsterer I., Schillert M., Klemz S. und Früstück C. (2010). Lehrfilm "Ihr Kind hatte einen Unfall" & Evaluation auf www.lmudle.de. München, Moodle - Lernplattform der Medizinischen Fakultät - **ME**dizinisches **CU**rriculum - LMU **M**ünchen.
- Gallagher T. J., Hartung P. J., Gerzina H., Gregory Jr. S. W. und Merolla D. (2004). Further analysis of a doctor-patient nonverbal communication instrument. *Patient Education and Counseling*, 57, 262-271.

- Heath C., Luff P. und Sanchez Svensson M. (2007). Video and qualitative research: analysing medical practice and interaction. *MEDICAL EDUCATION*, 41, 109-116.
- Hobma S., Ram P. und Muijtjens A. (2006). Effective improvement of doctor-patient communication: a randomised controlled trial. *British Journal of General Practice*, 56, 580-586.
- Hulsman R. L., Ros W. J., Winnubst J. A. und Bensing J. M. (1999). Teaching clinically experienced physicians communication skills. A review of evaluation studies. *MEDICAL EDUCATION*, 33, 655-668.
- Ishikawa H., Hashimoto H., Kinoshita M., Fujimori S., Shimizu T. und Yano E. (2006). Evaluating medical students' non-verbal communication during the objective structured clinical examination. *MEDICAL EDUCATION*, 40, 1180-1187.
- Jank W. und Meyer H. (1991). *Didaktische Modelle*. Frankfurt: Cornelsen.
- Jenkins V. und Fallowfield L. (2002). Can Communication Skills Training Alter Physicians' Beliefs and Behavior Clinics? *Journal of Clinical Oncology*, 20, 3, 765-769.
- Jenkins V., Fallowfield L. und Saul J. (2001). Information needs of patients with cancer: results from a large study in UK cancer centres. *British Journal of Cancer*, 84, 48-45.
- Kopecky-Wenzel M., Maier E. M., Muntau A. C. und Frank R. (2007). Wie sage ich es den Eltern?! „Überbringen schlechter Nachrichten“ im Medizinischen Curriculum München, MeCuM LMU. *Hauner Journal (Zeitschrift des Dr. v. Haunerschen Kinderspitals der Ludwig-Maximilians-Universität München)*, 2-2007,
- Kopecky-Wenzel M., Maier E. M., Muntau A. C., Reinhardt D. und Frank R. (2009). Überbringen schlechter Nachrichten - videogestützte Trainingseinheit für Medizinstudenten. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 37, 139-144.
- Kurtz S., Silverman J., Benson J. und Draper J. (2003). Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. *Academic medicine : journal of the Association of American Medical Colleges*, 78, 802-809.
- Langewitz W. A., Eich P., Kiss A. und Wössmer B. (1998). Improving Communication Skills - A Randomized Controlled Behaviorally Oriented Intervention Study for Residents in Internal Medicine. *Psychosomatic Medicine*, 60, 268-276.
- Madder H. (1997). Existential autonomy: why patients should make their own choices. *Journal of medical ethics*, 23, 221-225.
- Medizinische Fakultät der LMU München S. (2009). *MeCuM - Medizinisches Curriculum München*. München,
- Meredith C., Symonds P., Webster L., Lamont D., Pyper E., Gillis C. R. und Fallowfield L. (1996). Information needs of cancer patients in West Scotland: cross sectional survey of patients' views. *BMC Medical Education*, 313, 724-726.
- Miller S. J., Hope T. und Talbot D. C. (1999). The development of a structured rating schedule (the BAS) to assess skills in breaking bad news. *British Journal of Cancer*, 80, 792-800.
- Muntau A. C. und Maier E. M. (2007). Die Bedeutung des Hochschullehrertrainings auf Frauenchiemsee für das Dr. von Haunersche Kinderspital und für die Lehre in der Pädiatrie. *Hauner Journal (Zeitschrift des Dr. v. Haunerschen Kinderspitals der Ludwig-Maximilians-Universität München)*, 01-07,

- Novack D. H., Plumer R., Smith R. L., Ochitill H., G.R. M. und Bennett J. M. (1979). Changes in physicians' attitudes toward telling the cancer patient. *The Journal of the American Medical Association*, 241, 897-900.
- Oken D. (1961). What to tell cancer patients. *The Journal of the American Medical Association*, 175, 1120-1128.
- Ramirez A. J., Graham J., Richards M. A., Cull A. und Gregory W. M. (1996). Mental health of hospital consultants: the effects of stress and satisfaction at work. *THE LANCET*, 347, 724-728.
- Roter D. und Larson S. (2002). The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. *Patient Education and Counselling*, 46, 243-251.
- Satterfield J. M. und Hughes E. (2007). Emotion skills training for medical students: a systematic review. *MEDICAL EDUCATION*, 41, 935-941.
- Schildmann J., Härlein J., Buchardi N., Schlögl M. und Vollmann J. (2006). Breaking bad news: evaluation study on self-perceived competences and views of medical and nursing students taking part in a collaborative workshop. *Support Care Center*, 14, 1157-1161.
- Schmidt-Mast M., Kindlimann A. und Langewitz W. A. (2005). Recipients' perspective on breaking bad news: How you put it really makes a difference. *Patient Education and Counselling*, 58, 244-251.
- Simmenroth-Nayda A., Nolte C., Fischer T. und Himmel W. (2008). *Der Calgary Cambridge Observation Guide: Validierung einer ins Deutsche übersetzten Version Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA)*, 2008,
- Thiel v., J., Ram P. und Dalen v., J. (2000). "MAAS-Global 2000". Maastricht University, Netherlands,
- Thierer T., Handelman S. und Black P. (2001). Effects of perceived patient attributes on dentist communication behavior. *Special care in dentistry*, 21, 1, 21-26.
- Wand S., Schildmann J., Buchardi N. und Vollmann J. (2008). Die Aufklärungsgesprächbewertungsskala (AGBS): Ein Instrument zur Bewertung kommunikativer Kompetenzen bei der Aufklärung von Patienten über Tumorerkrankungen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 101, 645-651.
- Weber M., Bohler E. und Kohler E. (2003). Kann Kommunikation mit unheilbar kranken Patienten gelehrt werden? Evaluation eines Kursmodells. *Medizinische Klinik*, 98, 477-483.
- Wissow L. S., Larson S., Anderson J. und Hadjiisky (2005). Pediatric Residents' Responses That Discourage Discussion of Psychosocial Problems in Primary Care. *PEDIATRICS*, 115, 1569-1578.

9 **Abbildungsverzeichnis**

Abbildung 1	Manual der <u>A</u> ufklärungs <u>G</u> esprächs <u>B</u> ewertungs <u>S</u> kala – <i>Kurzfassung</i>	(5)
Abbildung 2	Beispiel psychosozialer Themen	(12)
Abbildung 3	Mittlere Bewertung der Einzelfragen durch RATER 1 - <i>Stichprobe 1</i>	(36)
Abbildung 4	Mittlere Bewertung der Einzelfragen durch RATER 2 - <i>Stichprobe 1</i>	(36)
Abbildung 5	Mittlere Bewertung der Einzelfragen durch RATER 1 - <i>Pre-Test Stichprobe 2</i>	(37)
Abbildung 6	Mittlere Bewertung der Einzelfragen durch RATER 2 - <i>Pre-Test Stichprobe 2</i>	(37)
Abbildung 7	Mittlere Bewertung der Einzelfragen durch RATER 1 - <i>Re-Test Stichprobe 2</i>	(38)
Abbildung 8	Mittlere Bewertung der Einzelfragen durch RATER 2 - <i>Re-Test Stichprobe 2</i>	(38)
Abbildung 9	Studienbedingungen in der Untersuchungsreihe zur Gesprächsbewertung	(53)
Abbildung 10	Testvalidität im Verlauf aller Untersuchungsreihen	(62)
Abbildung 11	Abschätzung von Trainingseffekten (ICC im Verlauf)	(63)
Abbildung 12	Auszug Manual der AufklärungsGesprächsBewertungsSkala	(66)
Abbildung 13	ICC im Verlauf (Pre-/Re-Test)	(70)
Abbildung 14	ICC im Verlauf (Hauptauswertung/Re-Test)	(70)
Abbildung 15	Prüfungsfragen zum Lehrfilm „Ihr Kind hatte einen Unfall“	(82-84)
Abbildung 16	Appendizes zur AGBS	(83-87)

10 Tabellenverzeichnis

Tabelle 1	Bewertung der 22 Einzelfragen der AGBS für die Stichprobe 1 (n=40)	(34)
Tabelle 2	Mittlere Bewertung der Themenblöcke beider Rater für die Stichprobe 1 (n=40)	(39)
Tabelle 3	Mittlere Bewertung der Themenblöcke beider Rater für die Stichprobe 2 (n=10)	(40)
Tabelle 4	Mittlere Bewertung der Themenblöcke beider Rater und Stichproben	(41)
Tabelle 5	Testvalidität für die Stichprobe 1 (n=40)	(43)
Tabelle 6	Trennschärfe in den Themenblöcken für die Stichprobe 1 (n=40)	(43)
Tabelle 7	Interraterreliabilität, Inhomogenität, Interaktion für die Stichprobe 1 (n=40)	(44)
Tabelle 8	Varianz der Leistung, Inhomogenität, Interaktion für d. Stichprobe 1 (n=40)	(45)
Tabelle 9	Testvalidität für die Stichprobe 2 (n=10)	(46)
Tabelle 10	Validität des Pre-/Re-Tests für die Stichprobe 2 (n=10)	(46)
Tabelle 11	Interraterreliabilität in den Themenblöcken für die Stichprobe 2 (n=10)	(47)
Tabelle 12	Re-Testreliabilität in den Themenblöcken für die Stichprobe 2 (n=10)	(48)
Tabelle 13	Testvalidität für drei Durchläufe und zwei Stichproben - <i>Gesamtbewertung</i>	(49)
Tabelle 14	Interraterreliabilität für drei Durchläufe und zwei Stichproben - <i>Themenblöcke</i>	(50)
Tabelle 15	Vergleich der Bewertung Pre-Test mit Erstveröffentlichung AGBS	(55)
Tabelle 16	Vergleich der Testvalidität Pre-Test mit Erstveröffentlichung AGBS	(55)
Tabelle 17	Vergleich d. Interraterreliabilität für d. Pre-Test mit Erstveröffentlichung AGBS	(55)
Tabelle 18	Transformation Score (<i>BAS</i>) \approx Noten (<i>AGBS</i>)	(57)
Tabelle 19	Vergleich der Gesamtbewertung für die Stichprobe 1 mit den Erstveröffentlichungen BAS und AGBS	(58)
Tabelle 20	Vergleich der Themenblöcke Stichprobe 1 mit Erstveröffentlichung AGBS	(58)
Tabelle 21	Vergleich der Varianz der Stichprobe 1 mit Erstveröffentlichung AGBS	(60)
Tabelle 22	Vergl. d. Testvalidität Stichprobe 1 mit Erstveröffentlichungen BAS und AGBS	(61)
Tabelle 23	Vergleich der Interraterreliabilität für den Pre-Test (Stichprobe 2) mit den Erstveröffentlichungen BAS und AGBS	(58)
Tabelle 24	Vergl. d.Trennschärfe für die Stichprobe 1 mit Erstveröffentlichung AGBS	(67)

11 Anhang

Abbildung 15

Prüfungsfragen zum Lehrfilm „Ihr Kind hatte einen Unfall“ (Frank 2010; Frank et al. 2010) / (Stand 23.11.2009)
(Wichtiges und die richtige Antwort sind *kursiv*)

1. Welche Reihenfolge des Vorgehens ist *am ehesten richtig*, sobald der Arzt den Eltern gegenübersteht:

- a) passenden Raum finden, begrüßen, Platz anbieten, setzen, Gespräch beginnen
 - b) passenden Raum finden, Platz anbieten, begrüßen, setzen, Gespräch beginnen
 - c) begrüßen, Gespräch beginnen, passenden Raum finden, Platz anbieten, setzen
 - d) *begrüßen, passenden Raum finden, Platz anbieten, setzen, Gespräch beginnen*
 - e) begrüßen, passenden Raum finden, Gespräch beginnen, Platz anbieten, setzen
-

2. Welche äußeren Umstände beeinflussen die Gesprächssituation zwischen den Eltern und dem Arzt besonders *positiv*?

- a) *das Gespräch findet in einem separater Raum statt*
 - b) der Dienstfunk piepst
 - c) vorbeigehende Menschen können mithören
 - d) das Gespräch findet auf dem Flur statt
 - e) der Arzt ist in Eile und nicht über die aktuelle Situation informiert
-

3. Welche äußeren Umstände beeinflussen die Gesprächssituation zwischen den Eltern und dem Arzt besonders *negativ*?

- a) das Gespräch findet in einem separater Raum statt
 - b) der Dienstfunk piepst
 - c) vorbeigehende Menschen können mithören
 - d) *das Gespräch findet auf dem Flur statt*
 - e) die moderne Möblierung lenkt ab
-

4. Worauf sollte der Arzt bei der Begrüßung *am meisten* achten?

- a) vorher Hände waschen
 - b) intensiver Blickkontakt mit allen Beteiligten
 - c) *Vorstellen mit Namen und Funktion*
 - d) Ein Kompliment an die Mutter, um ins Gespräch zu kommen
 - e) zuerst in einen separaten Raum führen
-

5. Worauf sollte der Arzt *besonders* achten, wenn das Gespräch *mit beiden* Eltern stattfindet?

- a) Blickkontakt mit beiden Eltern aufnehmen und halten
 - b) *Beide Eltern begrüßen*
 - c) Ich spreche nur mit dem Familienoberhaupt
 - d) Ich bitte einen Elternteil vor der Türe zu warten um mich völlig auf einen konzentrieren zu können
 - e) Ich wende mich hauptsächlich der Mutter zu, da sie das Kind am Besten kennt
-

6. Wie bewerten sie die folgenden Aussagen?

- A Der Arzt sollte sich *zu Beginn* des Gesprächs nach dem momentanen Befinden der Eltern erkundigen,...
(Verknüpfung) ...um...
B ...herauszufinden, ob der/die Gesprächspartner die intellektuelle Fähigkeit besitzen, sich mit einem Arzt zu unterhalten.

- a) A und B richtig, Verknüpfung richtig
 - b) *Nur A richtig*
 - c) Nur B richtig
 - d) A und B richtig, Verknüpfung falsch
 - e) A und B falsch
-

7. Wie bewerten sie die folgenden Aussagen?

- A Der Arzt sollte den Eltern zu Beginn des Gesprächs anbieten, das momentane Befinden zu schildern,...
(Verknüpfung) ...um...
B ... den Gesprächspartnern und ihrer Stimmungslage angemessen begegnen zu können und eine, nach Möglichkeit, verständnis- und vertrauensvolle Basis zu schaffen.

- a) *A und B richtig, Verknüpfung richtig*
 - b) Nur A richtig
 - c) Nur B richtig
 - d) A und B richtig, Verknüpfung falsch
 - e) A und B falsch
-

8. Was sollte der Arzt *am ehesten* tun/fragen, um eine Grundlage zur Überbringung der schlechten Nachricht, bzw. zur Informationsübermittlung, zu schaffen?

- a) Er sollte nach dem Kenntnisstand der Eltern, bezüglich des Unfallhergangs und der bislang veranlassten Maßnahmen erfragen
 - b) Er muss sich vergewissern, dass das verunfallte Kind krankenversichert ist
 - c) Er kann sich darauf verlassen, dass die Eltern durch die Rettungskräfte und das übrige Stationspersonal umfassend informiert wurden und erläutert in einfachen, kurzen Sätzen die intensivmedizinischen Maßnahmen
 - d) Er beschränkt sich ausschließlich auf die detaillierte Beschreibung des Unfallhergangs mit der Klärung der Schuldfrage bei dem Verkehrsunfall
 - e) Er konfrontiert die Eltern sofort mit der prognostischen Einschätzung, da er sich rechtlich absichern muss
-

9. Wie sollte der Arzt vorgehen, um die schlechte Nachricht zu überbringen?

- a) Er sollte einen „Warnschuss“ abgeben, um auf eine schlechte Nachricht vorzubereiten
 - b) Um die Eltern unter allen Umständen zu schonen, sollte es der Arzt vermeiden über die Erkrankung genauer zu informieren
 - c) Er sollte möglichst „kurz und schmerzlos“ auf den Punkt kommen und die Eltern mit den Fakten konfrontieren
 - d) Die Eltern haben kein Recht auf Information, so dass der Arzt völlig frei entscheiden kann, was er ihnen mitteilt
 - e) Der sogenannte „Warnschuss“, um die Eltern vorzubereiten, sollte unbedingt vermieden werden, da es zur unnötigen Verunsicherung führt
-

10. Wie bewerten sie die folgenden Aussagen?

- A Der Arzt verwendet beim Überbringen der schlechten Nachricht eine für die Eltern verständliche Sprache... (Verknüpfung) ...und...
- B ...überlässt es ihnen, wie detailliert auf Einzelheiten eingegangen wird.

- a) A und B richtig, Verknüpfung richtig
 - b) Nur A richtig
 - c) Nur B richtig
 - d) A und B richtig, Verknüpfung falsch
 - e) A und B falsch
-

11. Woran erkennt man im Allgemeinen *am ehesten*, dass der Arzt das Gespräch *gut* führt?

- a) Er hat den größten Redeanteil
 - b) Er spricht lauter als die Eltern
 - c) Er kündigt den Gesprächsverlauf an
 - d) Er lässt sich nicht unterbrechen und führt seinen Satz immer zu Ende
 - e) Er deckt alle Fehler schonungslos auf
-

12. Welches Element des Gespräches ist *am wichtigsten*?

- a) Tempo
 - b) akzentfreie Sprache
 - c) offene Fragen stellen
 - d) gleicher kultureller Hintergrund von Eltern und Arzt
 - e) Alter und Geschlecht der Gesprächsteilnehmer
-

13. Hat der Arzt während des Gespräches, insbesondere nach dem Überbringen der schlechten Nachricht eine *Pause* gemacht, bzw. sollte er das tun?

- a) Ja, möglichst lange Sprechpausen (<1 Min) sind das beste Mittel, um unangenehme Themen zu vermeiden
 - b) Nein, Sprechpausen kosten den Arzt nur unnötig Zeit und sollten deshalb grundsätzlich vermieden werden
 - c) Ja, Sprechpausen sind ein probates Mittel, um auf die gesetzlich vorgeschriebene Mindestgesprächsdauer kommen zu können
 - d) Nein, Sprechpausen vermitteln den Eltern das Gefühl eines inkompetenten Gesprächspartners und sollten daher möglichst vermieden werden
 - e) Ja, es wurden Sprechpausen von ca. 4 Sekunden gemacht, um den Eltern die Möglichkeit zu geben auf die Nachricht zu reagieren
-

14. Wie gelingt es dem Arzt, den Eltern Raum für Nachfragen zu gewähren?

- a) Der Arzt versichert sich, ob die Eltern alles verstanden haben
 - b) Gesprächspausen werden konsequent vermieden
 - c) Der Arzt lässt keinen Raum für Nachfragen seitens der Eltern
 - d) Der Arzt verlässt längere Zeit den Raum, um den Eltern Zeit zu geben, sich zu beratschlagen
 - e) Der Arzt fordert die Eltern auf, mit ihm gemeinsam eine detaillierte Liste der medizinischen Fachbegriffe und Prozeduren zu erörtern
-

15. Was macht die Eltern im ersten Gespräch *am ehesten* ärgerlich?

- a) Der Arzt nimmt die Sorgen der Eltern wahr
 - b) Der Arzt vermittelt den Eltern, dass er gerade keine Zeit hat
 - c) Der Arzt weiß anfangs nicht, welche Eltern gerade vor ihm stehen
 - d) Der Arzt lässt sich kurz unterbrechen
 - e) Der Arzt hat den Kittel nicht geschlossen
-

16. Welche nonverbalen Signale des Arztes vermitteln den Gesprächspartnern eine akzeptierende, Anteil nehmende, einfühlsame Haltung?

- a) lautes Lachen
 - b) *eine zugewandte Körperhaltung*
 - c) Blick schweifen lassen
 - d) Augenbrauen hochziehen
 - e) Kratzen am Hinterkopf
-

17. Welche Möglichkeiten gibt es, angemessen auf die Emotionen der Eltern zu reagieren?

- a) *direktes Ansprechen*
 - b) der Versuch die Eltern abzulenken
 - c) das Gespräch möglichst schnell beenden und den Raum verlassen
 - d) der Versuch die Emotionen zu übersehen
 - e) *eine möglichst lange Gesprächspause machen und dann nicht weiter darauf eingehen*
-

18. Welche sozialen Kompetenzen (empathische Fähigkeiten) sind für ein gutes Gespräch am meisten hilfreich?

- a) Fachliche Distanz
 - b) Intelligenz
 - c) (Welt-)Offenheit
 - d) Ehrgeiz
 - e) Einfühlungsvermögen
-

19. Welche Aussage(-n) ist/sind *am ehesten* richtig?

- 1) Ich konzentriere mich auf den Elternteil, der sich aktiver am Gespräch beteiligt
 - 2) Ich halte möglichst wenig Blickkontakt, damit sich keiner der Elternteile benachteiligt fühlt?
 - 3) Ich bemühe mich den traurigen Blick der Mutter zu übersehen, weil ich die fachliche Distanz wahren möchte.
 - 4) Ich verliere die Zeit nicht aus den Augen und verdeutliche dies den Eltern mit Blicken auf meine Armbanduhr, um meine weiteren Termine pünktlich wahrnehmen zu können.
 - 5) Ich bemühe mich um eine verständliche Sprache und ein angemessenes Tempo.
 - 6) Ich orientiere mich bei der Vermittlung von Details an den verbalen und nonverbalen Signalen der Eltern.
 - 7) Ich stelle mich zu Beginn des Gespräches mit Namen, Wohnort, Familienstand und Karriereabsichten vor.
 - 8) Ich mache nach jedem Satz eine Pause, um den Eltern die Möglichkeit zu geben meine Informationen aufzunehmen und zu überdenken.
- a) alle Aussagen sind richtig
 - b) 5, 6, 7 und 8 sind richtig
 - c) *5, 6 und 8 sind richtig*
 - d) 4, 5 und 8 sind richtig
 - e) alle Aussagen sind falsch

Abbildung 16

Appendizes zur AGBS (Wand et al. 2008)

(„Zusatzinformationen“)

Zusatz-Informationen **Zum Rollenspiel** **„Stimmklappen-Karzinom“**

Aufstellung von Sorgen/Bedenken des Patienten (10) + Ranking (11):

1. „Werde ich wieder gesund? Kann ich wieder normal sprechen? Muss ich daran sterben?“ (dies ist auch die wichtigste Sorge, die in 11 gemeint ist)
2. „Wie werden meine Frau/mein Mann, meine Kinder reagieren, wenn ich vielleicht nicht mehr mit Ihnen normal sprechen kann?“
3. „Wie soll das jetzt weitergehen mit meinem Job, wenn ich nicht wieder richtig sprechen kann, kann ich das vergessen!“
4. „Werde ich für die Behandlung lange im Krankenhaus bleiben müssen?“
5. „Was muss ich beachten während der Behandlung?“

Zu Frage 15: Infos, die der Arzt hat:

Herr/Frau Ingelmann ist 35 Jahre alt und arbeitet als Moderatorin beim Radio. Ihr/sein Hausarzt hat sie/ihn nach über Monate bestehender Heiserkeit trotz Stimm schonung zur Abklärung in die HNO-Klinik überwiesen. Der laryngoskopische Befund zeigt einem **ca. 1x2 cm großen Tumor der rechten Stimmklappe** mit eingeschränkter Stimmklappen-Beweglichkeit. **Höchstwahrscheinlich handelt es sich um ein Karzinom**, wofür auch spricht, dass Herr/ Frau Ingelmann seit 30 Jahren zwei Packungen Zigaretten pro Tag raucht. **Sonographisch sind keine vergrößerten LK im Halsbereich** nachweisbar. Sollte sich bei den weiteren Voruntersuchungen (Oberbauch-, Lungen-, Hals-CT,

Panendoskopie) kein Hinweis auf einem Zweittumor oder Metastasen finden, so sollte eine möglichst **funktionserhaltende Kehlkopfteilresektion mit Ausräumung der Halslymphknoten** (Neck dissection) rechts und anschließender Nachbestrahlung angestrebt werden.

Dauer des ersten stationären Aufenthalts: ca. ergänzen!

Prognose (vollständige Resektion des Tumors ohne Absiedlungen): **5-JÜR ca. 80%**.

Sollte sich der Tumor intraoperativ größer darstellen, als zunächst angenommen, muss **möglicherweise eine Totalresektion** vorgenommen werden, was den Verlust des Organs und die Notwendigkeit einer **prothetischen Versorgung** (Sprechkanüle o.ä.) zur Folge hätte. Aber **auch bei einer Teilresektion** ist mit einem **Verlust der normalen Sprache** zu rechnen, die aber innerhalb der folgenden Monate mit logopädischer Hilfe wiedererlernt werden kann, so dass der Patient im optimalen Fall seinen Beruf fortsetzen könnte. Eine **weitere Komplikation postoperativ** sind **narbige Schrumpfung**, die **Schluckbeschwerden** verursachen können.

Psychosoziale Aspekte, die vom Patient (falls nicht erfragt) signalisiert werden sollen (17, vergleiche auch Frage 10):

„Wie soll das jetzt weitergehen mit meinem Job, wenn ich nicht wieder richtig sprechen kann, kann ich das vergessen?“

„Wie werden mein/e Partner/in, meine Kinder reagieren, wenn ich vielleicht nicht mehr mit Ihnen normal sprechen kann?“

„Gibt es hier eine Psychologin/einen Seelsorger, mit der/dem ich sprechen kann?“

„Wie werden meine Freunde auf die Krankheit reagieren?“

„Was ist mit meinem Hund Kira, werde ich mich noch um sie kümmern können, wird sie mich noch verstehen?“

Zusatz-Informationen Zum Rollenspiel „Oligodendrogliom“

Aufstellung von Sorgen/Bedenken des Patienten (10) + Ranking (11):

1. „Werde ich wieder gesund? Kann ich wieder normal sprechen? Muss ich daran sterben?“ (dies ist auch die wichtigste Sorge, nie in 11 gemeint ist)
2. „Wie werden meine Frau/ mein Mann, meine Kinder reagieren, wenn ich vielleicht nicht mehr mit Ihnen normal sprechen kann?“
3. „Wie soll das jetzt weitergehen mit meinem Job, wenn ich nicht wieder richtig sprechen kann, kann ich das vergessen!“
4. „Werde ich für die Behandlung lange im Krankenhaus bleiben müssen?“
5. „Was muss ich beachten während der Behandlung?“

Zu Frage 15: Infos, die der Arzt hat:

Herr/Frau Martin ist 35 Jahre alt und Filialleiter/in einer Sparkasse. Er/sie leidet seit etwa 2 Monaten an Kopfschmerzen. Vor 10 Tagen erlitt Frau/Herr Martin einen zerebralen Krampfanfall und wurde zur weiteren Abklärung in die Neurologie des Klinikums eingewiesen. Es wurden ein EEG, ein CT, ein MRT und ein Angiogramm gemacht. Im CT zeigte sich links frontoparietal ein hypodenser Rundherd. Zwecks Abklärung der Entität wurde in der Abteilung für Neurochirurgie eine Probeexcision durchgeführt. Der Patient wurde nach dem Eingriff für einige Tage unter einer medikamentösen Anfallsprophylaxe entlassen bis das endgültige Histologieergebnis vorliegt.

Heute wurde Frau/Herr Martin zu einem Gespräch auf die neurochirurgische Station einbestellt, um die Befunde zu besprechen. Es handelt sich dem histologischen Bericht zu Folge um ein **Oligodendrogliom**, nach dem bereits durchgeführten Ganzkörper-CT gibt es **keinen Anhalt für eine Metastasierung**. **Therapeutisch** kommt nur die **transkranielle Tumorexstirpation** in Frage. Leider können Oligodendrogliome in einem Teil der Fälle nicht vollständig entfernt werden und **rezidivieren** daher **möglicherweise**.

Dauer des stationären Aufenthalts zunächst ca. noch ergänzen

Prognose (makroskopisch vollständige Resektion des Tumors): 5-JÜR: ca. 80%.

Es besteht ein gewisses **OP-Risiko** durch die möglicherweise großen **Blutverluste** und **Verletzung angrenzender Strukturen** (u.a. Broca'sches Sprachzentrum; **Risiko von motorischen Aphasien** (Sprechstörungen)). **Eine postoperative Chemotherapie ist nicht nötig**. Bei unkompliziertem Operations- und Wundheilungsverlauf ist **in etwa zwei Monaten an einen vorsichtigen Wiedereinstieg in den Beruf zu denken**.

Psychosoziale Aspekte, die vom Patient (falls nicht erfragt) signalisiert werden sollen (17):

„Wie soll das jetzt weitergehen mit meinem Job, wenn ich nicht wieder richtig sprechen kann, kann ich das vergessen?“

„Wie werden mein/e Partner/in, meine Kinder reagieren, wenn ich mich vielleicht durch die OP verändere?“

„Gibt es hier eine Psychologin/einen Seelsorger, mit der/dem ich sprechen kann?“

„Wie werden meine Freunde auf die Krankheit reagieren?“

„Was ist mit meinem Hund Kira, werde ich mich noch um sie kümmern können, wird sie mich noch verstehen?“

**Zusatz-Informationen
Zum Rollenspiel
„Non-Hodgkin-Lymphom“**

Aufstellung von Sorgen/Bedenken des Patienten (10) + Ranking (11):

1. „Werde ich wieder gesund oder muss ich daran sterben?“ (*dies ist auch die wichtigste Sorge, die in 11 gemeint ist*)
2. „Wie werden meine Frau/mein Mann, meine Kinder reagieren“
3. „Wie soll das jetzt weitergehen mit meinem Job, wenn ich solange aussetzen muss“
4. „Werde ich für die Behandlung lange im Krankenhaus bleiben müssen“
5. „Was muss ich beachten während der Behandlung“

Zu Frage 15: Infos, die der Arzt hat:

Herr/Frau Maier ist AbteilungsleiterIn in einem der örtlichen metallverarbeitenden Betriebe. Er/sie leidet seit etwa 2 Monaten an Bauchschmerzen, Verstopfung und außerdem seit zwei Wochen an Nachtschweiß, der den/die PatientIn zwingt, den Pyjama zu wechseln. Beim Besuch der/des PatientIn beim hausärztlich tätigen Internisten vor 10 Tagen erfolgte eine sonografische Untersuchung des Abdomens. Aufgrund der dort sichtbaren vergrößerten Lymphknoten und der klinischen Beschwerden des Patienten erfolgte die Vereinbarung eines stationären Aufnahmetermins.

Nach Einweisung in die Hämato/Onkologie des Klinikums wurde ein Ganzkörper-CT sowie zwei Tage später eine Punktion eines suspekt vergrößerten Lymphknotens im Bauchraum durchgeführt. Der/die PatientIn wurde nach dem Eingriff für einige Tage entlassen, bis das endgültige Histologieergebnis vorliegt.

Heute wurde Frau/Herr Maier zu einem Gespräch auf die hämato-onkologische Abteilung einbestellt, um die Befunde zu besprechen. Es handelt sich dem histologischen Bericht zu Folge um ein **hochmalignes diffus großzelliges NHL Stadium II** (zwei vergrößerte Lymphknotenregionen abdominell unterhalb des Zwerchfells). **Therapeutisch werden zunächst vier bis sechs Kurse eine kombinierte Immuno (Antikörper)-Chemotherapie durchgeführt.** In Abhängigkeit vom Befund in der Kontrolluntersuchung (Re-Staging) wird dann ggf. über weitere Therapieansätze entschieden. **Dauer des ersten stationären Aufenthalts ca. 4-6 Wochen**

Prognose: Die vorgeschlagene Therapie ist ein kurativer Behandlungsansatz, **5-JÜR: ca. 60-70%.**

Die kombinierte Immun-Chemotherapie birgt **unerwünschte Wirkungen und Risiken:** Kurzfristig kann es neben **Übelkeit und Erbrechen** zu Unverträglichkeiten bis hin zum **anaphylaktischen Schock** unter der Gabe der Immuntherapie (Antikörper gegen Tumorzellen) kommen. Mittel- und langfristig können die eingesetzten Chemotherapeutika **Schädigung** unter anderem an **Nerven und Herz** verursachen. Weiterhin besteht ein erhöhtes Risiko an sogenannten „**Zweitmalignomen**“ (bösartige Tumore, die durch die Chemotherapeutika ausgelöst werden) zu erkranken. **Bei unkompliziertem Behandlungsverlauf ist in etwa einem halben Jahr wieder an die Ausübung des Berufes zu denken.**

Psychosoziale Aspekte, die vom Patient (falls nicht erfragt) signalisiert werden sollen (17 vergleiche auch Frage 10):

- „Wie soll das jetzt weitergehen mit meinem Job?“
„Wie werden mein/e Partner/in, meine Kinder reagieren, wenn ich mich vielleicht durch die Chemotherapie verändere?“
„Gibt es hier eine Psychologin/einen Seelsorger, mit der/dem ich sprechen kann?“
„Wie werden meine Freunde auf die Krankheit reagieren?“
„Was ist mit Moritz, meinem Kater, ist er für mich gefährlich, wenn ich durch die Chemo geschwächt bin, muss er weg?“
-

**Zusatz-Informationen
Zum Rollenspiel
„Morbus Hodgkin“**

Aufstellung von Sorgen/Bedenken des Patienten (10) + Ranking (11):

1. „Werde ich wieder gesund oder muss ich daran sterben?“ (*dies ist auch die wichtigste Sorge, die in 11 gemeint ist*)
2. „Wie werden meine Frau/mein Mann, meine Kinder reagieren“
3. „Wie soll das jetzt weitergehen mit meinem Job, wenn ich solange aussetzen muss“
4. „Werde ich für die Behandlung lange im Krankenhaus bleiben müssen“
5. „Was muss ich beachten während der Behandlung“

Zu Frage 15: Infos, die der Arzt hat:

Herr/Frau Müller ist GeschäftsführerIn in einem örtlichen Schreibwarenhandel. Er/sie leidet seit etwa 3 Monaten an einem sich zunehmend vergrößerndem Lymphknoten am Hals (links) sowie Gewichtsverlust (9 kg in den letzten 6 Wochen). Beim Besuch der/des PatientIn beim hausärztlich tätigen Internisten vor 10 Tagen erfolgte eine sonografische Untersuchung des Halses und des Abdomens. Aufgrund der an beiden Orten vergrößerten Lymphknoten und der klinischen Beschwerden des Patienten erfolgte die Vereinbarung eines stationären Aufnahmetermins.

Nach Einweisung in die Hämato/Onkologie des Klinikums wurde ein Ganzkörper-CT

sowie zwei Tage später eine Punktion des Lymphknotens am Hals durchgeführt. Der/die PatientIn wurde nach dem Eingriff für einige Tage entlassen bis das endgültige Histologieergebnis vorliegt.

Heute wurde Frau/Herr Maier zu einem Gespräch auf die hämato-onkologischen Abteilung einbestellt, um die Befunde zu besprechen. Es handelt sich dem histologischen Bericht zu Folge, um **ein M. Hodgkin Stadium IIIB** (zwei vergrößerte Lymphknotenregionen abdominell unterhalb des Zwerchfells sowie ein Lymphknotenkonglomerat links cervical). **Therapeutisch** werden zunächst **sechs bis acht Kurse einer Chemotherapie** durchgeführt. In Abhängigkeit vom Befund in der Kontrolluntersuchung (Re-Staging) wird dann ggf. über weitere Therapieansätze entschieden.

Dauer des ersten stationären Aufenthalts ca. 4-6 Wochen

Prognose: Die vorgeschlagene Therapie ist ein kurativer Behandlungsansatz, **5-JÜR: ca. 65-75%**.

Die Chemotherapie birgt **unerwünschte Wirkungen und Risiken**: Kurz und Mittelfristig kann es neben **Übelkeit und Erbrechen** zur **Schwächung der Abwehr** kommen. In Abhängigkeit von der Entwicklung des Blutbildes kann die Gabe von Erythrozyten- und Thrombozytenkonzentraten notwendig sein. Mittel- und langfristig können die eingesetzten Chemotherapeutika **Schädigung unter anderem an Nerven und Herz** verursachen. Weiterhin besteht ein erhöhtes Risiko an sogenannten **„Zweitmalignomen“** (bösartige Tumore, die durch die Chemotherapeutika ausgelöst werden) zu erkranken. **Bei unkompliziertem Behandlungsverlauf ist in etwa einem halben Jahr wieder an die Ausübung des Berufes zu denken.**

Psychosoziale Aspekte, die vom Patient (falls nicht erfragt) signalisiert werden sollen (17):

„Wie soll das jetzt weitergehen mit meinem Job?“

„Wie werden mein/e Partner/in, meine Kinder reagieren, wenn ich mich vielleicht durch die Chemotherapie verändere?“

„Gibt es hier eine Psychologin/einen Seelsorger, mit der/dem ich sprechen kann?“

„Wie werden meine Freunde auf die Krankheit reagieren?“

„Was ist mit Moritz, meinem Kater, ist er für mich gefährlich, wenn ich durch die Chemo geschwächt bin, muss er weg?“

12 Danksagung

Mein herzlichster Dank gilt Herrn Professor Reiner Frank für die herausragende Betreuung und das großartige Engagement. Die stets aufmunternde, positive Art mit denen Herr Frank all seinen Kollegen und Doktoranden gleichermaßen begegnet, wird für mich auch über die Zusammenarbeit hinaus noch Vorbildfunktion haben.

Außerdem möchte ich mich bei Frau Sabine Klemz für ihre tatkräftige Unterstützung bei der Gewinnung der Datengrundlage und die allgemein gute Zusammenarbeit bedanken.

Dem gesamten Team der Kinder- und Jugendpsychiatrischen Ambulanz Lindwurmstraße bin ich schließlich für die allzeit unkomplizierte und freundliche Atmosphäre, sowie der konstruktiven Unterstützung meiner Arbeit sehr dankbar.

Schließlich gilt auch sämtlichen Mitwirkenden der Videoaufzeichnungen mein Dank, die einer wissenschaftlichen Verwendung der Aufnahmen zugestimmt haben.

13 Lebenslauf

Name: Clemens Frühstück

Geburtstag: 24.09.1979 in Starnberg

Staatsangehörigkeit: deutsch

Familienstand: verheiratet

Eltern: Michael Frühstück, Buchhalter
Elisabeth Frühstück, geb. Pasterny, Hausfrau

Geschwister: Almuth Andreae
Benedikt Frühstück
Agnes Onyemaeke

Ehefrau: Susanna Frühstück, geb. Breitenfelder, Erzieherin

Kinder: Leni Frühstück, Ava Frühstück

Schulbildung: 1986- 1990 Grundschule an der Guldeinstrasse,
München
1990- 1999 Wittelsbacher Gymnasium, München
Abiturfächer: Biologie, Physik, Deutsch, Ethik
Abiturnote: 2,5

Werdegang: Oktober 1999 bis März 2002 Tätigkeit bei Vogel-
Burda Communications, Redaktion CHIP und CHIP
online

September 2002 bis Juli 2003 Studium an der
Medizinischen Universität Innsbruck

September 2003 bis November 2010
Medizinstudium an der Ludwig-Maximilian-
Universität München

Seit April 2010 Anstellung als Assistenzarzt in der
Klinik und Poliklinik für Psychiatrie und
Psychotherapie der Ludwig-Maximilians-Universität
München

Fremdsprachenkenntnisse: Englisch und Französisch