Aus der Abteilung für Allgemein-, Viszeral- und Kinderchirurgie
(Prof. Dr. med. B. M. Ghadimi)
der Medizinischen Fakultät der Universität Göttingen

Standardisierte und qualitätsgesicherte Video-Prüfung: Aufklärungsgespräch vor der Operation

INAUGURAL – DISSERTATION
zur Erlangung des Doktorgrades
der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von

Christoph Kiehl

aus

Achim

Göttingen 2018

Dekan:	Prof. Dr. rer. nat. H. K. Kroemer
Referent/in:	Prof. Dr. med. S. König
Ko-Referent/in:	i.V. PD Dr. Philipp Kauffmann
Drittreferent/in:	…….....

Tag der mündlichen Prüfung: 24.07.2019

Hiermit erkläre ich, die Dissertation mit dem Titel „Standardisierte und qualitätsgesicherte Video-Prüfung: Aufklärungsgespräch vor der Operation“ eigenhändig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den 12.08.2018 …….

Inhaltsverzeichnis

Abbildungsverzeichnis III

Tabellenverzeichnis IV

Abkürzungsverzeichnis V

1. Einleitung 1

1.1 Kommunikation in der Arzt-Patientenbeziehung 1

1.2 Vermittlung der Kommunikationskompetenz und Auswahl der geeigneten Prüfungsform 3

1.3 Objective Structured Clinical Examination 5

1.4 Möglichkeiten und Vorteile des Einsatzes einer Videoaufzeichnung 8

2. Zielsetzung und Fragestellung 10

3. Methoden 11

3.1 Studiendesign und die Teilnehmer 11

3.2 Prüfungsgruppen und Szenarien 12

3.3 Ablauf der Prüfung 13

3.4 Bewertungsbögen (= Checklisten) 14

3.5 Rater 15

3.6 Statistische Auswertung 15

4. Ergebnisse 19

4.1 Deskriptive Statistik und Szenarienvergleich 19

4.1.1 Deskriptive Statistik über alle Szenarien 19

4.1.2 Deskriptive Statistik für das Szenario Appendektomie 22

4.1.3 Deskriptive Statistik für das Szenario Cholezystektomie 24

4.1.4 Deskriptive Statistik für das Szenario Leistenhernienverschluss 26

4.1.5 Szenarienvergleich 28

4.2 Checklistenevaluation 32

4.2.1 Interne Konsistenz (Cronbachs α) ... 32
4.2.2 Itemschwierigkeit ... 33
4.2.3 Trennschärfe ... 38
4.3 Interrater-Reliabilität ... 40
4.4 Kalkulation der Bewertungszeit ... 42
5. Diskussion ... 43
5.1 Etablierung eines reliablen und qualitativ hochwertigen Bewertungs-systems für die Rater 43
5.2 Vorzüge der Video-Aufzeichnung ... 46
5.3 Einschränkungen und Ausblick ... 48
5.4 Einordnung der Studie in den wissenschaftlichen Kontext ... 49
6. Schlussfolgerung ... 51
7. Zusammenfassung ... 52
7.1 Hintergrund ... 52
7.2 Methoden ... 52
7.3 Ergebnisse ... 52
7.4 Diskussion und Ausblick ... 53
8. Anhang ... 54
8.1 Checklisten ... 54
8.2 Einverständiserklärung ... 57
9. Literaturverzeichnis ... 58

Abbildungsverzeichnis

Abbildung I: Pyramide der Prüfungsstufen nach Miller in Anlehnung an das Original (Miller 1990) neu erarbeitet von C. Kiehl 4
Abbildung II: Histogramm der Gesamtbewertung 20
Abbildung III: Q-Q-Diagramm der Gesamtbewertung 21
Abbildung IV: Histogramm zum Szenario Appendektomie 23
Abbildung V: Histogramm zum Szenario Cholezystektomie 25
Abbildung VI: Histogramm zum Szenario Leistenhernienverschluss 27
Abbildung VII: Szenarienvergleich 31
Abbildung VIII: Bewertungsvergleich nach Rater und Szenario 41
Abbildung IX: Rater-Vergleich 42
Abbildung X: Checkliste „Akute Appendizitis“ 54
Abbildung XI: Checkliste „Symptomatische Cholezystolithiasis" 55
Abbildung XII: Checkliste „Leistenhernie" 56
Abbildung XIII: Einverständniserklärung 57

Tabellenverzeichnis

Tabelle I: Trennschärfen 17
Tabelle II: Deskriptive Statistik über alle Szenarien 19
Tabelle III: Deskriptive Statistik für das Szenario Appendektomie 22
Tabelle IV: Deskriptive Statistik für das Szenario Cholezystektomie 24
Tabelle V: Deskriptive Statistik für das Szenario Leistenhernienverschluss 26
Tabelle VI: Levene-Test 28
Tabelle VII: Einfaktorielle ANOVA 28
Tabelle VIII: Scheffé-Prozedur (Mehrfachvergleich) 29
Tabelle IX: Scheffé-Prozedur (Subgruppenstratifizierung) 30
Tabelle X: Bestimmung von Cronbachs α 32
Tabelle XI: Itemstatistiken für das Szenario Appendektomie Abschnitt A 33
Tabelle XII: Itemstatistiken für das Szenario Cholezystektomie Abschnitt A 34
Tabelle XIII: Itemstatistiken für das Szenario Leistenhernienverschluss Abschnitt A 34
Tabelle XIV: Itemstatistiken für das Szenario Appendektomie Abschnitt B 35
Tabelle XV: Itemstatistiken für das Szenario Cholezystektomie Abschnitt B 36
Tabelle XVI: Itemstatistiken für das Szenario Leistenhernienverschluss Abschnitt B 37
Tabelle XVII: Interrater-Reliabilität 40

Abkürzungsverzeichnis

ÄApprO	=	Approbationsordnung für Ärzte
ANOVA	=	Analysis of variance (Varianzanalyse)
BGBL	=	Bundesgesetzblatt
ggf.	=	gegebenenfalls
ggü.	=	gegenüber
ICC	=	Intraclass-Correlation
n.b.	=	nicht berechenbar
OP	=	Operation
OSCE	=	Objective Structured Clinical Examination
PAL	=	Peer Assisted Learning
V-OSCE	=	Video-assisted Objective Structured Clinical Examination

1. Einleitung

„Reden ist Gold, aber…“ so ließ Frau Dr. med. Zylka-Menhorn in ihrem Artikel im Deutschen Ärzteblatt verlauten und stellte fest, dass die Arzt-Patienten-Kommunikation von großer Wichtigkeit sei, aber noch nicht alle Fragen, wie sie sich in den Alltag der Ärzte besser integrieren lasse, geklärt seien (Zylka-Menhorn 2013). Insbesondere in der Chirurgie sei da noch Klärungsbedarf.

Diese Dissertation soll eine Methode beleuchten, den Fokus der Studentinnen und Studenten (im Folgenden werden für personenbezogene Nomina männliche Formen gebraucht, die stets auch die weibliche Form beinhalten) am Ende ihres klinischen Studienabschnittes auf die Arzt-Patienten-Kommunikation zu lenken und ihre diesbezügliche Kompetenz in einem präoperativen Aufklärungsgespräch zu prüfen, um sie damit auf den Alltag im Berufsleben vorzubereiten. Auszüge dieser Dissertation wurden bereits im “The Journal of Surgical Research” im Jahre 2014 publiziert (Kiehl et al. 2014).

1.1 Kommunikation in der Arzt-Patientenbeziehung

Seit den 1990er Jahren gibt es immer mehr Bestrebungen, die praktischen Fertigkeiten in der Medizin auch in den Fokus der medizinischen Ausbildung zu stellen (Pabst 1995). Diesem wurde 2002 durch die Änderung der Approbationsordnung für Ärzte Rechnung getragen, indem dort dieser Anteil stark erhöht wurde (ÄApprO 2002). So wurden in den darauffolgenden Jahren immer mehr praxisorientierte Lehrinhalte in die Curricula der Medizinischen Fakultäten integriert (Langewitz 2012). In einer aktuellen Umfrage von Härtl et al. an allen deutschen Medizinischen Fakultäten zeigte sich, dass mittlerweile an allen Hochschulen kommunikative Fertigkeiten gelehrt werden. Bei den meisten findet sich diese auch niedergeschrieben im Curriculum wieder (Härtl et al. 2015).

Die Lehre der Kommunikation lohnt sich, wenn man folgendes bedenkt: Positive Auswirkungen einer guten Arzt-Patienten-Kommunikation auf den Behandlungserfolg und damit langfristig auf die Gesundheit der Patienten konnten bereits belegt werden (Kaplan et al. 1989). Dieser Zusammenhang ist dabei unter anderem auf die sogenannte Compliance des Patienten – also die Bereitschaft des Patienten, einer medizinischen Empfehlung Folge zu leisten – zurückzuführen. Forschungen zur Compliance bzw. der Non-Compliance – also der Nichtbefolgung der ärztlichen Empfehlung – ergaben, dass selbst lebenswichtige Medikamente

aufgrund eines gestörten Vertrauensverhältnisses zwischen Arzt und Patient von weniger als 50% der Patienten regelmäßig eingenommen wurden (Geisler 1992). Für ein intaktes Vertrauensverhältnis zwischen den Parteien wiederum ist eine ausreichende und vertrauensfördernde Kommunikation unerlässlich (Schmitt 2009).

Insbesondere im chirurgischen Tätigkeitsfeld, in welchem die körperliche Integrität der Patienten eine besondere Bedeutung einnimmt, ist eine zugewandte und umfassende Kommunikation – somit auch im Aufklärungsgespräch – von Vorteil, um sich die Kooperation des Patienten zu sichern (Theuer et al. 2011).

Dieser Umstand erwächst nicht nur aus dem vernünftigen Menschenverstand heraus, sondern ist in erster Linie eine gesetzliche Pflicht des Arztes gegenüber seinem Patienten. Die vollständige Aufklärung eines Patienten soll dem Patienten dabei sowohl Informationen über das Krankheitsbild, die verschiedenen Therapiemöglichkeiten inklusive deren Vor- und Nachteile als auch die daraus erwachsenden Risiken und Prognosen vermitteln. Dies alles soll in einem für den Patienten verständlichen Rahmen stattfinden. Darin inkludiert ist vor allem die sprachliche Verständlichkeit (BGBL 2002). Bestärkt wurde dies noch einmal durch die Einführung des neuen Patientenrechtegesetzes am 26. Februar 2013, welches die im Bürgerlichen Gesetzbuch dargelegten Pflichten des Arztes noch einmal erweitert (BGBL 2013).

Mit dieser Festschreibung der Aufklärungspflicht verbunden ist die nun gesetzlich eindeutiger geregelte Haftbarkeit des Arztes. Die Kommunikation bekommt in diesem Zusammenhang somit eine rechtlich bedeutsame Komponente, da entsprechende Konsequenzen für den Arzt durch eine ausreichende Kommunikation und ein stabiles Vertrauensverhältnis abwendbar sind (Dillschneider et al. 2012).

Über die gesetzlichen Vorschriften hinaus erfordert eine umfassende und den gesetzlichen Auflagen entsprechende Aufklärung die adäquate Dauer und den richtigen Rahmen (Shah et al. 2011). Der Patient sollte im Mittelpunkt stehen und sich verstanden und respektiert fühlen. Dabei liegt der Fokus neben den Inhalten auch auf der Vermittlung eines Gefühls ausreichender Zuwendung durch den Arzt gegenüber dem Patienten. Gelingt dies dem Arzt nicht, kann dies zu einer geringeren Motivation des Patienten zur Mitarbeit (Compliance) mit den oben beschriebenen Konsequenzen führen (Geisler 1992).

Studien zeigen hierfür jedoch ein anderes Bild: Patienten haben das Gefühl, der Arzt – insbesondere der Chirurg – stehe unter enormem Zeitdruck. Die Folge davon ist, dass der Patient seine Fragen verschweigt, weil er dem Arzt nicht dessen kostbare Zeit stehlen will

(Barthel et al. 2005). Um dies zu umgehen, sollte der Arzt in der Lage sein, ein suffizientes Gespräch zu führen. Dies bedeutet nicht, ein Gespräch in die Länge zu ziehen, sondern vielmehr zu verkürzen, ohne den Eindruck von Zeitmangel entstehen zu lassen (Geisler 1992). Dafür ist es erforderlich zu wissen, wie man ein Gespräch inhaltlich und sprachlich straff und dennoch informativ strukturiert (Etrillard 2009). Die Erlernbarkeit dafür notwendiger Strategien konnten Maguire und Pitceathly bereits nachweisen (Maguire und Pitceathly 2002).

Nachdem nun die rechtliche und patientenzentrierte Wertigkeit der Kommunikation in der Arzt-Patienten-Beziehung beschrieben wurde, soll im Folgenden dargestellt werden, wie man einerseits den Studenten diese Fähigkeit vermitteln kann. Andererseits soll auch gezeigt werden, wie eine ausreichende Kompetenz am Ende der Ausbildung evaluiert werden kann.

1.2 Vermittlung der Kommunikationskompetenz und Auswahl der geeigneten Prüfungsform

Zur Vermittlung der Kommunikationskompetenz gibt es mannigfaltige Ansätze (Hulsman et al. 1999). Dazu gehört unter anderem die klassische Vorlesung. Diese stellt die wohl am weitesten verbreitete Methode dar. Der professionelle Lehrer vermittelt einer großen Gruppe von Studenten mit oder ohne Hilfsmittel den Lehrstoff in einer Einzelvorlesung oder in aufeinander aufbauenden Sitzungen. Eine gut strukturierte und ansprechend gehaltene Vorlesung kann eine sowohl leistungsfähige als auch hochwertige Lehrmethode darstellen – jedoch vorwiegend für die theoretische Wissensvermittlung (Brown und Manogue 2001).

Für die Vermittlung von praktischen Kompetenzen eignet sich z. B. das „Peer Assisted Learning“ (= PAL). Hier werden durch speziell geschulte studentische Tutoren Inhalte in Kleingruppen vermittelt. Das PAL hat sich aufgrund seiner guten empirischen Datenlage mittlerweile als weit verbreitete Methode in der medizinischen Ausbildung etabliert (Cate und Durning 2007). Im Zusammenhang mit kommunikativen Kompetenzen konnten Ringel et al. 2015 in einer Studie nachweisen, dass sowohl die Tutoren als auch die Studenten deutlich von diesem Lehrformat profitierten (Ringel et al. 2015).

Außerdem wird die Simulation zum besseren Wissenstransfer genutzt. Hier werden speziell geschulte (Laien-) Schauspieler verwendet, um sowohl alltägliche als auch besondere Situationen der ärztlichen Tätigkeit nachzustellen und die Studenten somit im Umgang mit dem Patienten und den situationsbedingten Erforderlichkeiten zu schulen (Wallace 1997). Auch dies

ist eine mittlerweile gut untersuchte Methode, welche sowohl in der Lehre als auch in der Prüfung Anwendung findet (McNaughton et al. 2008).

Ebenso kommt das Selbststudium unter angeleiteter Zielsetzung und Bereitstellung von Lehrmaterial zum Einsatz. Diese Methode zeichnet sich durch seinen hohen Grad an wachsendem Selbstvertrauen und Autonomie des Studenten aus, erfordert aber gleichzeitig ein hohes Maß an Motivation (O'Shea 2003). Zur noch stärkeren Motivation und Partizipation am autonomen Lernen können die Studenten zusätzlich auch in Kleingruppen unter Verzicht auf einen professionellen Tutor zusammengestellt werden. Dies zeichnet sich bei gleichem Lernerfolg zu Kleingruppen, welche durch einen Tutor angeleitet werden, durch eine bessere Gruppendynamik und Partizipation der einzelnen Teilnehmer aus (Hoffman et al. 2014).

Allerdings ist eine wissenschaftlich fundierte Lehrform noch keine Garantie für einen Lehr- bzw. Lernerfolg. Um einen Lehr-/Lernerfolg nachzuweisen, kann z. B. eine summative Prüfung am Ende einer Lehrphase genutzt werden. Dabei kann eine solche Prüfung insbesondere dazu führen, dass der Lehr-/Lernerfolg überhaupt erst stattfindet, da auch wenig motivierte Studenten zum Lernen gezwungen werden, oder den Lehr-/Lernerfolg auch noch zu steigern (Van der Vleuten 1996). Es gibt nach Millers Modell von 1990 dazu insgesamt vier Stufen der Fähigkeitsabfrage und daraus resultierend geeignete Prüfungsformen wie auch in Abbildung I als erweiterte Form des Originals von Miller dargestellt (Miller 1990, Abbildung I).

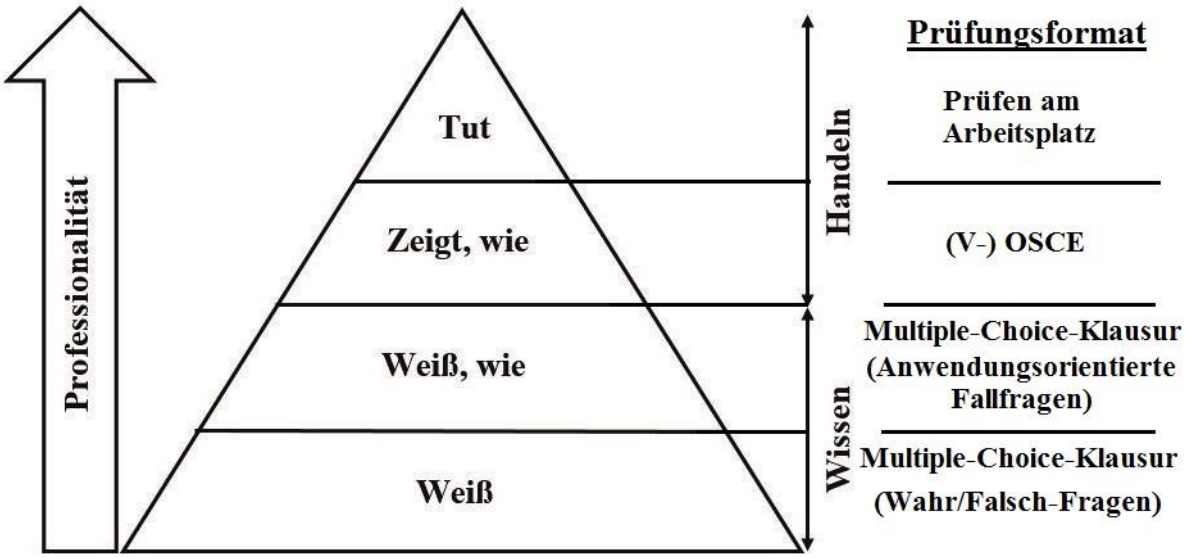


Abbildung I: Pyramide der Prüfungsstufen nach Miller in Anlehnung an das Original (Miller 1990) neu erarbeitet von C. Kiehl

Auf der ersten Stufe steht das reine Faktenwissen (Weiß) und dessen Abfrage. Als Beispiel soll der Student die Symptome einer Krankheit erlernt haben und diese in einer „Wahr/Falsch"-Abfrage z. B. im Rahmen einer Multiple-Choice-Klausur wiedergeben können. Auf der zweiten Stufe soll der Student über das reine Wissen hinaus Zusammenhänge interpretieren können (Weiß, wie). Er soll somit z. B. in einer Multiple-Choice-Klausur bei anwendungsorientierten Fragen, welche sich beispielsweise auf einen komplexen Fall aus dem medizinischen Alltag beziehen, richtige Antworten unter Distraktoren erkennen können. Die dritte Stufe überschreitet dann das rein kognitive Zusammenhangswissen und kombiniert es mit der Fähigkeit eines Studenten, das Gelernte in einem ausgewählten Setting – z. B. einer Simulation – anzuwenden (Zeigt, wie). Hierfür ist als exemplarische Prüfungsform die Objective Structured Clinical Examination (OSCE) zu nennen, welche in jüngster Zeit zunehmend in die medizinische Ausbildung Eingang findet (Patricio et. 2013, Laidlaw et al. 2014). Diese wird zu einem späteren Zeitpunkt noch genauer erläutert. Auf der vierten und höchsten Stufe sollen die gelernten Inhalte und Fähigkeiten in einem realen Setting angewendet werden (Tut), wie es zum Beispiel durch eine direkte Beobachtung am Arbeitsplatz möglich wäre.

Der Inhalt der Prüfung entscheidet dabei über das Design der Prüfung (Al-Wardy 2010). In unserem Fall soll der Student seine Fähigkeit beweisen, ein Aufklärungsgespräch vor einer Operation mit einem simulierten Patienten durchzuführen. Ein Aufklärungsgespräch vor einer Operation stellt einen komplexen Sachverhalt dar. Zur Rekapitulation sind hier die wichtigsten Stichpunkte der Miller-Pyramide aufgeführt: Die Studenten sind angehalten, sowohl über die theoretischen Hintergründe der zugrundeliegenden Erkrankung (Wissen) als auch über die alternativen Behandlungsmethoden und deren technische Umsetzung, Folgen und Risiken referieren zu können (Weiß wie). Zu diesem Wissenshintergrund tritt die Herausforderung hinzu, diese Aufklärung in einem auf den Patienten zugeschnittenen sprachlichen und sachlichen Rahmen zu gestalten (Zeigt, wie), also die dritte Stufe nach Miller. Es stehen hier also die Beobachtung einer frei gewählten Simulation oder die OSCE als standardisierte Simulation als Prüfungsformen zur Verfügung. Aufgrund der besseren Vergleichbarkeit einer Prüfungsleistung durch Standardisierung (Harden et. al 1975) haben wir uns für eine OSCE entschieden.

1.3 Objective Structured Clinical Examination

Das Prinzip einer OSCE (Objective Structured Clinical Examination) basiert auf einer Prüfung mit mehreren Stationen. Der Student durchläuft diese nacheinander mit einer festgelegten Zeit

pro Station und wird auf standardisierten Prüfungsbögen (Checklisten) von Bewertern (Rater) vor Ort bewertet. Die Vergabe der Punkte richtet sich dabei nach einheitlich formulierten Bedingungen – erstmals so beschrieben von R. Harden und Kollegen (Harden et al. 1975). Das Zusammenspiel der Stationen mit den standardisierten Prüfungsszenarien und Checklisten ergibt dann die hohe Objektivität, Reliabilität und Validität dieser Prüfungsform. Dies wird möglich durch die eingehende Schulung von Ratern und die repetitive Simulation, was zu einer annähernd gleichen Prüfungssituation für alle Prüflinge führt (Khan et al. 2013, Pell et al. 2010). Dennoch muss das Design der Prüfung sehr genau geprüft werden, um die Güte der Prüfung positiv zu beeinflussen (Turner und Dankoski 2008, Brannick et al. 2011). Reliable OSCE-Prüfungen sollten aus mindestens 10 Stationen bestehen, da erst durch die steigende Zahl von Stationen eine adäquat hohe Reliabilität erreicht werden kann (Brannick et al. 2011). Hinsichtlich der Güte der Prüfung ist auch zu berücksichtigen, dass bei OSCE-Prüfungen häufig Schauspielpatienten zum Einsatz kommen. Diese müssen ebenfalls konkret geschult werden, damit allen Studenten eine äquivalente und adäquate Prüfungssituation geschaffen wird (Cleland et al. 2009).

Dies ist besonders wichtig, wenn die Kommunikationskompetenz zwischen Patient und Student geprüft werden soll. Die Studenten müssen auf die Fragen und Emotionen der Patienten reagieren. Dies wiederum bewirkt, dass der Patient – in diesem Fall der Schauspieler – in gewissem Maße steuern kann, wie das Gespräch verläuft, indem er z. B. eher zurückhaltend oder sehr offen hinterfragend agiert. Über diesen Weg hat der Patient/Schauspieler dann letztlich auch Einfluss auf den Schwierigkeitsgrad der Prüfung und somit auch auf die Bewertung des Studenten (Levine und McGuire 1970). Ein gut geschulter Schauspieler ist daher unabdingbar, um jedem Studenten identische Prüfungsbedigungen zu gewähren (Collins und Harden 1998).

Insgesamt ist die OSCE mittlerweile weit verbreitet und anerkannt als ein objektives, reliables und valides Prüfungsformat und ihre Verbreitung nimmt stetig zu (Barzansky und Etzel 2003, Brannick et al. 2011, Turner und Dankoski 2008, Patrício et al. 2013). Im Fach Chirurgie allerdings sind die Beschreibungen und evaluierten Prüfungsszenarien jedoch noch rar (Kalbitz et al. 2010).

Allerdings gibt es im Bereich der Arzt-Patienten-Kommunikation im Fach Chirurgie ein paar Berichte über die Anwendung einer OSCE: So wurde von Chipman und Kollegen eine OSCE mit dem Ziel entwickelt, das Überbringen schlechter Nachrichten an den Patienten und seine Angehörigen in einem chirurgisch-intensivmedizinischen Setting zu üben und zu bewerten

(Chipman et al. 2007). Dabei kamen zwei verschiedene Szenarien zum Einsatz: Zum einen ein Gespräch am Lebensende des Patienten und zum anderen das „Eingeständnis“ einer iatrogenen Komplikation. Die Studie konnte mit einem positiven Lerneffekt für die Teilnehmer aufwarten.

Yudkowski und Kollegen implementierten eine OSCE zur Simulation der patientenzentrierten Kommunikation im chirurgischen Alltag. Die Besonderheit bei dieser Studie war, dass die simulierten Patienten gleichzeitig als Rater fungierten. Die Teilnehmer bewerteten diese Prüfung als gute Möglichkeit ihre kommunikativen Fähigkeiten auszubauen und schätzten die gerechte Bewertung (Yudkowski et al. 2004).

Auch die Gesprächsführung in komplexen und anspruchsvollen Szenarien wurde bereits in einer Prüfung etabliert. Unter anderem beinhaltete die Prüfung ein Szenario über die Gesprächsführung am Lebensende in einem Erwachsenen- und Kinderhospiz. Außerdem sollte eine Familienkonferenz zur weiteren Therapieentscheidung simuliert werden. Letztere forderte die Studenten besonders (Tchorz et al. 2013).

Eine OSCE zum Aufklärungsgespräch vor der Operation wurde von Govindan 2008 beschrieben. Hier war der das Aufklärungsgespräch Teil einer prä- und postinterventionellen OSCE mit insgesamt vier Stationen. Als Intervention erhielten die Teilnehmer zwischen den beiden Prüfungen Schulungen hinsichtlich der Arzt-Patienten-Kommunikation, was zu einer signifikanten Verbesserung des Abschneidens in der postinterventionellen OSCE führte. Dabei nahmen alle Teilnehmer an beiden Prüfungen teil. Eine Kontrollgruppe ohne Intervention wurde nicht untersucht (Govindan 2008).

Eine weitere OSCE zu diesem Thema wurde von Srinivasan 1999 durchgeführt. Hier wurde neben dem klinischen Wissen über das Krankheitsbild auch explizit die Empathie des Kandidaten gegenüber dem Patienten bewertet. Dabei wiesen die Kandidaten insbesondere im Bereich der Empathie Defizite auf (Srinivasan 1999).

Abschließend zu diesem Thema lässt sich zusammenfassen, dass die OSCE an sich bereits vielfach untersucht wurde und sich insbesondere auch als Prüfungsform für die Kommunikationskompetenz in der medizinischen Ausbildung etabliert hat. Die vorliegende Arbeit soll sich daher insbesondere mit dem Aspekt der klinisch-chirurgischen Prüfung befassen. Darüber hinaus sollen im Folgenden die Möglichkeiten einer Implementierung einer Videoaufzeichnung in eine OSCE näher beleuchtet werden.

1.4 Möglichkeiten und Vorteile des Einsatzes einer Videoaufzeichnung

Videoaufzeichnungen dienen sowohl als Prüfungsbestandteil (Zeigen von Videos) als auch zur Prüfungsbewertung (Aufzeichnung der Prüfung). Im Folgenden soll ein Überblick über die Einsatzmöglichkeiten gegeben werden.

In Bezug auf kommunikative Aspekte wurden beispielsweise Videos mit Arzt-Patienten-Interaktionen vorab aufgenommen und später in einer OSCE-Prüfung Studenten vorgespielt. Die Studenten sollten die Aufzeichnungen hinsichtlich guter und schlechter Kommunikationsaspekte anhand von Checklisten bewerten (Humphris und Kaney 2008). Hier war das Video somit Bestandteil der Prüfung und diente nicht der Dokumentation der Prüfung an sich.

Ähnlich diesem Konzept wurden Videos von Baribeau und Kollegen verwendet (Baribeau et al. 2012). Die Autoren erstellten 5 Videos mit Arzt-Patienten-Interaktionen und ließen diese von Studenten hinsichtlich der in den Videos vom Arzt gezeigten kommunikativen Kompetenzen analysieren. Dabei werteten die Studenten die Videos zunächst nur mit den individuellen Vorkenntnissen aus. Daraufhin erfolgte eine Intervention in Form eines Kurses zur Vermittlung von kommunikativen Kompetenzen. Im Anschluss wurden die Videos erneut von den Studenten ausgewertet. Die Studenten zeigten hier nur in Teilen der abgefragten Kompetenzen signifikante Verbesserung durch die Intervention. Die Autoren sahen daher noch Verbesserungsbedarf.

Aber auch für nicht-kommunikative Aspekte wurden bereits Videos in OSCEs eingesetzt. So wurde z. B. in der Intensivmedizin eine OSCE mit dem Ziel der Verbesserung des Umgangs mit Ultraschall im Bereich des Thorax etabliert. In der Prüfungsstation wurden Videos von Ultraschalluntersuchungen vorgespielt, die dann von den Prüflingen interpretiert werden mussten (Breitkreutz et al. 2013).

Anders wurden Videos von Woodward-Kron und Kollegen verwendet. Sie verwendeten Videoaufzeichnungen von einer vorangegangenen OSCE für die Entwicklung einer Trainings-Webseite. Bei der im Vorfeld veranstalten OSCE sollten Ärzte in 4 Stationen ihre Fähigkeiten in der Kommunikation mit Patienten zeigen. Diese Arzt-Patienten-Interaktionen wurden gefilmt und diese Filme hinsichtlich beobachtbarer Schwächen und Stärken in der Kommunikation analysiert. Aus den Erkenntnissen dieser Untersuchung wurde eine Trainingswebsite zur Verbesserung der kommunikativen Kompetenzen erstellt. (Woodward-Kron et al. 2015).

Aufzeichnungen einer OSCE wurden ebenfalls bei Vivekananda-Schmidt und Kollegen verwendet (Vivekananda-Schmidt et al. 2007). Hier erfolgte bei einer orthopädischen OSCE der direkte Vergleich von Bewertungen anwesender Prüfer mit denen von Prüfern, die im Nachhinein die Leistungen der Studenten anhand der Videoaufzeichnungen bewerteten. Damit konnten die Kollegen die Vergleichbarkeit der beiden Bewertungsmodalitäten nachgewiesen werden.

In einer anderen Studie von Chen et al. wurden OSCE-Videos mehreren Ratern vorgelegt. Deren Bewertungen wurden dann verglichen und damit Rater identifiziert, die weiteren Schulungsbedarf hatten, um die OSCE möglichst reliabel zu gestalten (Chen et al. 2013).

Eine weitere Möglichkeit des Einsatzes von Videoaufzeichnungen wurden von Maloney und Kollegen beleuchtet. Während eine Kontrollgruppe zu einem prüfungsrelevanten Thema lediglich mittels fremder Videos angeleitet wurde, filmte die Interventionsgruppe eigene Videos zum Prüfungsthema zur Selbstreflektion. In der abschließend stattfindenden OSCE konnte gezeigt werden, dass die Interventionsgruppe signifikant besser abschnitt (Maloney et al. 2013b).

Fast man die Ergebnisse zusammen, so gibt es mannigfaltige Möglichkeiten, Videomaterial in einer OSCE-Prüfung einzusetzen. Sowohl als Prüfungsobjekts als auch zur Bewertung der Prüfung wurden Videos bereits verwendet. In dieser Arbeit soll der Einsatz von Videos zur Bewertung der Prüfung genauer beleuchtet werden.

2. Zielsetzung und Fragestellung

Die kommunikativen Kompetenzen im Arztberuf allgemein und im chirurgischen Tätigkeitsfeld im Speziellen erlangen eine immer größere Bedeutung (vgl. Kapitel 1.1). Dabei ist die Wahl der Prüfungsform von entscheidender Bedeutung, da Prüfungen das Lernverhalten steuern können (vgl. Kapitel 1.2).

In der vorliegenden Arbeit haben wir den Einsatz einer OSCE mit Videoaufzeichnung zur zeitversetzten Bewertung der studentischen Leistungen untersucht. Ziel war es, eine qualitätsgesicherte Prüfung zu etablieren. Dabei wurde folgende Fragestellung zu Grunde gelegt:

1) Sind die verwendeten Prüfungsszenarien vergleichbar schwer?
2) Sind die verwendeten Checklisten inhaltlich konsistent? Gibt es zu schwierige/zu leichte Items? Weisen die Items eine hohe Trennschärfe auf? Gibt es Unterschiede bezüglich der Gütekriterien zwischen den verwendeten Szenarien?
3) Ist die Bewertung der Rater der studentischen Prüfungsleistung kongruent?

3. Methoden

3.1 Studiendesign und die Teilnehmer

Die Studie wurde als Querschnittsstudie über eine Kohorte eines sechsten klinischen Semesters durchgeführt. Eingeschlossen wurden alle Studenten, welche am Pflichtmodul 6.1 „Operative Medizin“ entsprechend dem Göttinger Lernzielkatalog (Pflichtmodule 2014, Lernzielkatalog klinischer Studienabschnitt 2008, Projektion Lernzielkatalog 2008) im Sommersemester 2011 teilnahmen und ihre Zustimmung zur wissenschaftlichen Datenauswertung gaben.

Das Curriculum der Universitätsmedizin Göttingen der Georg-August-Universität Göttingen sieht ein sechsjähriges theoretisch-praktisches Studium gefolgt von einem Jahr mit arbeitsplatzbasierter Ausbildung (Praktisches Jahr = PJ) an einem Lehrkrankenhaus der Göttinger Universitätsmedizin oder im Universitätsklinikum Göttingen selbst vor. Es teilt sich dabei in einen vorklinischen Abschnitt von vier Semestern und einen klinischen Abschnitt mit sechs Semestern (Lernzielkatalog klinischer Studienabschnitt 2008, Projektion Lernzielkatalog 2008). Im klinischen Abschnitt erfolgt die Lehre modularisiert. Hier werden Organe- und Funktionssysteme, die entsprechenden Krankheitsbilder und deren verschiedene Therapieoptionen fächerübergreifend gelehrt und in der Regel mit einer Modulklausur im Multiple-Choice-Verfahren (Vyas und Supe 2008) abgeschlossen.

Zum Zeitpunkt der hier untersuchten Prüfung befanden sich die Studenten kurz vor dem Ende ihrer modularisierten Ausbildung im sechsten klinischen Semester. Bereits im ersten klinischen Semester haben die Studenten im Modul 1.1 „Ärztliche Basisfertigkeiten“ praktische Lehre zum Thema Arzt-Patienten-Kommunikation mit speziellem Fokus auf die Anamneseerhebung erhalten. Hier hatten sie bereits Kontakt mit Schauspielpatienten und mussten in Rollenspielen die Rolle des Arztes übernehmen. In den folgenden Semestern besuchten die Studenten die Module – insbesondere in Modul 4.3 „Erkrankungen des Verdauungsapparates“, in denen sie umfassend neben den internistischen Herangehensweisen auch die chirurgischen Inhalte zu den Krankheitsbildern erlernten. Zur weiteren Vertiefung der chirurgischen Vorgehensweise absolvierten die Studenten in der vorlesungsfreien Zeit des vierten klinischen Semesters ein einwöchiges Blockpraktikum. Hier kamen die Studenten sowohl mit unfallchirurgischen und orthopädischen als auch mit allgemein- und viszeralchirurgischen Patienten in Kontakt.

Im Modul 6.1 „Operative Medizin“ wurden die bereits erlernten Grundlagen für die Durchführung eines Aufklärungsgespächs durch repetitive Vorlesungen und Seminare sowie angeleitetes Selbststudium – Effektivität belegt u. a. bei Brydges et al. 2015 – vertieft.

In der Einführungsveranstaltung des Moduls erhielten die Studenten u. a. einen Überblick über die Anforderungen, die Lernziele und Prüfungsthemen sowie den organisatorischen und technischen Ablauf der Prüfung. Außerdem wurde eine Plenarveranstaltung/Vorlesung über die die Grundlagen des chirurgischen Aufklärungsgesprächs gehalten. Des Weiteren wurden den Studenten auf den Stud.IP-Servern, welche in der Universität Göttingen allgemein als Plattform zur Verbreitung von veranstaltungsbegleitenden Informationsmaterial genutzt wurden, Aufklärungsbögen von Diomed von Thieme Compliance zur Vorbereitung auf die Prüfung zur Verfügung gestellt. Diese enthielten detaillierte Darstellungen der Krankheitsbilder, operativen Abläufe, Komplikationen und Nachsorge. Zusätzlich standen den Studenten ebenfalls auf den Stud.IP-Servern die aktuellen und aus vorherigen Modulen stammenden Vorlesungs- und Seminarpräsentationsfolien zur Verfügung.

Das universitätseigene Ethikkommitee und der Datenschutzbeauftragte zeigten nach Konsultation keinerlei Bedenken und erklärten den Verzicht auf ein schriftliches Genehmigungsverfahren mit Verweis auf die vorgeschlagene schriftliche Einverständniserklärung. Hierin stimmten die Studenten sowohl der Videoaufzeichnung als auch deren Speicherung und Auswertung zu und traten die Besitzansprüche zu Gunsten der Universität ab (Abbildung XIII). Des Weiteren versicherten die Studenten, keine Kopien zu erstellen.

3.2 Prüfungsgruppen und Szenarien

Die 155 Studenten teilten sich selbstständig zu Beginn des Moduls in 77 Zweier-Prüfungsgruppen und eine Einzelprüfung auf und wählten einen der verfügbaren 78 Termine jeweils an einem Donnerstagnachmittag während der fünf Modulwochen zur Prüfung. An jedem Termin fanden zu 6 Zeitpunkten jeweils 3 Prüfungen gleichzeitig statt. Dabei wurden die Szenarien nach einem Zufallsprinzip rotiert, so dass den Studenten das Prüfungsthema von der vorangegangenen Prüfungsgruppe nicht mitgeteilt werden konnte. Alle drei Prüfungsthemen waren zuvor im Curriculum in den Lehrformaten Vorlesung und Seminar unterrichtet worden: Akute Appendizitis, symptomatische Cholezystolithiasis und Leistenhernie mit den Operationsverfahren der laparoskopischen Appendektomie, laparoskopischen Cholezystektomie und dem offenen Leistenhernienverschluss mit Netzeinlage (Operationsverfahren nach Lichtenstein).

In der Prüfung hatten die Studenten die Aufgabe, einen Schauspielpatienten (SP) über die bevorstehende Operation aufzuklären. Dabei wurde vorausgesetzt, dass der SP über die jeweilige Operationsindikation informiert und damit einverstanden war, so dass die Operationsaufklärung im Vordergrund stand. Die Schauspielpatienten stammten aus einem Pool von teils professionellen, teils Laien-Schauspielern. Diese waren im Vorfeld für ihre Rolle mit kurzer Biographie und für die drei Krankheitsbilder in Hinsicht auf deren Symptome und Verläufe geschult worden. Um eine Interaktion mit den Studenten zu erleichtern, wurden fünf Leitfragen für die SP definiert.

Für die SP wurde ebenfalls eine Einverständniserklärung in die Videoaufzeichnung, Speicherung und Auswertung erstellt.

3.3 Ablauf der Prüfung

In Abwandlung zur klassischen OSCE wurde die hier untersuchte Prüfung – die Video-Objective Structured Clinical Examination (V-OSCE) – als Einzelstations-OSCE ausgeführt. Sie sollte als Pilotprojekt für eine noch zu entwickelnde OSCE vor dem PJ dienen.

Die studentischen Zweiergruppen hatten ein Zeitfenster von 30 Minuten und damit Gelegenheit, je ein Aufklärungsgespräch von maximal 10 Minuten durchzuführen. Zwei studentische Hilfskräfte assistierten bei technischen Problemen und stellten den Rücklauf der unterzeichneten Einverständnisbögen sicher. Während der Prüfung standen den Studenten die Aufklärungsbögen von Diomed zur Verfügung sowie ggf. selbst erstellte Notizen.

Nach der Vorstellung beim Schauspieler wurde die Prüfung gestartet. Dazu begann ein Student mit dem Aufklärungsgespräch, während der zweite die Kamera bediente. Nach dem Abschluss der Aufklärung wurde gewechselt und ein zweites Video gedreht.

Nach Aufzeichnung der beiden Videos erhielten die Studenten auf Wunsch ein direktes Feedback durch den Schauspieler – sofern noch ein ausreichendes Zeitfenster bestand. Zur Auswertung der Videos wechselten die Studenten in einen benachbarten Raum und hatten ein weiteres Zeitfenster von 30 Minuten zur Durchsicht beider Videos, gaben sich Peer-Feedback (Cho und MacArthur 2011) und konnten ein Video auswählen, das dann anschließend als Dateitestat auf zwei externe Festplatten (redundante Speicherung) überspielt wurde. Hiermit galt die Prüfung als beendet. Das andere Video wurde gelöscht und ging nicht in die Bewertung ein.

3.4 Bewertungsbögen (= Checklisten)

Die Checklisten enthielten insgesamt 26 Items (Abbildung X, Abbildung XI, Abbildung XII). Die sich in einen A- und einen B-Abschnitt untergliederten. Im Abschnitt A (7 Items) wurden die kommunikative Kompetenz und die Patienteninteraktion erfasst.

Die Bewertung erfolgte mit einer sechsstufigen Likert-Skala (6 = positiver Pol, 1 = negativer Pol) – so beschrieben bei Cox und Matell (Cox 1980, Matell und Jacoby 1972). Die erreichbare Maximalpunktzahl betrug 42 Punkte im Abschnitt A.

Abschnitt B umfasste insgesamt 19 Items. Davon wurden zwei Items wiederum durch eine sechsstufige Likert-Skala zur Bewertung des Aufklärungsgesprächs abgebildet. Die übrigen 17 Items wurde auf dem Nominalskalenniveau als: „genannt bzw. nicht genannt“ bewertet.

Die Checklisten waren einheitlich gestaltet – soweit inhaltlich möglich. Dabei waren die allgemeinen Operationsrisiken mit acht Items bei allen Szenarien gleich. Die operationsspezifischen Risiken umfassten fünf Items. Dabei erfasste jeweils ein Item die Verletzung von Nachbarorganen und von Nachbarstrukturen passend zum Szenario. Die übrigen drei Items waren für das Szenario Appendektomie und Cholezystektomie gleich und wurden nur für das Szenario Leistenhernienverschluss speziell angepasst. Die möglichen Erweiterungen des Eingriffs und das postoperative Management waren mit je zwei Items zahlenmäßig vertreten.

Insgesamt konnten im Abschnitt B der Checkliste maximal 46 Punkte erreicht werden.

Die erreichten Punkte aus Abschnitt A und B wurden im Verhältnis 3:7 addiert. Daraufhin wurden die Rohpunkte in Prozentpunkte umgerechnet und dann mit der definierten Bestehensgrenze von 70% abgeglichen. Die Prüfung wurde als Bestanden- oder Nicht-Bestanden-Prüfung durchgeführt, wobei das Bestehen der Prüfung für den erfolgreichen Abschluss des Leistungsnachweises im Fach Chirurgie vorausgesetzt wurde.

Die Gewichtung der beiden Abschnitte in 3:7 erfolgte aufgrund der Feststellung aus dem vorangegangenen Semester, dass ein Bestehen der Prüfung allein durch sicheres Auftreten und freundliche Zuwendung zum Patienten aufgrund der dadurch bedingten hohen Punktzahl in Abschnitt A ergänzt durch nur die Nennung des Operationsverfahrens in Abschnitt B möglich war; ohne dabei auf weitere Eckpunkte der Operation eingehen zu müssen.

3.5 Rater

Die beiden Rater waren Assistenzärzte im 2. bis 3. Weiterbildungsjahr an der Universitätsmedizin Göttingen, die sich freiwillig für den Einsatz als Rater gemeldet hatten. Dabei befand sich Rater 1 in der Weiterbildung zum Facharzt für Allgemein- und Viszeralchirurgie und Rater 2 war ein Zahnarzt in der Weiterbildung zum Oralchirurgen. Beide Rater hatten keine Vorerfahrung in der Bewertung von OSCEs, wurden aber im Vorhinein in der Bewertung geschult und auf die drei Szenarien vorbereitet.

Die Bewertungen erfolgten außerhalb der Dienstzeit und unabhängig voneinander anhand einer digitalen Kopie der Videos, die auf einer externen Festplatte zur Verfügung gestellt wurden. Die Rater erhielten eine finanzielle Aufwandsentschädigung.

3.6 Statistische Auswertung

Die statistische Auswertung erfolgte mit dem Programm SPSS Statistics der Firma IBM in der Version 21. Für die gesamte statistische Auswertung wurde der α-Fehler und somit das Signifikanzniveau auf 0,05 festgelegt. Zur leichteren Veranschaulichung wird eine Signifikanz von < 0,05 in den Tabellen ohne weitere Nachkommastellen angegeben.

Die deskriptive Statistik umfasste eine Auswertung der globalen Bewertung untergliedert in Abschnitt A, Abschnitt B und das Gesamtabschneiden. Dabei wurden zunächst alle Rohpunkte auf 0 als Basis normiert und in Prozent umgerechnet und somit einer Intervallskala angepasst. Es wurden dann die Mittelwerte der Bewertungen in Prozent von Rater 1 und 2 nach Umrechnung auf Intervallskalenniveau – wie oben beschrieben – errechnet. Zur Feststellung einer Normalverteilung wurde ein Kolmogorov-Smirnov-Test für die Gesamtverteilung durchgeführt und außerdem deren Schiefe und Kurtosis bestimmt. Der Kolmogorov-Smirnov-Test dient als Indikator für eine Normalverteilung gegenüber einer nicht normalverteilten Grundgesamtheit, während Schiefe und Kurtosis Rückschlüsse auf einen zu schwierigen oder zu leichten Test zulassen. Die Nullhypothese des Kolmogorov-Smirnov-Tests laute, dass es sich bei den Werten um eine Normalverteilung handele. Die Alternativhypothese dementsprechend laute, dass es sich nicht um eine Normalverteilung handele. Bei einem Signifikanzniveau von 0,05 gilt der Test mit Signifikanzen >0,05 als positiv und die Werte damit als normalverteilt. Werte für die Schiefe von >0 zeigen eine rechtsschiefe, also linkssteile Verteilung an und geben den Hinweis auf einen zu schweren Test. Werte von <0 deuten mit einer linksschiefen, also rechtssteilen Verteilung einen zu leichten Test an. Ein Wert von 0

entspricht der Normalverteilung und somit einem angemessenen Schwierigkeitsniveau. Ein Wert von 0 bei der Kurtosis entspricht ebenfalls eine Normalverteilung. Werte von >0 deuten auf eine spitze Verteilung mit Sammlung der Werte im Zentrum hin. Werte von <0 zeigen eine flache Verteilung mit Sammlung von Werten an den Seiten. Des Weiteren wurden Mittelwerte, Standardabweichungen und Konfidenzintervalle zur Darstellung der durchschnittlichen globalen Leistung berechnet – unterteilt nach den drei Szenarien und der Gesamtleistung (Janssen und Laatz 2013, Kuckartz et al. 2013).

Zur genaueren Betrachtung, ob es zwischen den Szenarien signifikante Bewertungsunterschiede gab, wurde eine einfache Varianzanalyse (ANOVA) als Erweiterung des T-Tests für mehr als zwei Gruppen und Scheffé-Prozeduren zur weiteren Unterteilung nach Subgruppen innerhalb der Varianzgesamtheit durchgeführt. Zur Prüfung der Grundvoraussetzungen für diese beiden Berechnungen wurde ein Levene-Test auf Varianzhomogenität durchgeführt (Kuckartz et al. 2013, Scheffé 1999). Die ANOVA habe dabei als Nullhypothese, dass es keinen Unterschied in der Bewertung zwischen den einzelnen Szenarien gebe. Die Alternativhypothese sei, dass es einen Unterschied in der Bewertung zwischen mindestens zwei der drei Szenarien gebe. Die Scheffé-Prozedur habe die gleiche Nullhypothese wie die ANOVA, bezogen auf den Vergleich von zwei Szenarien zu einem anderen Szenario. Der Levene-Test habe die Nullhypothese, dass Varianzhomogenität herrsche. Die Alternativhypothese sei, dass Varianzheterogenität herrsche.

Zur weiteren Item-Analyse nach der klassischen Test-Theorie wurden die Itemschwierigkeit und die Trennschärfe der Items bestimmt. Diese stellen die Gütekriterien eines Items dar. Dabei trifft die Itemschwierigkeit auf einer Skala zwischen 0 und 1 eine relative Aussage darüber, wie viele Studenten die abgefragte Aufgabe (das Item) erfüllen konnten. Dabei heißt ein Wert von 1, dass alle Studenten die Aufgabe gelöst haben. Das Item also sehr leicht zu bewältigen war. Ein Wert von 0 hingegen bedeutet, dass kein Student die Aufgabe bewältigt hat. Werte um 0,5 sprechen für ein Item, das am besten zwischen guten und schlechten Studenten unterscheiden kann. Dabei sollten die Items gleichmäßig über einen Bereich von 0,05 bis 0,95 verteilt sein mit einer Häufung zwischen 0,2 und 0,8 (Kelava und Moosbrugger 2012). Für diese und weitere Berechnungen wurden zur leichteren Darstellbarkeit die Skalen der Checklisten (Skalenwerte 1-2 in Abschnitt B bzw. Skalenwerte 1-6 in Abschnitt A und B) auf 0 als Basis transformiert (neue Skalenwerte 0-1 in Abschnitt B und neue Skalenwerte 0-5 in Abschnitt A und B) und die Checklisten wurden getrennt nach Abschnitt A und Abschnitt B betrachtet.

Ebenso wurde die Trennschärfe bestimmt, welche eine Aussage darüber zulässt, inwiefern das Abschneiden eines Probanden in einem einzelnen Item mit dem Abschneiden des Probanden hinsichtlich der Gesamtheit aller Items korreliert. Die Trennschärfe kann Werte von -1 bis 1 annehmen. Dabei heißt ein Wert von 1, dass dieses Item dasselbe Kriterium misst wie alle anderen Items zusammen. Ein Wert von -1 würde bedeuten, dass es das genaue Gegenteil misst. Bei der Erstellung von Fragebögen sind in der Regel Trennschärfen zwischen 0,4 und 0,7 anzustreben (Kelava und Moosbrugger 2012). Eine Einteilung der Güte einer Trennschärfe für den medizinischen Prüfungskontext wurde von Möltner und Kollegen veröffentlicht (Tabelle I, Möltner et al. 2006).

Trennschärfe	Qualität	weiteres Verfahren
<0,100	schlecht	verwerfen respektive genau überprüfen
0,101 - 0,200	moderat	beibehalten, aber nochmal überprüfen
>0,200	gut	Möglichkeiten zur Verbesserung?

Tabelle I: Trennschärfen

In dieser Dissertation sollen daher ebenfalls die o.g. Grenzwerte gelten.

Zur Einschätzung der Reliabilität der Prüfung wurden die Checklisten einer internen Konsistenzprüfung (Cronbachs α) unterzogen. Die Reliabilität der Prüfung steigt, je homogener die Items gemessen werden. Das Cronbachs α kann Werte zwischen 0 (keine Homogenität) und 1 (vollständige Homogenität) annehmen. Werte von >0,7 werden als gut angesehen (Kelava und Moosbrugger 2012, Cronbach 1951, Schmitt 1996).

Für die Bestimmung des Übereinstimmungsgrades beider Rater wurden mehrere Testverfahren genutzt. Zuerst wurde die absolute Mittelwertabweichung zwischen den Ratern errechnet. Anschließend wurde eine einfache Korrelation ρ nach Pearson vorgenommen. Diese kann Werte zwischen -1 und 1 annehmen. Ein Wert von 1 spricht für eine vollkommene Übereinstimmung und ein Wert von -1 für divergierende Bewertungen. Der ermittelte Wert kann durch Multiplikation mit 100 als prozentuale Übereinstimmung wiedergegeben werden. Dabei sprechen Werte für ρ >0,7 für eine sehr hohe Korrelation (Kuckartz et al. 2013).

Des Weiteren wurde eine Intraklassen-Korrelation (= ICC) durchgeführt. Neben der einfachen Bestimmung des Übereinstimmungsgrads werden hier noch die Varianzen der Beobachtung einzelner Items mit den Varianzen der Gesamtheit verglichen. Der ICC kann zur Reliabilitätsbetrachtung Werte zwischen 0 (keine Übereinstimmung) und 1 (vollständige Übereinstimmung) annehmen. Für hohe Korrelationen ergeben sich für Werte >0,7. Es gibt

verschiedene Modelle zur Berechnung des ICC, welche sich nach den Parametern des Tests richten. In der vorliegenden Arbeit wurde das Modell ICC 2.1 verwandt. Dies ergab sich aus der Konstellation, dass die Rater zufällig ausgesucht wurden, alle Videos getrennt angeschaut wurden und anschließend die Einzelbewertungen verglichen wurden. Es wurde dabei das justierte Modell im Sinne der milderen Prüfung zu Grunde gelegt (Shrout und Fleiss 1979, Wirtz und Caspar 2002).

Die Berechnung der Korrelation nach Pearson und des ICC erfolgte einmal für alle Szenarien zusammen und jeweils für die einzelnen Szenarien getrennt.

4. Ergebnisse

4.1 Deskriptive Statistik und Szenarienvergleich

4.1.1 Deskriptive Statistik über alle Szenarien

	Deskriptive Statistik über alle Szenarien			
	Mittelwert [%]	Standard-abweichung [%]	Konfidenzintervall [%]	
			untere Grenze	obere Grenze
Abschnitt A	96,5	3,4	95,9	97,0
Abschnitt B	85,9	6,0	85,0	86,9
Gesamt	88,9	4,6	88,2	89,6
	Minimum [%]	**Maximum [%]**	**Schiefe**	**Kurtosis**
Abschnitt A	88,1	100,0	-0,700	-0,577
Abschnitt B	68,5	97,8	-0,903	0,777
Gesamt	76,6	98,4	-0,902	0,662
	Kolmogorov-Smirnov-Test			
	Signifikanz			
Abschnitt A	<0,05			
Abschnitt B	<0,05			
Gesamt	<0,05			

Tabelle II: Deskriptive Statistik über alle Szenarien

Die durchschnittliche Bewertung der Studenten durch beide Rater war mit 88,9% (±4,6%) hoch. Das Gesamtspektrum der erbrachten Leistungen rangierte von 76,6% bis 98,4%. Die Durchführung des Kolmogorov-Smirnov-Tests deutete mit Signifikanzen <0,05 darauf hin, dass es sich bei den Werten nicht um eine Normalverteilung handelte. Die Bestimmung von Schiefe und Kurtosis ergab ebenfalls Werte von ≠0 (Tabelle II).

Betrachtete man die Abweichung der Kurtosis- und Schiefe- Werte und die Darstellung im Histogramm der Gesamtbewertung, so zeigte sich, dass es sich um eine leicht linksschiefe – also rechtssteile – und etwas spitze Verteilung handelte (Abbildung II, Tabelle II).

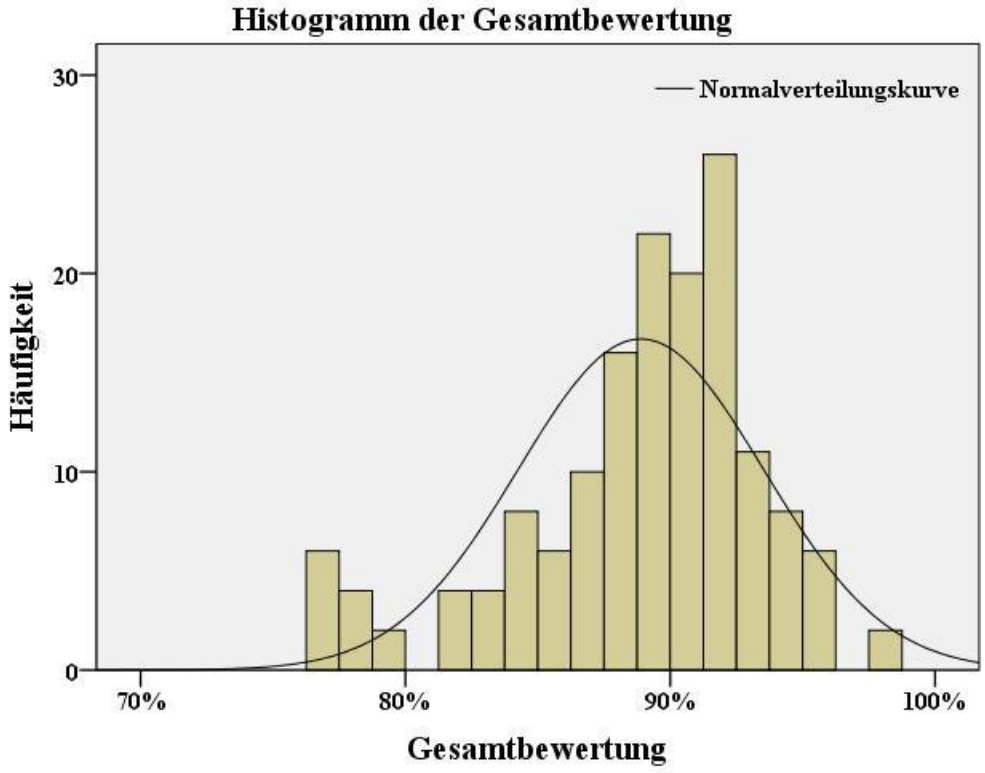


Abbildung II: Histogramm der Gesamtbewertung

Aufgetragen wurde im Histogramm der Gesamtbewertung die absolute Häufigkeit des Mittelwertes aus den Bewertungen beider Rater in Prozent. Die Normalverteilungskurve veranschaulicht dabei die Form einer theoretischen Normalverteilung.

Im Quantile-Quantile-Diagramm (Q-Q-Diagramm) wurden die von uns gemessenen Werte der Größe nach geordnet und einer theoretischen Normalverteilung gegenüber aufgetragen. Die aufgetragenen Wertepaare lagen hier nahe an der interpolierten Geraden einer perfekten Normalverteilung (Abbildung III).

Es ließ sich somit schlussfolgern, dass auch unsere Werte trotz des negativen Ergebnisses im Kolmogorov-Smirnov-Test einer Normalverteilung entsprachen. Betrachtete man dazu die absoluten Werte der Schiefe und Kurtosis von <1, so war dies hinreichend mit einer Normalverteilung vereinbar (Abbildung III, Miles und Shevlin 2006). Vor diesem Hintergrund erachteten wir die einzelnen Szenarien als Teil der Gesamtheit als ebenfalls hinreichend normalverteilt und verzichteten im Folgenden auf den Kolmogorov-Smirnov-Test. In Hinblick auf die Schwierigkeit der Prüfung sprachen Schiefe und Kurtosis bei Werten <1 für eine insgesamt weder zu schwierige noch zu leichte Prüfung (Lienert und Raatz 1998). Unter der

Fragestellung, ob dies auch für die einzelnen Szenarien gelten würde, wurden Schiefe und Kurtosis für die einzelnen Szenarien neu berechnet.

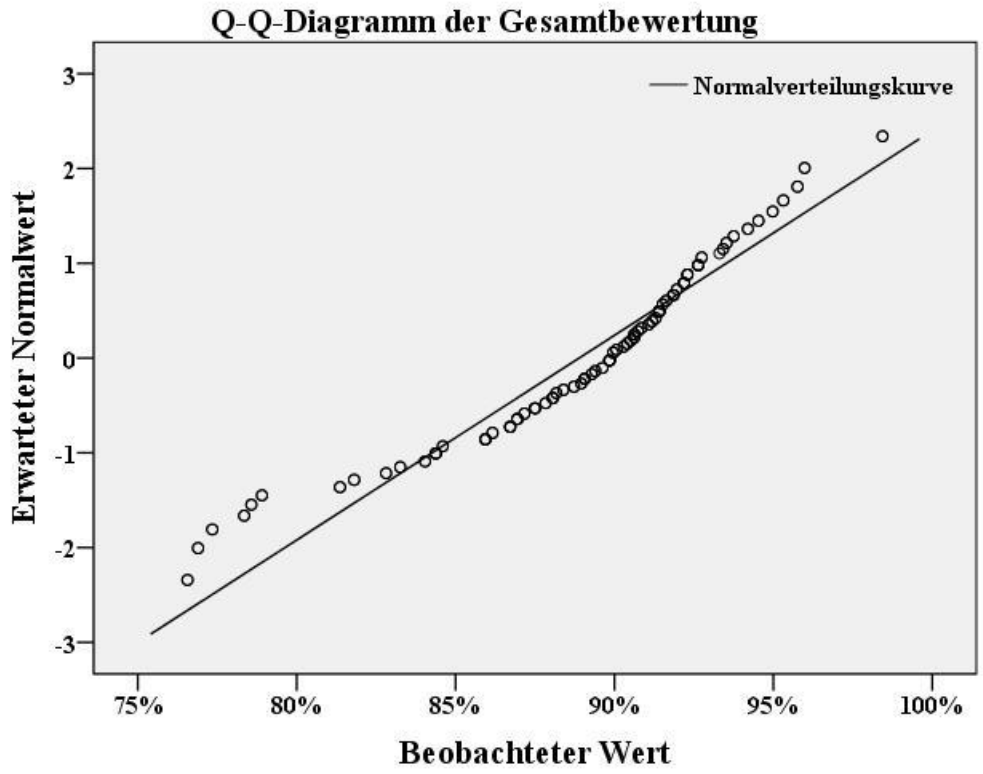


Abbildung III: Q-Q-Diagramm der Gesamtbewertung

Aufgetragen wurden im Quantilen-Quantilen-Diagramm (Q-Q-Diagramm) die von uns gemessenen Werte der Größe nach geordnet und gegenüber dem rechnerisch erwarteten Normalwert einer theoretischen Normalverteilung aufgereiht. Die Normalverteilungskurve spiegelt dabei die Punkteverteilung einer perfekten Normalverteilung wider.

4.1.2 Deskriptive Statistik für das Szenario Appendektomie

	Deskriptive Statistik für das Szenario Appendektomie			
	Mittelwert [%]	Standard-abweichung [%]	Konfidenzintervall [%]	
			untere Grenze	obere Grenze
Abschnitt A	97,8	2,2	97,2	98,4
Abschnitt B	88,4	5,4	86,9	89,8
Gesamt	91,0	4,0	90,0	92,1
	Minimum [%]	**Maximum [%]**	**Schiefe**	**Kurtosis**
Abschnitt A	92,9	100,0	-0,650	-0,850
Abschnitt B	71,7	97,8	-1,044	1,963
Gesamt	78,4	98,4	-1,079	2,298

Tabelle III: Deskriptive Statistik für das Szenario Appendektomie

Für das Szenario Appendektomie ließen sich insgesamt im Vergleich zu allen anderen Szenarien mit 91,0% (± 4,0%) etwas höhere Mittelwerte in der Gesamtbewertung berechnen (Tabelle III, Tabelle IV, Tabelle V). Das Spektrum der Bewertungen war mit 78,4% bis 98,4% kleiner. Passend dazu zeigte die Gesamtbewertung durch die linksschiefe Verteilung einen leichteren Test und eine stärkere Häufung im Zentrum der Verteilung (Tabelle III). Die aufgetragene Verteilung folgt weiterhin annähernd dem Verlauf der Normalverteilungskurve (Abbildung IV).

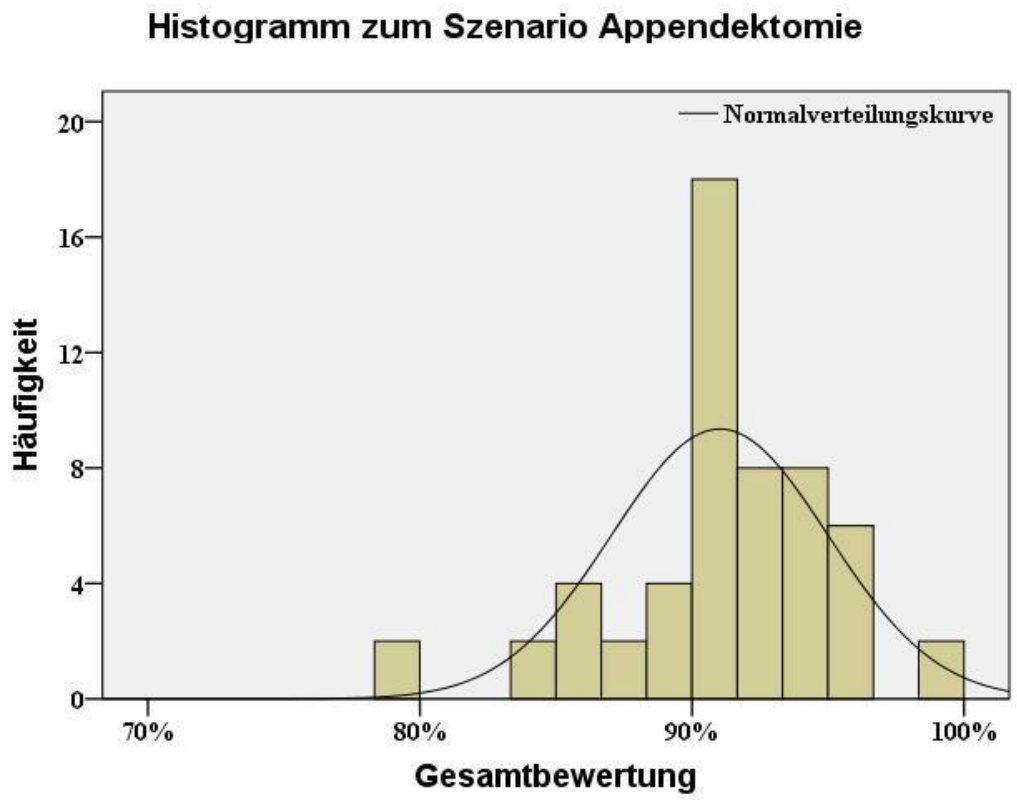


Abbildung IV: Histogramm zum Szenario Appendektomie

Aufgetragen wurde im Histogramm zum Szenario Appendektomie die absolute Häufigkeit des Mittelwertes aus den Bewertungen beider Rater in Prozent. Die Normalverteilungskurve veranschaulicht dabei die Form einer theoretischen Normalverteilung.

4.1.3 Deskriptive Statistik für das Szenario Cholezystektomie

	Deskriptive Statistik für das Szenario Cholezystektomie			
	Mittelwert [%]	Standard-abweichung [%]	Konfidenzintervall [%]	
			untere Grenze	obere Grenze
Abschnitt A	95,8	3,5	94,8	96,8
Abschnitt B	85,5	5,6	83,8	87,1
Gesamt	88,4	4,4	87,1	89,6
	Minimum [%]	**Maximum [%]**	**Schiefe**	**Kurtosis**
Abschnitt A	89,3	100,0	-0,252	-1,420
Abschnitt B	70,7	92,4	-1,121	0,755
Gesamt	76,9	94,2	-1,256	0,863

Tabelle IV: Deskriptive Statistik für das Szenario Cholezystektomie

Hier zeigten sich mit 88,4% (± 4,4%) bei der Gesamtbewertung ähnliche Werte im Vergleich zur Gesamtbewertung aller Szenarien, jedoch im Mittel schlechtere Bewertungen als im Szenario Appendektomie (Tabelle II, Tabelle III, Tabelle IV). Trotzdem zeigte sich dieses Szenario in der Gesamtbewertung noch linksschiefer als das Szenario Appendektomie, dafür aber breiter verteilt (Tabelle III, Tabelle IV). Dieser Indikator für die Schwierigkeit ließ sich auch graphisch im Histogramm darstellen (Abbildung V).

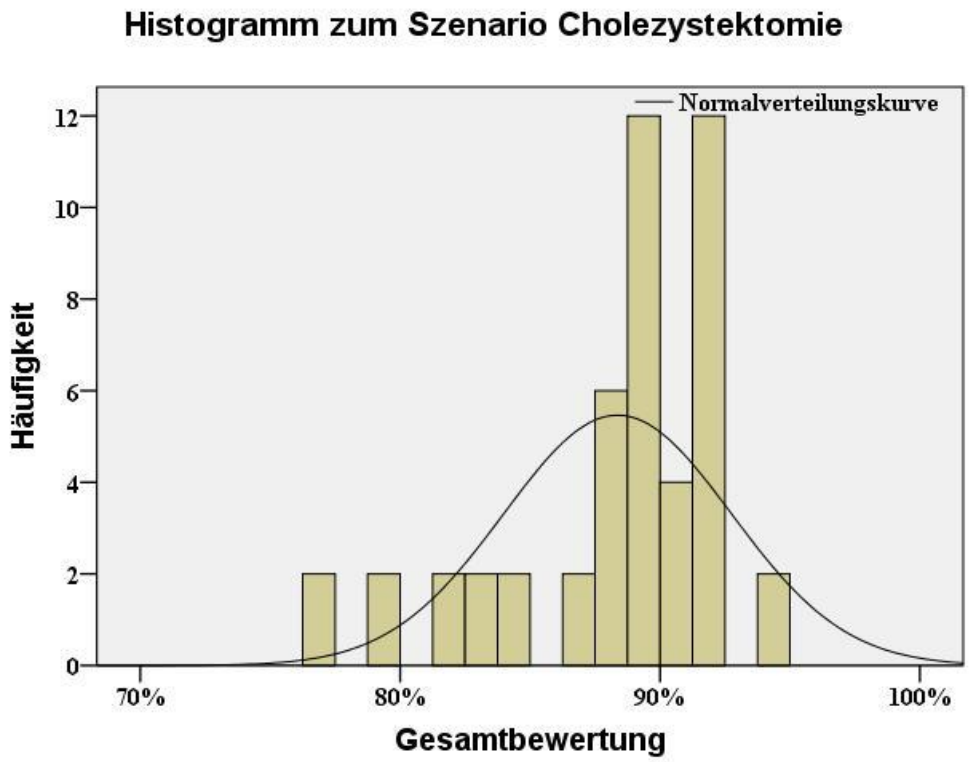


Abbildung V: Histogramm zum Szenario Cholezystektomie

Aufgetragen wurde im Histogramm zum Szenario Cholezystektomie die absolute Häufigkeit des Mittelwertes aus den Bewertungen beider Rater in Prozent. Die Normalverteilungskurve veranschaulicht dabei die Form einer theoretischen Normalverteilung.

4.1.4 Deskriptive Statistik für das Szenario Leistenhernienverschluss

	Deskriptive Statistik für das Szenario Leistenhernienverschluss			
	Mittelwert [%]	Standard-abweichung [%]	Konfidenzintervall [%]	
			untere Grenze	obere Grenze
Abschnitt A	95,7	3,9	94,6	96,7
Abschnitt B	83,7	6,1	81,9	85,4
Gesamt	87,0	4,7	85,7	88,4
	Minimum [%]	**Maximum [%]**	**Schiefe**	**Kurtosis**
Abschnitt A	88,1	100,0	-0,492	-0,959
Abschnitt B	68,5	93,5	-0,849	0,463
Gesamt	76,6	95,0	-0,739	0,093

Tabelle V: Deskriptive Statistik für das Szenario Leistenhernienverschluss

Die mittlere Gesamtbewertung war mit 87,0% (± 4,7%) niedriger als der Durchschnitt und niedriger als für die beiden anderen Szenarien (Tabelle III, Tabelle IV, Tabelle V). Dies ließ den Schluss zu, dass dieses Szenario am schwierigsten war, was sich auch in der geringsten Linksschiefe mit nur marginaler Häufung im Zentrum widerspiegelte (Tabelle V). Entsprechend zeigte auch das Histogramm eine annähernd normale Verteilung (Abbildung VI).

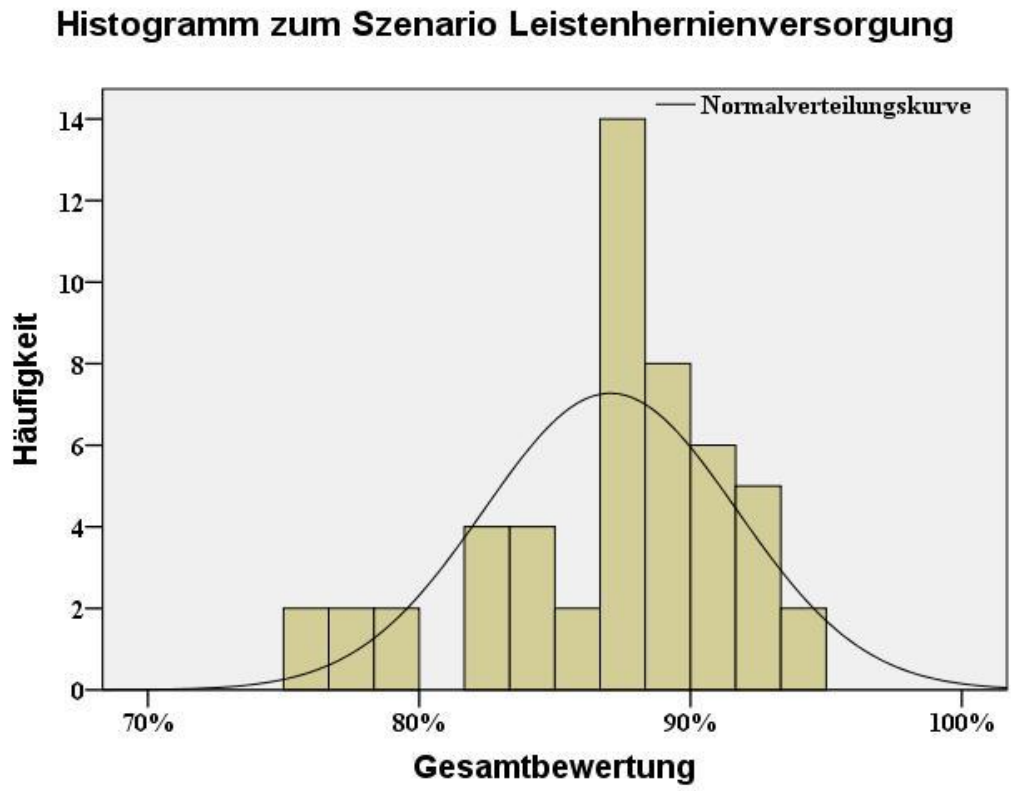


Abbildung VI: Histogramm zum Szenario Leistenhernienverschluss

Aufgetragen wurde im Histogramm zum Szenario Cholezystektomie die absolute Häufigkeit des Mittelwertes aus den Bewertungen beider Rater in Prozent. Die Normalverteilungskurve veranschaulicht dabei die Form einer theoretischen Normalverteilung.

4.1.5 Szenarienvergleich

Zunächst zeigte der Levene-Test, dass im Abschnitt B der Checklisten und in der Gesamtbewertung Varianzhomogenität herrschte. Lediglich in Abschnitt A waren Heterogenitäten in der Varianz nachweisbar (Tabelle VI). Wir führten mit dem Ziel, Unterschiede in der Bewertung der einzelnen Szenarien zu detektieren, weitere Rechenoperationen (einfache Varianzanalyse (ANOVA), Scheffé-Prozeduren mit Subgruppen-stratifizierung) für die beiden Abschnitte und die Gesamtbewertung durch.

Levene-Test		
	Statistik	**Signifikanz**
Abschnitt A	12,242	<0,05
Abschnitt B	0,553	0,576
Gesamt	0,992	0,373

Tabelle VI: Levene-Test

Die einfaktorielle ANOVA zeigte, dass es einen Unterschied in der Bewertung zwischen den einzelnen Szenarien gab (Tabelle VII).

Einfaktorielle ANOVA		
	F-Wert	**Signifikanz**
Abschnitt A	7,622	<0,05
Abschnitt B	9,154	<0,05
Gesamt	11,716	<0,05

Tabelle VII: Einfaktorielle ANOVA

In der Scheffé-Prozedur zeigten sich im Mehrfachvergleich deutliche Unterschiede in der Bewertung des Szenarios Appendektomie im Vergleich zu den beiden anderen Szenarien. So ergab sich in Abschnitt A mit einem relativen Bewertungsunterschied von 0,13% zwischen den Szenarien Leistenhernienverschluss und Cholezystektomie kein Anhalt für signifikante Bewertungsunterschiede. Hingegen wies das Szenario Appendektomie mit 2,05% gegenüber dem Szenario Cholezystektomie und mit 2,17% gegenüber dem Leistenhernienverschluss signifikante Bewertungsunterschiede auf (Tabelle VIII).

Gleiches galt für Abschnitt B. Hier wichen die Szenarien Cholezystektomie und Leistenhernienverschluss zwar mit 1,79% stärker voneinander ab. Aber diese Abweichung war

nicht signifikant. Das Szenario Appendektomie wich hingegen mit 2,89% signifikant vom Szenario Cholezystektomie und mit 4,68% ebenfalls signifikant vom Szenario Leistenhernien-verschluss ab (Tabelle VIII).

Scheffé-Prozedur (Mehrfachvergleich)					
	abhängige Variable		**Mittlere Differenz (I-J) [%]**	**Standard-fehler [%]**	**Signi-fikanz**
	I	J			
Abschnitt A	Appendektomie	Cholezystektomie	2,05	0,64	<0,05
		Leistenhernien-versorgung	2,17	0,63	<0,05
	Cholezystektomie	Appendektomie	-2,05	0,64	<0,05
		Leistenhernien-versorgung	0,13	0,65	0,982
	Leistenhernien-versorgung	Appendektomie	-2,17	0,63	<0,05
		Cholezystektomie	-0,13	0,65	0,982
Abschnitt B	Appendektomie	Cholezystektomie	2,89	1,13	<0,05
		Leistenhernien-versorgung	4,68	1,11	<0,05
	Cholezystektomie	Appendektomie	-2,89	1,13	<0,05
		Leistenhernien-versorgung	1,79	1,15	0,302
	Leistenhernien-versorgung	Appendektomie	-4,68	1,11	<0,05
		Cholezystektomie	-1,79	1,15	0,302
Gesamt	Appendektomie	Cholezystektomie	2,65	0,85	<0,05
		Leistenhernien-versorgung	3,97	0,84	<0,05
	Cholezystektomie	Appendektomie	-2,65	0,85	<0,05
		Leistenhernien-versorgung	1,32	0,87	0,321
	Leistenhernien-versorgung	Appendektomie	-3,97	0,84	<0,05
		Cholezystektomie	-1,32	0,87	0,321

Tabelle VIII: Scheffé-Prozedur (Mehrfachvergleich)

Für die Gesamtbewertung ergab sich das gleiche Bild. Die Szenarien Cholezystektomie und Leistenhernienverschluss wichen mit 1,32% nicht signifikant voneinander ab, während das Szenario Appendektomie mit 2,65% gegenüber dem Szenario Cholezystektomie und mit 3,97% gegenüber dem Szenario Leistenhernienverschluss signifikant differierte (Tabelle VIII).

Scheffé-Prozedur (Subgruppenstratifizierung)			
Abschnitt A			
Szenario	**N**	**Subgruppe**	
		1	**2**
Leistenhernien-versorgung	51	95,7%	
Cholezystektomie	48	95,8%	
Appendektomie	56		97,8%
Signifikanz		0,981	1,000
Abschnitt B			
Szenario	**N**	**Subgruppe**	
		1	**2**
Leistenhernien-versorgung	51	83,7%	
Cholezystektomie	48	85,5%	
Appendektomie	56		88,4%
Signifikanz		0,288	1,000
Gesamt			
Szenario	**N**	**Subgruppe**	
		1	**2**
Leistenhernien-versorgung	51	87,0%	
Cholezystektomie	48	88,4%	
Appendektomie	56		91,0%
Signifikanz		0,307	1,000

Tabelle IX: Scheffé-Prozedur (Subgruppenstratifizierung)

In der Subgruppenstratifizierung zeigte sich bei direktem Vergleich der Szenarien Leistenhernienverschluss und Cholezystektomie vs. Appendektomie, dass die Nullhypothese der Scheffé-Prozedur hinsichtlich der Existenz dieser Subgruppen nicht zu verwerfen war (Tabelle IX). Folglich waren die Szenarien Leistenhernienverschluss und Cholezystektomie einheitlicher bewertet, als das Szenario Appendektomie.

Die Boxplot-Darstellung bestätigte ebenfalls Subgruppen (Abbildung VII). Die vollständige Überlappung der T-Balken bezogen auf die y-Achse in den Szenarien Cholezystektomie und Leistenhernienverschluss zeigten deren Homogenität hinsichtlich der Bewertungen, während das Szenario Appendektomie nur teilweise Überschneidungen bot.

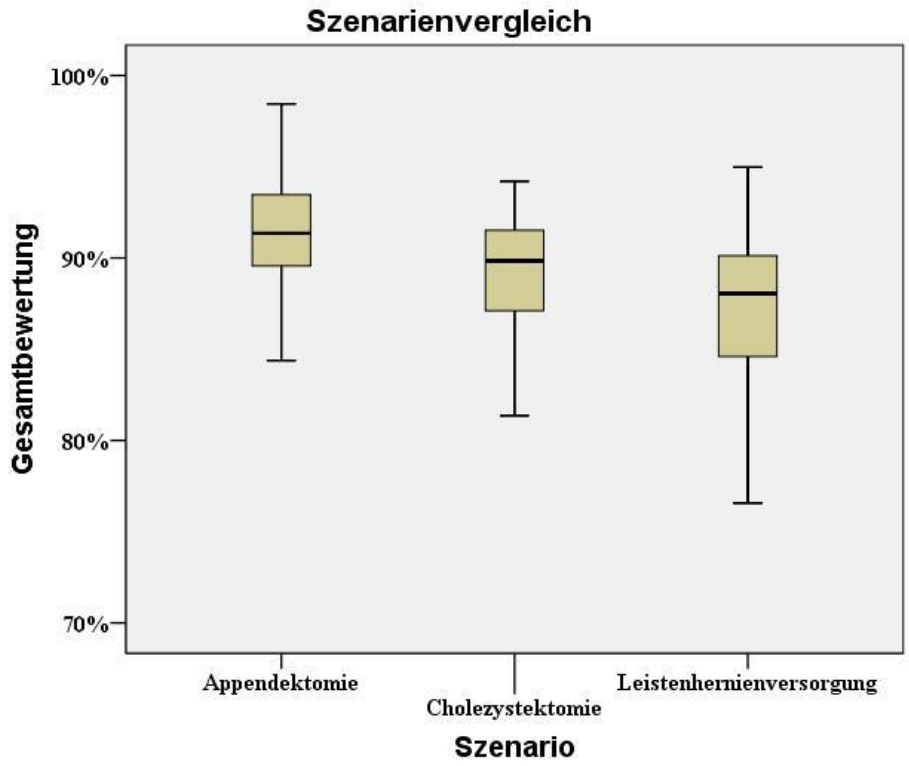


Abbildung VII: Szenarienvergleich

Aufgetragen wurde im Szenarienvergleich der Median der Bewertungen von beiden Ratern als Boxplot-Diagramm unterteilt nach Szenario. Dabei gibt die Box den einfachen Interquartilenabstand an. Die T-Whisker geben den zweifachen Interquartilenabstand an.

4.2 Checklistenevaluation

4.2.1 Interne Konsistenz (Cronbachs α)

Cronbachs α wurde für die drei Szenarien getrennt nach Abschnitt A, B und beide Abschnitte gemeinsam berechnet (Tabelle X). Es rangierte dabei zwischen Werten von 0,201 bis 0,384 für Abschnitt A, 0,583 bis 0,623 für Abschnitt B und 0,565 und 0,605 für die gemeinsame Betrachtung. Für die gemeine Betrachtung waren diese Werte als akzeptabel zu beschreiben, insofern man berücksichtigte, dass es sich hier um eine Einzelstations-OSCE handelte. Für weitere Details sei hier auf die Diskussion verwiesen.

Bestimmung von Cronbachs α				
	Items	**Appendektomie**	**Cholezystektomie**	**Leistenhernien-versorgung**
		Cronbachs α		
Abschnitt A	7	0,201	0,229	0,384
Abschnitt B	19	0,623	0,583	0,596
Gesamt	26	0,605	0,565	0,571

Tabelle X: Bestimmung von Cronbachs α

Die niedrigen Werte für den Teil A mussten vor dem Hintergrund interpretiert werden, dass in diesem Abschnitt sehr homogen bewertet wurde. Beim Item bezüglich des Raums für offene Fragen wurde sogar von beiden Ratern für alle Studenten die volle Punktzahl vergeben, womit volle Varianzhomogenität bei diesem Item herrschte. Für die Betrachtung aller Items dieses Abschnitts fielen aufgrund der so hohen Homogenität selbst die kleinsten Varianzen verstärkt ins Gewicht. Wir gingen daher von möglicherweise falsch niedrigen Werten für Cronbachs α aus. Zur Sicherstellung einer erhöhten Reliabilität in der nächsten Semesterkohorte wurden diese Items aufgrund des niedrigen Cronbachs α dennoch einer genaueren Betrachtung und Überarbeitung zugeführt. Dies war jedoch nicht mehr Bestandteil der vorliegenden Dissertation.

4.2.2 Itemschwierigkeit

Itemstatistiken für das Szenario Appendektomie Abschnitt A			
Item	**Item-schwierigkeit**	**Trenn-schärfe**	**Cronbachs α, wenn Item weggelassen**
Patientenbegrüßung	0,96	0,133	0,138
Name und Funktion genannt	0,94	-0,008	0,331
Respekt ggü. dem Patienten gezeigt	0,99	0,183	0,161
Sprache angemessen	0,99	0,235	0,128
Logische Reihenfolge	0,98	-0,005	0,236
Ausstrahlung von Sicherheit	0,96	0,174	0,075
Raum für offene Frage	1,00	n.b.	n.b.
Mittelwerte aller Items	0,97	0,119	

Tabelle XI: Itemstatistiken für das Szenario Appendektomie Abschnitt A

Mit einer durchschnittlichen Schwierigkeit von 0,96 waren die Items des Abschnitts A sehr leicht. Dabei lag die durchschnittliche Schwierigkeit im Szenario Appendektomie mit 0,97 leicht über den anderen beiden Szenarien mit je 0,95 (Tabelle XI, Tabelle XII, Tabelle XIII). Insbesondere das Item „Raum für offene Fragen“ wurde bei allen Studenten mit voller Punktzahl bewertet (→ Itemschwierigkeit = 1,00). Daher konnte aufgrund der fehlenden Varianz keine Trennschärfe für dieses Item berechnet werden.

Itemstatistiken für das Szenario Cholezystektomie Abschnitt A			
Item	**Item-schwierigkeit**	**Trenn-schärfe**	**Cronbachs α, wenn Item weggelassen**
Patientenbegrüßung	0,97	0,057	0,233
Name und Funktion genannt	0,88	-0,093	0,429
Respekt ggü. dem Patienten gezeigt	0,99	0,021	0,244
Sprache angemessen	0,97	0,315	0,098
Logische Reihenfolge	0,92	0,156	0,146
Ausstrahlung von Sicherheit	0,90	0,288	-0,002
Raum für offene Frage	1,00	n.b.	n.b.
Mittelwerte aller Items	0,95	0,124	

Tabelle XII: Itemstatistiken für das Szenario Cholezystektomie Abschnitt A

Itemstatistiken für das Szenario Leistenhernienverschluss Abschnitt A			
Item	**Item-schwierigkeit**	**Trenn-schärfe**	**Cronbachs α, wenn Item weggelassen**
Patientenbegrüßung	0,96	0,081	0,433
Name und Funktion genannt	0,82	0,330	0,259
Respekt ggü. dem Patienten gezeigt	1,00	n.b.	n.b.
Sprache angemessen	0,96	0,147	0,400
Logische Reihenfolge	0,97	0,262	0,342
Ausstrahlung von Sicherheit	0,92	0,291	0,292
Raum für offene Frage	1,00	n.b.	n.b.
Mittelwerte aller Items	0,95	0,222	

Tabelle XIII: Itemstatistiken für das Szenario Leistenhernienverschluss Abschnitt A

Im Abschnitt B zeigten sich die Werte stärker verteilt. Die Spannbreite lag hier zwischen 0,14 und 1,00 (Tabelle XIV, Tabelle XV, Tabelle XVI).

Im oben definierten Bereich von 0,2 bis 0,8 für Items (Kelava und Moosbrugger 2012) lagen im Szenario Appendektomie 42% der Items und waren somit geeignet für die Differenzierung zwischen guten und schlechten Studenten. Ein Item wies einen Wert von <0,2 und war daher sehr schwierig. Die restlichen Werte lagen über 0,8 und waren sehr leicht (Tabelle XIV).

Itemstatistiken für das Szenario Appendektomie Abschnitt B			
Item	**Item-schwierigkeit**	**Trenn-schärfe**	**Cronbachs α, wenn Item weggelassen**
Operationsindikation	0,96	0,249	0,607
Operationsverfahren	0,90	0,209	0,616
Umstieg auf offenes Verfahren	0,96	0,119	0,621
Ausdehnung der Operation	0,45	0,181	0,618
Lagerungsschäden	0,27	0,321	0,596
Thrombose/Embolie	0,86	0,238	0,609
Blutung	0,98	0,082	0,623
Infektion	0,98	0,385	0,610
Verletzung von Gefäßen und Nerven	0,93	0,204	0,614
Narbenbildung	0,84	0,362	0,593
Verwachsungen/Bridenileus	0,80	0,336	0,595
Nahtbruch/Narbenbruch	0,86	0,595	0,566
Verletzung von Gefäßen beim Setzen der Trokare	0,45	0,328	0,594
Verletzung von Nachbarorganen	0,91	0,325	0,602
Hautemphysem	0,82	0,091	0,627
Schmerzen im Schulterbereich	0,61	0,122	0,627
Pneumothorax	0,43	0,122	0,628
Zügiger Kostaufbau und Mobilisation	0,70	0,149	0,622
Keine Einschränkung in der Ernährung	0,52	0,142	0,625
Mittelwert aller Items	0,75	0,240	

Tabelle XIV: Itemstatistiken für das Szenario Appendektomie Abschnitt B

Für das Szenario Cholezystektomie lagen 53% der Werte zwischen 0,2 und 0,8 hinsichtlich der Itemschwierigkeit. Zu schwierige Items gab es hier nicht. Gegenüber dem Szenario Appendektomie ließ sich somit eine etwas bessere Unterscheidung zwischen guten und schlechten Prüfungsteilnehmern vornehmen (Tabelle XIV, Tabelle XV).

Itemstatistiken für das Szenario Cholezystektomie Abschnitt B			
Item	**Item-schwierigkeit**	**Trenn-schärfe**	**Cronbachs *α*, wenn Item weggelassen**
Operationsindikation	0,94	0,184	0,573
Operationsverfahren	0,87	0,189	0,575
Umstieg auf offenes Verfahren	0,71	0,253	0,560
Ausdehnung der Operation	0,25	-0,117	0,616
Lagerungsschäden	0,17	0,038	0,590
Thrombose/Embolie	0,88	0,275	0,561
Blutung	0,98	0,071	0,583
Infektion	0,94	0,391	0,556
Verletzung von Gefäßen und Nerven	0,94	0,035	0,586
Narbenbildung	0,73	0,298	0,553
Verwachsungen/Bridenileus	0,44	0,303	0,551
Nahtbruch/Narbenbruch	0,65	0,291	0,553
Verletzung von Nachbarorganen	0,94	0,130	0,578
Verletzung der Gallenwege	0,85	0,272	0,560
Hautemphysem	0,52	0,277	0,555
Schmerzen im Schulterbereich	0,44	0,185	0,572
Pneumothorax	0,38	0,483	0,518
Zügiger Kostaufbau und Mobilisation	0,67	0,104	0,585
Vermeidung fettreicher, opulenter Speisen	0,79	0,157	0,575
Mittelwert aller Items	0,69	0,201	

Tabelle XV: Itemstatistiken für das Szenario Cholezystektomie Abschnitt B

Im Szenario Leistenhernienverschluss lagen 53% der Werte in dem empfohlenen Rahmen der Schwierigkeit zwischen 0,2 und 0,8 (Kelava und Moosbrugger 2012). Wie im Szenario Appendektomie war ein Item sehr schwierig. Die Szenarien Leistenhernienverschluss und Cholezystektomie wiesen somit die gleiche Anzahl von Items auf, die eine Unterscheidung zwischen guten und schlechten Studenten zuließen (Tabelle XIV, Tabelle XV, Tabelle XVI).

Itemstatistiken für das Szenario Leistenhernienverschluss Abschnitt B			
Item	**Item-schwierigkeit**	**Trenn-schärfe**	**Cronbachs α, wenn Item weggelassen**
Operationsindikation	0,89	0,297	0,576
Operationsverfahren	0,89	0,068	0,622
Eröffnung der Bauchhöhle	0,31	0,314	0,569
Ausdehnung der Operation	0,18	0,250	0,580
Lagerungsschäden	0,14	0,125	0,595
Thrombose/Embolie	0,73	0,383	0,559
Blutung	0,97	0,068	0,598
Infektion	0,90	0,432	0,565
Verletzung von Gefäßen und Nerven	1,00	n. b.	n. b.
Narbenbildung	0,65	0,215	0,584
Verwachsungen/Bridenileus	0,59	0,080	0,604
Nahtbruch/Narbenbruch	0,69	0,435	0,550
Verletzung von Haut-/Muskelnerven und Gefäßen	0,96	0,035	0,601
Verletzung von Nachbarorganen	0,76	0,251	0,579
Unverträglichkeit des Kunststoffnetzes	0,80	0,507	0,546
Bewegungseinschränkung/chron. Schmerzen	0,51	0,268	0,575
Rezidiv/Dislokation des Netzes	0,69	-0,113	0,630
Zügiger Kostaufbau und Mobilisation	0,59	0,293	0,571
Keine Einschränkung in der Ernährung	0,14	0,125	0,595
Mittelwerte aller Items	0,65	0,224	

Tabelle XVI: Itemstatistiken für das Szenario Leistenhernienverschluss Abschnitt B

Im Vergleich mit dem Szenario Appendektomie mit einer durchschnittlichen Schwierigkeit des Checklistenabschnitts B von 0,75 zeigten sich die beiden anderen Szenarien mit 0,69 für die Cholezystektomie und 0,65 für den Leistenhernienverschluss schwieriger (Tabelle XIV, Tabelle XV, Tabelle XVI).

4.2.3 Trennschärfe

Für die Bewertung der Trennschärfen sei hier auf Tabelle I verwiesen. Für den Abschnitt A im Szenario Appendektomie zeigten sich 43% der Items als nicht berechenbar bei fehlender Varianz oder <0,100 und somit als schlecht. Sie sollten daher verworfen werden. 43% zeigten eine moderate Trennschärfe und 14% waren gute Items (Tabelle XI). Im Szenario Cholezystektomie in Abschnitt A waren 57% der Items schlecht. 14% waren moderat und 29% zeigten gute Trennschärfen (Tabelle XII). Im Szenario Leistenhernienverschluss waren 43% der Trennschärfen nicht berechenbar oder <0,100, 14% wiesen moderate Trennschärfen auf und 43% waren gut (Tabelle XIII).

In Abschnitt B wiesen im Szenario Appendektomie 11% der Items schlechte Trennschärfen auf, 32% der Items waren moderat und 57% gut (Tabelle XIV). Im Szenario Cholezystektomie waren 21% der Trennschärfen auf einem schlechten Niveau, 32% auf moderatem und 47% auf gutem Niveau (Tabelle XV). Im Szenario Leistenhernienverschluss waren 32% der Trennschärfen schlecht oder nicht berechenbar, 11% der Items wiesen moderate Trennschärfen auf und 57% gute Trennschärfen (Tabelle XVI).

Kombiniert man nun die Erkenntnisse aus Itemschwierigkeiten und Trennschärfen (Tabelle XI, Tabelle XII, Tabelle XIII), so zeigte sich für Teil A insgesamt eine zu leichte Aufgabenstellung mit geringer Trennschärfe. Insbesondere das Item „Raum für offene Fragen“ war in allen drei Szenarien für alle Studenten mit der vollen Punktzahl bewertet worden. Ebenso war das Item „Respekt gegenüber dem Patienten gezeigt“ zu leicht und wies ebenfalls keine gute Trennschärfe auf. Hingegen besaß das Item: „Ausstrahlung von Sicherheit“ im Vergleich eine höhere Schwierigkeit und konnte zumindest in den Szenarien Cholezystektomie und Leistenhernienverschluss gute Trennschärfen erzeugen.

Für Abschnitt B (Tabelle XIV, Tabelle XV, Tabelle XVI) zeigte sich das Item „Operationsindikation“ über alle Szenarien zwar etwas zu leicht (Itemschwierigkeit >0,80), aber von moderater bis guter Qualität im Bereich der Trennschärfe. Hinsichtlich beider Gütekriterien konnten die Items „Narbenbildung“ und „Nahtbruch/Narbenbruch“ über alle Szenarien hinweg

gute Itemschwierigkeiten mit lediglich kleinen Ausreißern im Szenario Appendektomie und gute Trennschärfen aufweisen. Somit lagen insgesamt für das Szenario Appendektomie 89% der Items im Bereich moderater bis guter Trennschärfen. Im Szenario Cholezystektomie waren es 78% und im Szenario Leistenhernienverschluss 68%.

Für die einzelnen Szenarien betrachtet, trugen für das Szenario Appendektomie acht Items („Ausdehnung der Operation“, „Lagerungsschäden“, „Verwachsungen/Bridenileus“, „Verletzung von Gefäßen beim Setzen der Trokare“, „Schmerzen im Schulterbereich“, „Pneumothorax“, „Zügiger Kostaufbau und Mobilisation“ & „Keine Einschränkung in der Ernährung“) mit gutem Niveau hinsichtlich der Kombination von Trennschärfe und Itemschwierigkeit zu der hohen Qualität hinsichtlich eines Cronbachs α von 0,605 bei (Tabelle XI, Tabelle XIV). Genau zu überprüfen, waren in dieser Checkliste – neben den oben bereits genannten – die Items „Name und Funktion genannt“ und „Logische Reihenfolge“, weil diese eine negative Trennschärfe und zu große Leichtigkeit aufwiesen.

Im Szenario Cholezystektomie konnten neun Items („Umstieg auf offenes Verfahren“, „Narbenbildung“, „Verwachsungen/Bridenileus“, „Nahtbruch/Narbenbruch“, „Hautemphysem“, „Schmerzen im Schulterbereich“, „Pneumothorax“, „Zügiger Kostaufbau und Mobilisation“ & „Vermeidung fettreicher, opulenter Speisen“) zu einer hohen Qualität mit einem Cronbachs α von 0,565 beitragen (Tabelle XII, Tabelle XV). In diesem Szenario mussten – neben den oben bereits genannten – die Items „Name und Funktion genannt“ und „Ausdehnung der Operation“ vom Prüfungsverantwortlichen inhaltlich überprüft werden, da alle negative Trennschärfen aufwiesen. Dabei zeigte sich das erste Item als zu leicht (Itemschwierigkeit >0,80), das zweite als gerade noch akzeptabel mit einer Itemschwierigkeit von 0,25.

Im Szenario Leistenhernienverschluss waren es acht Items („Eröffnung der Bauchhöhle“, „Thrombose/Embolie“, „Narbenbildung“, „Nahtbruch/Narbenbruch“, „Verletzung von Nachbarorganen“, „Unverträglichkeit des Kunststoffnetzes“, „Bewegungseinschränkungen/ chronische Schmerzen“ & „Zügiger Kostaufbau und Mobilisation“), die zu einem guten Cronbachs α von 0,571 beitrugen (Tabelle XIII, Tabelle XVI). Auf den Prüfstand mussten hier – neben den oben bereits genannten – die Items „Verletzungen von Gefäßen und Nerven“ und „Rezidiv/Dislokation des Netzes“.

4.3 Interrater-Reliabilität

Interrater-Reliabilität								
Szenario	**Mittelwertabweichung**			**Intraklassen-Korrelation**			**Korrelation nach Pearson**	
	Total [%]	**Konfidenz-intervall**		**ICC**	**Konfidenz-intervall**		**ρ**	**Signifikanz**
		untere Grenze	**obere Grenze**		**untere Grenze**	**obere Grenze**		
Appendektomie	3,63	2,98	4,28	0,83	0,73	0,90	0,84	<0,05
Cholezystektomie	1,49	0,47	2,50	0,73	0,56	0,84	0,76	<0,05
Leistenhernien-versorgung	1,63	0,65	2,60	0,76	0,61	0,85	0,77	<0,05
Gesamtbewertung	2,31	1,79	2,83	0,78	0,71	0,83	0,80	<0,05

Tabelle XVII: Interrater-Reliabilität

Verglich man die beiden Rater miteinander, so lagen die absoluten Abweichungen im Mittelwertvergleich für die einzelnen Szenarien zwischen 1,49% und 3,63% (Tabelle XVII). Dabei wichen die Rater beim absoluten Mittelwertvergleich am stärksten im Szenario Appendektomie voneinander ab. Im Szenario Cholezystektomie bewerteten sie am kongruentesten. Es zeigte sich trotz der höchsten absoluten Mittelwertabweichung von 3,63% im Szenario Appendektomie mit einem ICC von 0,83 (= 83% Übereinstimmung) und einem ρ von 0,84 (= 84% Übereinstimmung) die höchste prozentuale Übereinstimmung zwischen den Ratern. Der umgekehrte Fall zeigte sich im Szenario Cholezystektomie, welche beim absoluten Mittelwertvergleich nur einen Unterschied von 1,49% verzeichnete. Hingegen war die prozentuale Übereinstimmungsrate von 73% im ICC und 76% in der Korrelation nach Pearson am geringsten. Das Szenario Leistenhernienverschluss platzierte sich für alle Betrachtungsweisen mit einer absoluten Mittelwertabweichung von 1,63% und einem ICC von 0,76 und einem ρ von 0,77 zwischen den beiden anderen Szenarien (Tabelle XVII).

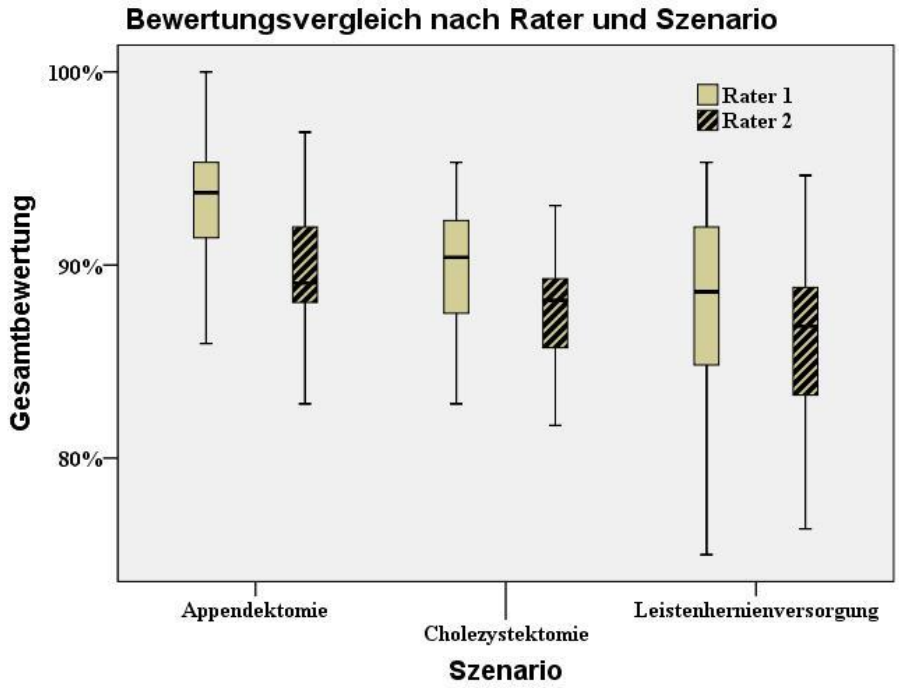


Abbildung VIII: Bewertungsvergleich nach Rater und Szenario

Aufgetragen wurde im Bewertungsvergleich nach Rater und Szenario als Boxplot-Diagramm der Mittelwert der Bewertungen der beiden Rater unterteilt nach Szenario als Median. Dabei gibt die Box den einfachen Interquartilenabstand an. Die T-Whisker den zweifachen Interquartilenabstand.

In der Gesamtbewertung lagen die Rater mit 2,31% auseinander und korrelierten mit einem ICC von 0,70 und einem ρ von 0,80 sehr hoch (Tabelle XVII).

Ersichtlich wurde dies auch aus den graphischen Auftragungen der Abbildungen VIII und IX. In Abbildung VIII überlappten sich die T-Balken in weiten Anteilen als Indiz hoher Übereinstimmung. In Abbildung IX veranschaulichte die Regressionsgerade, dass niedrige Bewertungen von Rater 1 niedrigen Bewertungen von Rater 2 entsprachen und umgekehrt.

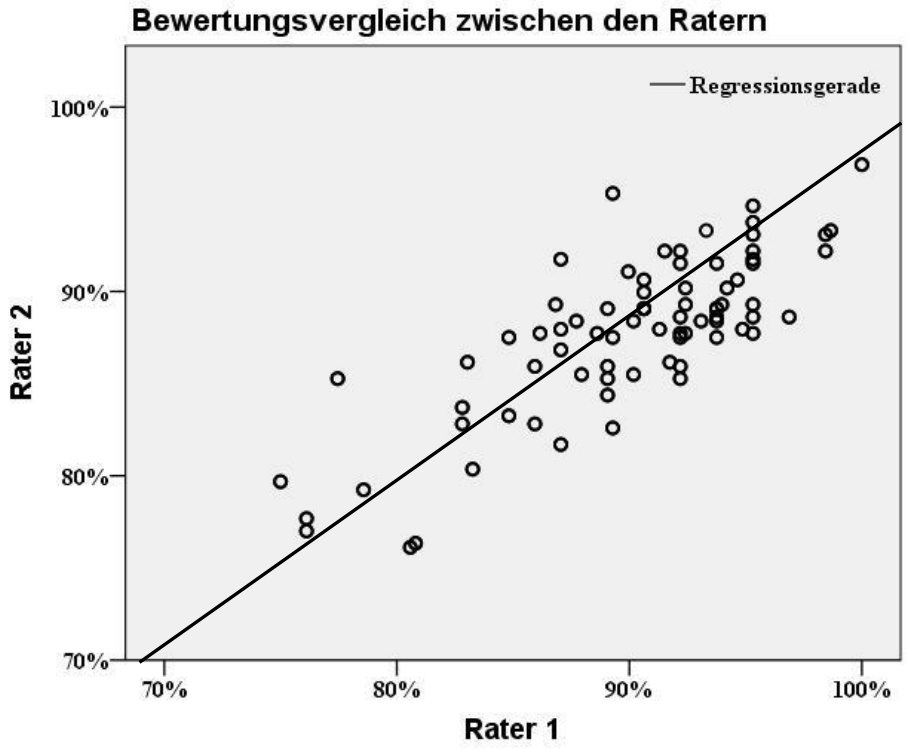


Abbildung IX: Rater-Vergleich

Aufgetragen wurden hier die einzelnen Gesamtbewertungen von Rater 1 gegen die entsprechenden Bewertungen von Rater 2 bei demselben Studenten. Die Regressionsgerade veranschaulicht eine lineare Steigung zwischen 0% und 100%.

4.4 Kalkulation der Bewertungszeit

Die Bewertung der Videos mit der studentischen Performance wurde von den zwei Ratern in fünf Sitzungen in bezahlten Überstunden durchgeführt. Der normale Ablauf einer solchen Sitzung bestand aus der Bewertung von acht bis zehn Videos in Folge und Integration einer kurzen Pause von ca. fünf bis zehn Minuten nach der Hälfte der Videos.

Die zwei Studenten nahmen während ihrer 30-minütigen Zeitspanne mit dem Schauspielpatienten je ein Video von ungefähr zehn Minuten Dauer auf. Dabei wurde ein Video ausschließlich zur Selbstreflektion verwendet, das andere zur Bewertung abgegeben. Somit entstanden in der gesamten Semesterkohorte 78 Videos von durchschnittlich zehn Minuten Dauer zur Bewertung. Inklusive Bewertung und Ausfüllen der Checklisten wurde eine Zeit von zwölf Minuten pro Video veranschlagt. Zum Bewerten der gesamten Semesterkohorte standen also 78 Mal zwölf Minuten an, was einer Gesamtzeit von 936 Minuten – also ca. 16 Stunden – pro Rater entsprach. Aus den oben genannten 5 Sitzungen ergaben sich also ca. drei Stunden pro Sitzung und Rater.

5. Diskussion

Die detaillierte Analyse der Bewertung von Prüfungsleistungen ist ein lebendiger und aufwändiger Prozess. Hohe Qualitätsmaßstäbe sollten angesetzt werden, um das bestehende Prüfungssystem durch qualitativ hochwertige und vor allem durchführbare Prüfungen zu ergänzen bzw. Teile daraus zu ersetzen.

Folglich sollten Strategien entwickelt werden, wie objektive, valide und reliable Prüfungen implementiert werden können. Ein Weg dahin kann die Standardisierung solcher Prüfungen sein, wie er mit den OSCEs bereits beschritten wurde. Dazu gehören vorab im Detail ausgearbeitete Checklisten und geschulte Rater, welche zu der hohen Qualität der OSCEs beitragen können (Khan et al. 2013, Pell et al. 2010).

Eine Ergänzung für diese hohen Qualitätsansprüche kann nun die Videoaufzeichnung darstellen. Mit ihr ist es möglich, die Prüfungssituation immer wieder Revue passieren zu lassen. Gelangt ein Rater bei der Durchsicht zu einem Item, welches er bislang nicht bewertet hat, ist es dem Prüfer möglich, das Video unter diesem Aspekt noch einmal durchzusehen. Insgesamt kann dies zu einer gerechteren und reliableren Prüfung beitragen. Entsprechende Untersuchungen zwischen Ratern vor Ort und Video-Ratern wurden bereits hinsichtlich der Interrater-Reliabilität durchgeführt (Vivekananda-Schmidt et al. 2007) und wurden durch die Ergebnisse dieser Dissertation hinsichtlich der Übereinstimmung von verschiedenen Video-Ratern ergänzt.

5.1 Etablierung eines reliablen und qualitativ hochwertigen Bewertungssystems für die Rater

In Bezug auf die technische Qualität der Bewertungsinstrumente zeigten alle drei Checklisten bei der internen Konsistenzprüfung gute Ergebnisse. Ebenso waren die Checklisten von einer im Durchschnitt guten Itemschwierigkeit und moderaten bis guten Trennschärfe geprägt. In der Gesamtbetrachtung ließ sich dies als Indiz für die hohe Qualität der drei Checklisten deuten. Anders ausgedrückt, zeugten sie durchschnittlich von einem angemessenen Schwierigkeitsgrad und gaben den Prüfern die Möglichkeit, gute von schlechten Studenten zu unterscheiden.

Unterschied man dabei nach Abschnitt A und Abschnitt B der Checklisten, musste man jedoch feststellen, dass der Abschnitt A insgesamt zu leicht war. Keines der Items konnte einen Schwierigkeitsgrad von <0,80 vorweisen. Die Trennschärfen hingegen ließen in der Mehrheit

noch eine moderate bis gute Unterscheidung zwischen guten und schlechten Studenten zu. Dieser Umstand wies somit auf eine zu leichte Prüfung hin. Möglichweise war dies jedoch auch Ausdruck der hohen Güte der bislang etablierten Lehre hinsichtlich der kommunikativen Fertigkeiten. Einschränkend zu diesem Punkt musste aber angemerkt werden, dass zur Bewertung immer die Auswahl zwischen zwei Videos bestand. Man könnte somit unterstellen, dass nur die sprachlich flüssigen und inhaltlich gelungenen Videos zum Testat abgegeben wurden. Hier könnte sich eine weitere Forschungsfrage anschließen, in welcher die nicht zum Testat abgegebenen Videos mit den testierten Videos verglichen werden.

Abschnitt B hingegen konnte durch eine angemessene Streuung von Schwierigkeitsgraden und Trennschärfen bei den einzelnen Items überzeugen. Aber auch hier gab es einige Items mit schlechten Qualitätsmerkmalen. Betrachtete man nur die schwierigeren Items mit einer Itemschwierigkeit von <0,40 und schlechten bis moderaten Trennschärfen von <0,200, so fanden sich hier zum Beispiel die Items „Lagerungsschäden“ in den Szenarien Cholezystektomie und Leistenhernienverschluss, „Ausdehnung der Operation“ im Szenario Cholezystektomie und „Keine Einschränkung in der Ernährung“ im Szenario Leistenhernienverschluss. So wie bei den Items aus Abschnitt A die Lehre bezüglich der Kommunikation außerordentlich gut gewesen sein könnte, könnte hier im Gegensatz dazu eine unterrepräsentierte Seite der Krankheitsbilder in der Lehre zu Tage getreten sein. Alle diese Items beschäftigten sich mit Themen des chirurgischen Alltags, welche von den Lehrkörpern möglicherweise als so selbstverständlich angesehen wurden, dass sie in der Lehre nicht ausreichend für die Studierenden angesprochen bzw. vermittelt wurden. Den testatisch vermeintlich schlechten Parametern entsprechend, sollte die universitäre Lehre neu fokussiert werden und zukünftig die damit verbundenen Unterrichtsinhalte verstärkt adressiert werden. Nach dem Durchlauf weiterer Semesterkohorten sollten diese Items dann einer erneuten Betrachtung unterzogen werden, ob durch die Verbesserung der Lehre ein Erfolg hinsichtlich der Machbarkeit dieser Items erzielt werden konnte.

In Bezug auf die Gesamtreliabilität, repräsentiert durch Cronbachs α zwischen 0,565 bis 0,605 für die drei Szenarien, könnte man dem Einspruch stattgeben, dass angestrebte Werte von >0,700 nicht erreicht wurden. Die Besonderheit hier war allerdings, dass es sich um eine Pilot-Station für eine noch zu etablierende Gesamt-OSCE als Vorbereitung auf das PJ handelte. Als Einzelstations-OSCE lagen die Werte im Rahmen des zu erwartenden und akzeptablen Wertebereichs, wenn man diese mit z. B. der Studie aus der Forschungsgruppe um Tudiver und Kollegen (Tudiver et al. 2009) verglich. Hier wurden für eine Einzelstation innerhalb einer

OSCE Werte von α = 0,580 gemessen und als akzeptabel bewertet. In der Literatur hatte sich allerdings für die Bewertung einer klassischen OSCE mit mehreren Stationen mindestens ein α von 0,6 (Vivekanda-Schmidt et al. 2007) oder besser noch α >0,7 durchgesetzt. Um eine Voraussage treffen zu können, wie hoch nun das rechnerische Cronbachs α bei Kombinationen mehrerer Einzelstationen der V-OSCE wäre, soll hier ein Rechenbeispiel folgen. Betrachtet man die Szenarien getrennt als einzelne Stationen einer OSCE, dann mitteln wir zunächst die einzelnen Cronbachs α von 0,565, 0,571 und 0,605 auf einen Einheitswert von 0,580. Mit diesen Werten kann man nach der Spearman-Brown-Formel die voraussichtliche Reliabilität r_m einer OSCE mit m Stationen, ausgehend von einer OSCE mit n Stationen und einer Reliabilität r_n, berechnen (Möltner et al. 2006):

$$r_m = \frac{m\, r_n}{n + (m - n) r_n}$$

Dabei spiegeln die Werte r_m und r_n das Cronbachs α wider. Somit ergibt sich hier für r_n = 0,580 bei n = 1 und m = 3 ein voraussichtliches Cronbachs α von 0,806 bei r_m = 0,806. Dies interpretierten wir als weiteren Beleg für die hohe Qualität der Checklisten der V-OSCE vor dem Hintergrund, dass eine Kombination dieser Einzelstation in einen OSCE-Parcours mit nur drei Stationen bereits den Ansprüchen bezüglich eines Cronbachs α von >0,700 genügen würde. Legte man eine OSCE mit zehn Stationen zu Grunde, ließe sich bereits auf Grundlage unseres Einheitswertes von 0,580 ein Cronbachs α von 0,932 vorhersagen.

Betrachtet man nun die Interrater-Reliabilität mit einem ICC von 0,78 und einem ρ von 0,80, so lagen wir hier auf einem guten Niveau – insbesondere vor dem Hintergrund, dass die durchschnittliche absolute Abweichung lediglich bei 2,31% lag. In vergleichbarer Literatur wurden ICC Werte zwischen 0,7 für eine OSCE mit Ultraschalluntersuchung am muskulo-skelettalen Apparat des Körpers (Kissin et al. 2014) und 0,96 bei einer OSCE zum Thema „Evidenz-basierte Medizin“ (Tudiver et al. 2009) gemessen. Dies waren allerdings Arbeiten zu einer klassischen OSCE mit Live-Ratern. Die in der Einleitung bereits erwähnte Arbeit der Arbeitsgruppe von Vivekananda-Schmidt verglich zwar Rater von Videobewertungen, allerdings nicht untereinander, sondern mit Ratern vor Ort. In diesem Fall konnten nur weitaus geringere Übereinstimmungsraten bei ICC-Werten zwischen 0,32 bis 0,58 und Korrelationen nach Pearson von 0,46 bis 0,66 bestimmt werden (Vivekananda-Schmidt et al. 2007). Ebenso lagen hier auch die mittleren absoluten Abweichungen der Bewertungen durch die Rater zwischen 7,0% und 17,8% und damit deutlich höher als in der vorliegenden Dissertation. Eine andere Arbeit von Chen und Kollegen bestimmte andere Parameter zur Überprüfung des

Übereinstimmungsgrades ihrer Rater. Diese waren zwar hoch, aber ein direkter Vergleich der Werte mit den hier erhobenen Werten konnte aufgrund der anderen Parameter und Rechenoperationen nicht durchgeführt werden (Chen et al. 2013).

Insgesamt rangierten unsere Übereinstimmungswerte auf einem hohen Niveau. Dabei war noch zu berücksichtigen, dass wir sogar einen fachfremden Rater herangezogen hatten. Wir konnten also nachweisen, dass durch die Verwendung der standardisierten Prüfungsszenarien und der im Vorfeld detailliert ausgearbeiteten Checklisten zusammen mit eingehenden Schulungen der Prüfer eine hohe Interrater-Reliabilität erzielt werden konnte.

Hinsichtlich der Validität dieser Prüfung wurden keine genaueren Überprüfungen vorgenommen. Gleichwohl wurde die Augenschein- und Inhaltsvalidität dadurch sichergestellt, dass die klinischen Szenarien, Schauspielerrollen und Checklisten von einer Fachexpertin (in der Lehre erfahrene, chirurgische Oberärztin) erstellt und von weiteren Fachärzten in einem Review-Verfahren gegengeprüft und überarbeitet wurden. Zur Konstruktvalidität ließ sich hingegen keine Aussage treffen, da ein Vergleich der studentischen Performance zu verschiedenen Zeitpunkten der Ausbildung nicht vorgesehen war und auch kein weiteres Prüfungsformat eingesetzt wurde.

5.2 Vorzüge der Video-Aufzeichnung

Die Implementierung einer Video-gestützten OSCE (V-OSCE) hatte sowohl für die Studiengangskoordination als auch die Lehrbeauftragten und nicht zuletzt für die Teilnehmer an der Prüfung mannigfaltige potentielle Vorzüge (Casey et al. 2009). Eine klassische OSCE geht mit einer anspruchsvollen Organisations- und Koordinationsarbeit einher, mit dem Ziel sicherzustellen, dass sich Studenten, Schauspieler und Rater zur gleichen Zeit am selben Ort einfinden. Insbesondere in Hinblick auf die Rater könnte dies eine große Herausforderung darstellen, da sie für die Prüfung aus dem klinischen Alltag mit Operationen, Stationsdienst, Funktionsdiensten und Forschung freigestellt werden müssen. Dabei sollten diese bestenfalls auch noch Erfahrungen auf dem Gebiet der Prüfungsbewertung haben bzw. geschult worden sein. Gerade in Hinblick auf den unterschiedlichen Erfahrungsstand der Rater im klinischen Alltag und in der Lehre könnte dies zu einem Problem für eine reliable Prüfung werden (Vivekananda-Schmidt et al. 2007). In diesem Zusammenhang konnte die hier verwendete Videoaufzeichnung zu einer Erleichterung des administrativen Ablaufs und zur Erhöhung der Bewertungsqualität beitragen.

Die Video-Aufnahmen der Prüfung ermöglichten die zeitlich versetzte Bewertung z. B. in bezahlten Überstunden bzw. außerhalb der Kernarbeitszeit eines Arztes. Dabei war diese Möglichkeit allerdings mit einer für den entsprechenden Rater erhöhten Arbeitsbelastung verbunden und somit der Work-Life-Balance nicht zuträglich (Kastner 2010). Es sollte somit keine generelle Lösung für eine gute Lehre sein, die Arbeitszeit auf diesem Gebiet in den außerdienstlichen Bereich zu verlagern. Gleichwohl erschien die Durchführung der Bewertungen im Modul 6.1 „Operative Medizin“ innerhalb bezahlter Überstunden als vorerst einzige praktikable Möglichkeit, eine klinische Prüfung für eine gesamte Semesterkohorte von derzeit 155 Studenten adäquat zu etablieren. Somit war es möglich, die Zahl der benötigten Rater auf nur 2 zu beschränken und somit die Anzahl der für den Regeldienst zur Verfügung stehenden Ärzte hoch zu halten. Weitere Lösungsansätze könnten sich hieraus dennoch ergeben, da die Personalstruktur durch den geringeren Personalaufwand angeglichen werden könnte. Hieraus könnten sich ebenfalls neue Forschungsfragen auf dem Gebiet der Personalentwicklung ableiten.

Auch aus der Sicht des Raters boten sich einige Vorteile. So konnte z. B. die Müdigkeit und Erschöpfung während einer OSCE-Bewertung in einer selbstbestimmten Umgebung reduziert oder auch vollständig vorgebeugt werden (McLaughlin et al. 2009). Im Rahmen der Video-Beurteilung bedeutete dies, dass Pausen und auch längere Unterbrechungen jederzeit in Abhängigkeit von den persönlichen Bedürfnissen eingelegt werden konnten. Als positiver Nebeneffekt konnte daraus eine potentiell erhöhte Bewertungskonsistenz erwachsen (Vivekananda-Schmidt et al. 2007).

Aus Sicht der Studenten ergab sich mit der Video-Aufzeichnung zusätzlich die Möglichkeit der Archivierung ihrer Leistungen in einem elektronischen Portfolio (Gómez et al. 2013). Hieraus ergaben sich Möglichkeiten zur Selbstreflektion und einem Feedback durch andere Studenten oder Lehrkräfte, wodurch Wissenslücken aufgedeckt werden konnten.

Um die benötigte Zeit für die Bewertungen noch weiter zu reduzieren, wurde nach Abschluss der Forschungsarbeiten zu dieser Dissertation eine weitere Pilotstudie durchgeführt. Dazu wurden die Rater gebeten, eine Woche nach Abschluss der kompletten Bewertungen sich pro Szenario vier Videos – insgesamt also zwölf Videos – noch einmal mit der 1,2-fachen Geschwindigkeit anzusehen und erneut zu bewerten. Die vorläufigen Daten ließen darauf schließen, dass es eine hohe Kongruenz zwischen der ersten und zweiten Bewertung gab. Folglich zeigte sich hier das Potential, durch die Verkürzung der Betrachtungszeit infolge beschleunigter Abspielgeschwindigkeit die benötigte Zeit für die Bewertung zu reduzieren.

Somit könnten die notwendigen bezahlten Überstunden reduziert und damit auch Personalkosten insgesamt für die V-OSCE verringert werden (Kelly und Murphy 2004, Rau et al. 2011). Es bedarf jedoch weiterer Studien zum Zweck der Feststellung, bei welcher Abspielgeschwindigkeit reliable Bewertungen noch möglich sind. Ebenso sollte zukünftig geprüft werden, ob elektronische Checklisten die Bewertung noch weiter beschleunigen können.

5.3 Einschränkungen und Ausblick

Eine Evaluation der V-OSCE durch die Studierenden war nicht Bestandteil der vorliegenden Dissertationsstudie. Auch wurde nicht erhoben, inwieweit sich die Studierenden durch die Videoaufnahme beeinflusst fühlten. Dennoch musste in Erwägung gezogen werden, dass das Filmen einen Einfluss auf die Leistungen im Rahmen der Prüfung gehabt haben könnte. Dennoch war anzunehmen, dass im Zuge der zunehmenden Digitalisierung Studenten heutzutage an die Implementierung von neuen Technologien gewöhnt sind und sich schnell an diese anpassen oder per se akzeptieren. Davon abgesehen waren die Video-Aufnahmen kein Novum mehr. Die Nutzung von Video-Aufnahmen zu Zwecken der Selbstreflektion, dem qualifiziertem Feedback durch Lehrkräfte oder der Selbstbeurteilung wurden in der Literatur bereits ausführlich beschrieben (Maloney et al. 2013a, Hawkins et al. 2012).

Zudem gab es keine Untersuchungen dahingehend, wie der Charakter der Prüfung die Studenten beeinflusst haben könnte. Die V-OSCE war in das Pflichtmodul 6.1 „Operative Medizin“ eingebettet. Die Teilnahme an der Video-Aufzeichnung erfolgte aber auf freiwilliger Basis. Die Studenten konnten eine Aufzeichnung auch ablehnen und sich durch einen Prüfer vor Ort bewerten lassen. Dies fand in der untersuchten Semesterkohorte nicht statt. Inwiefern dies im Rahmen eines freiwilligen Zwangs, um nicht hinter den anderen Studenten zurückzustehen, oder wirklich freiwillig geschah und wie dies die studentischen Leistungen beeinflusst haben könnte, war nicht Teil dieser Studie.

Eine weitere Limitation dieser Studie zeigt sich in der Tatsache, dass potentielle Störfaktoren nicht untersucht bzw. im Vorfeld nicht exkludiert werden konnten. Da alle Studenten der Semesterkohorte diese Prüfung ablegen mussten, konnten Studenten im Vorfeld nicht nach sozioökonomischen Gesichtspunkten oder Bildungshintergrund vorselektiert und randomisiert werden. Ebenso konnten einschlägige oder allgemeine Vorerfahrungen durch absolvierte Famulaturen oder berufsnahe Ausbildungen nicht erfasst und bei der Planung berücksichtigt werden. Zudem sind Lehrkräfte und Rater tagtäglich mit einer sehr heterogenen Gruppe von

Studenten konfrontiert, denen sie gerecht werden müssen. Gleichwohl bedeutet dies im Umkehrschluss, dass diese Prüfungsergebnisse und deren Auswertung einen realistischen Semesterquerschnitt zeigen.

Ebenso muss einschränkend konstatiert werden, dass es sich bei der Prüfungssituation nur um die Simulation eines Arzt-Patienten-Kontakts handelte. Somit entsprach die Prüfung der dritten Stufe der Miller-Pyramide und nicht der vierten und letzten Stufe. Somit limitiert sich die Generalisierbarkeit der vorliegenden Studienergebnisse in Hinblick auf den klinischen Arbeitsalltag.

Aufgrund dieser simulierten Prüfungssituation hielten manche Studenten im Video eher einen Monolog über die ihrer Meinung nach prüfungsrelevanten Inhalte. Ein Gespräch zwischen Arzt und Patient entwickelte sich selten.

Abschließend muss noch die Einschränkung auf zwei Rater in Betracht gezogen werden. Bei der Rekrutierung der Rater war die Bereitschaft im Lehrkörper, 78 Videos in bezahlten Überstunden anzusehen und zu bewerten, eher unterdurchschnittlich ausgeprägt. Nach Sichtung der Bewerbungen um diese vergütete, zusätzliche Arbeit blieben nur zwei Rater, deren Qualifizierung und Zeitkapazität ausreichend waren, um diese Studie durchführen zu können.

5.4 Einordnung der Studie in den wissenschaftlichen Kontext

Van der Vleuten (1996) benannte fünf Qualitätskriterien, die zu einer guten Prüfung beitragen und die auch am Format der V-OSCE nachvollzogen werden konnten. Diese sind die Reliabilität, die Validität, der Lerneffekt, die Akzeptanz und die Kosten.

In Hinblick auf die Reliabilität konnte nachgewiesen werden, dass im Kontext einer noch zu etablierenden OSCE zur Vorbereitung auf das PJ diese Einzelstations-V-OSCE gute Werte lieferte, auf deren Grundlage eine Multi-Stations-OSCE mit einem Cronbachs α von >0,90 und Interrater-Kongruenzen von >0,80 zu erwarten sind. Somit kann die V-OSCE als reliable Prüfung gelten.

Validitätsbezeugungen sind in dieser Studie limitiert, da keine größeren Untersuchungen dazu stattgefunden haben. Trotzdem wurden die Checklisten, Szenarien und Schauspielpatienten unter Expertise von Chirurgen und OSCE-Experten erstellt und ausgesucht und können somit als valide gelten.

Der Lerneffekt war hoch, da die gesamte Semesterkohorte mit einer durchschnittlichen Leitung von 88,9% bestanden hat. Dabei war die Akzeptanz der Studenten für diese Prüfung nicht Gegenstand dieser Studie.

Die Kosten für diese V-OSCE waren ein kritischer Punkt wie bei vielen Formen der klinisch-praktischen Prüfungen. Die bezahlten Überstunden als Mehrkosten standen dabei den niedrigeren administrativen Kosten gegenüber und wurden von uns abschließend als gerechtfertigt und einer normalen OSCE überlegen angesehen. Mit der Einbettung in die Multi-Stations-OSCE könnten darüber hinaus weitere Kosteneinsparungen entstehen, da sich die administrative Arbeit durch die Zusammenlegung der Pilot-Einzelprüfungen mutmaßlich weiter reduzieren wird. Darüber hinaus können möglicherweise aufgrund der qualitativ guten Grundlage der V-OSCE-Station in einer Multi-Stations-OSCE grundsätzlich Stationen eingespart werden, ohne dass die Qualität der Prüfung leidet. Dies könnte wiederum zur Kostensenkung beitragen.

Die Kosten-Nutzen-Abwägung einer neuen klinisch-praktischen Prüfung bedarf also – wie dargelegt – einer umfassenden Diskussion. Dennoch konnte mit dieser Studie eine Möglichkeit zur Vereinfachung der organisatorischen Arbeit einer OSCE gezeigt werden, während gleichzeitig die qualitativ gute Bewertung der studentischen Leistungen gewährleistet war. So kommen wir zu dem Schluss, dass die Kosten-Nutzen-Abwägung dieser Prüfung klar zu Gunsten der Nutzen-Seite ausfällt.

6. Schlussfolgerung

Abschließend können wir feststellen, dass die V-OSCE eine praktikable, objektive und reliable Prüfungsform ist und somit eine Alternative zum traditionellen Bewerten vor Ort darstellt. Wir bewegen uns damit auf einer Linie mit den aktuellen Entwicklungen im Bereich der medizinischen Lehre, die eine immer stärkere Einflechtung von praktischen Prüfungen in Form von OSCEs vorsehen (Patricio et. 2013, Laidlaw et al. 2014). Insbesondere im Kontext der nationalen Entwicklung entlang des neuen Lernzielkatalogs könnte dieses Projekt zu einer besseren Handhabung der klinisch-praktischen Prüfungen im chirurgischen Kontext beitragen (Kadmon et al. 2013).

Dennoch sind weitere Studien auf diesem noch recht unerforschten Feld der V-OSCEs notwendig. Zum einen sollten weitere Studien in Hinblick auf den studentischen Lerneffekt durchgeführt werden. Eine Video-OSCE setzt durch die Aufzeichnung der studentischen Leistung auf Video neue Aspekte. Diese Aufzeichnung ist jederzeit erneut abspielbar und kann zur Verdeutlichung der Stärken und Schwächen des Studenten jederzeit herangezogen werden. Somit entstehen einige neue Fragestellungen. Beeinflusst das Wissen um die Unvergänglichkeit der erbrachten Leistung den Studenten in seiner Vorbereitung auf die Prüfung? Können hierdurch neue Lernanreize gesetzt und das Lernverhalten der Studenten beeinflusst werden? Zum anderen sollten auch die Einflüsse des anschließenden Feedbacks durch die Schauspielpatienten für die Studenten weiter untersucht werden. Ebenso könnten die Möglichkeiten eines ratergestützten Feedbacks ausgelotet werden, welches bislang aus organisatorischen Gründen nicht vorgesehen war.

7. Zusammenfassung

7.1 Hintergrund

Für ein gutes und vertrauensbasiertes Arzt-Patienten-Verhältnis sind kommunikative Fertigkeiten grundlegend, um das Verständnis und die Compliance des Patienten sicherzustellen (Kaplan et al. 1989, Geisler 1992). Vor diesem Hintergrund war es unser Ziel, eine reliable Prüfung zu entwerfen, welche im Rahmen einer OSCE videoassistiert (V-OSCE) die Fähigkeiten der Studenten, ein präoperatives Aufklärungsgespräch zu führen, zum Inhalt hat.

7.2 Methoden

155 Studenten des 6. Klinischen Semester nahmen an der summativen V-OSCE im Rahmen des Pflichtmoduls „Operative Medizin“ teil. Sie wurden instruiert, zu je einem von drei Krankheitsbildern und deren operativen Behandlungsmöglichkeiten ein Aufklärungsgespräch mit einem Schauspielpatienten zu führen und dieses auf Video aufzunehmen. Anschließend wurden die Videos von zwei unabhängigen Prüfern ausgewertet. Die Ergebnisse sowie die verwendeten Checklisten wurden hinsichtlich der Gütekriterien (u. a. Deskriptive Statistik, Interrater-Reliabilität, Trennschärfe, Cronbachs α) mittels SPSS statistisch ausgewertet.

7.3 Ergebnisse

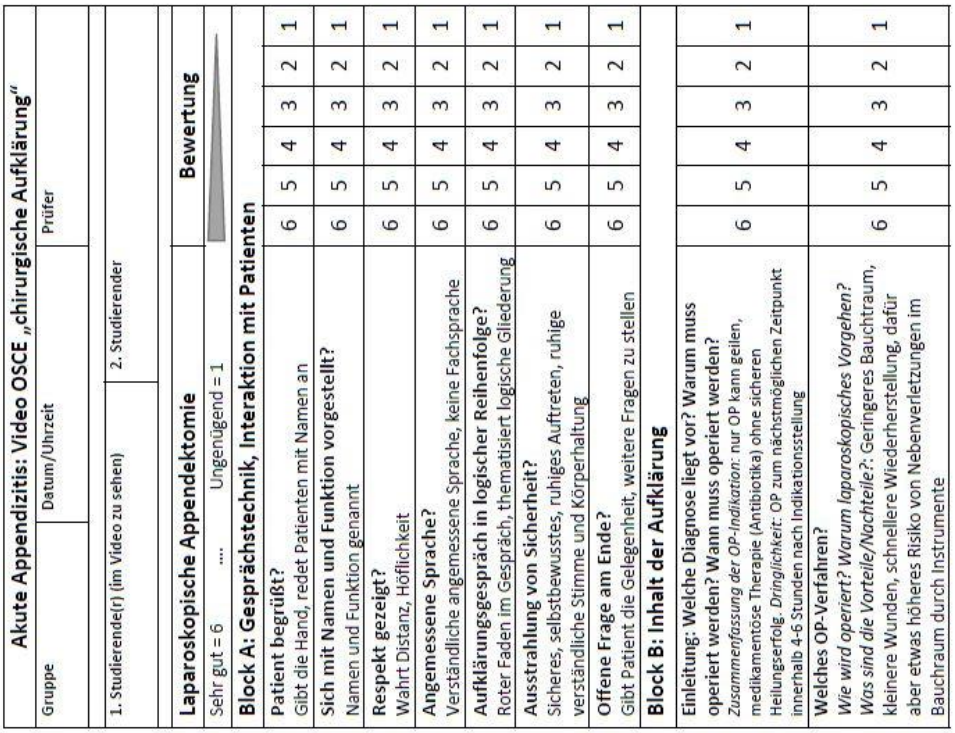
Alle Studenten haben die Prüfung bestanden mit durchschnittlichen Leistungen von 91,0% (± 4,0%), 88,4% (± 4,4%) und 87,0% (± 4,7%) für die drei Prüfungsthemen Appendektomie, Cholezystektomie und Leistenhernienverschluss. Die Trennschärfen der Checklisten waren überwiegend von moderater bis guter Qualität und die interne Konsistenz zeigte sich mit Werten zwischen 0,565 bis 0,605 für Cronbachs α passabel. Die Interrater-Reliabilität wies mit einem Pearson’s ρ von 0,80 und einem Intraklassen-Korrelations-Koeffizienten (ICC 2.1) von 0,78 hohe Werte an Übereinstimmung auf, obwohl lediglich ein Prüfer einen chirurgischen Hintergrund hatte, während der andere in der Zahnmedizin arbeitete.

7.4 Diskussion und Ausblick

Die V-OSCE ist eine praktikable und reliable Methode, Studenten kommunikative Fertigkeiten, welche zu einem guten Arzt-Patienten-Verhältnis im präoperativen Umfeld beitragen, zu vermitteln. Zusätzlich punktet sie mit ihrer Effizienz, da sowohl der organisatorische Aufwand verringert als auch fachfremde Ärzte als Prüfer eingesetzt werden können.

8. Anhang

8.1 Checklisten



Akute Appendizitis: Video OSCE „chirurgische Aufklärung“								
Gruppe	Datum/Uhrzeit	Prüfer						
1. Studierende(r) (im Video zu sehen)	2. Studierender							
Laparoskopische Appendektomie		**Bewertung**						
Sehr gut = 6 ..... Ungenügend = 1								
Block A: Gesprächstechnik, Interaktion mit Patienten								
Patient begrüßt? Gibt die Hand, redet Patienten mit Namen an		6	5	4	3	2	1	
Sich mit Namen und Funktion vorgestellt? Namen und Funktion genannt		6	5	4	3	2	1	
Respekt gezeigt? Wahrt Distanz, Höflichkeit		6	5	4	3	2	1	
Angemessene Sprache? Verständliche angemessene Sprache, keine Fachsprache		6	5	4	3	2	1	
Aufklärungsgespräch in logischer Reihenfolge? Roter Faden im Gespräch, thematisiert logische Gliederung		6	5	4	3	2	1	
Ausstrahlung von Sicherheit? Sicheres, selbstbewusstes, ruhiges Auftreten, ruhige verständliche Stimme und Körperhaltung		6	5	4	3	2	1	
Offene Frage am Ende? Gibt Patient die Gelegenheit, weitere Fragen zu stellen		6	5	4	3	2	1	
Block B: Inhalt der Aufklärung								
Einleitung: Welche Diagnose liegt vor? Warum muss operiert werden? Wann muss operiert werden? *Zusammenfassung der OP-Indikation:* nur OP kann geilen, medikamentöse Therapie (Antibiotika) ohne sicheren Heilungserfolg. *Dringlichkeit:* OP zum nächstmöglichen Zeitpunkt innerhalb 4-6 Stunden nach Indikationsstellung		6	5	4	3	2	1	
Welches OP-Verfahren? *Wie wird operiert? Warum laparoskopisches Vorgehen? Was sind die Vorteile/Nachteile?:* Geringeres Bauchtraum, kleinere Wunden, schnellere Wiederherstellung, dafür aber etwas höheres Risiko von Nebenverletzungen im Bauchraum durch Instrumente		6	5	4	3	2	1	

Mögliche Erweiterung des Eingriffs?	Genannt	Nicht genannt
Umsteigen auf offenes Verfahren mit Eröffnung Bauchhöhle	2	1
Operation an anderen Organen (Eierstöcke, Meckel-Divertikel)	2	1
Allgemeine OP-Risiken		
Lagerungsschäden	2	1
Thrombose/Embolie	2	1
Blutung	2	1
Infektion	2	1
Verletzung von Gefäßen oder Nerven	2	1
Narbenbildung	2	1
Verwachsungen/Darmverschluss	2	1
Nahtbruch/Narbenbruch	2	1
Operationsspezifische Komplikationen		
Verletzung von Gefäßen (Aorta, Becken) beim Setzen der Trokare	2	1
Verletzung von Nachbarorganen (Darm, Harnblase)	2	1
Hautemphysem	2	1
Schmerzen im Schulterbereich	2	1
Pneumothorax	2	1
Postoperative Verhaltenmaßnahmen		
Zügiger Kostaufbau und rasche Mobilisation	2	1
Langfristig keine Einschränkung in der Ernährung zu erwarten	2	1
Auswertung		
Punktzahl Block A x 0,3		
Punktzahl Block B x 0,7		
Gesamtpunktzahl		
OSCE bestanden?	NEIN	JA
Positivbeispiel für Abschlussbesprechung?		JA
Anmerkungen		

Abbildung X: Checkliste „Akute Appendizitis“

Abbildung XI: Checkliste „Symptomatische Cholezystolithiasis"

Symptomatische Cholezystolithiasis: Video OSCE „chirurgische Aufklärung"		
Gruppe	**Datum/Uhrzeit**	**Prüfer**

1. Studierende(r) (im Video zu sehen)	**2. Studierender**

Laparoskopische Cholezystektomie	**Bewertung**					
Sehr gut = 6 .... Ungenügend = 1						
Block A: Gesprächstechnik, Interaktion mit Patienten						
Patient begrüßt? Gibt die Hand, redet Patienten mit Namen an	6	5	4	3	2	1
Sich mit Namen und Funktion vorgestellt? Namen und Funktion genannt	6	5	4	3	2	1
Respekt gezeigt? Wahrt Distanz, Höflichkeit	6	5	4	3	2	1
Angemessene Sprache? Verständliche angemessene Sprache, keine Fachsprache	6	5	4	3	2	1
Aufklärungsgespräch in logischer Reihenfolge? Roter Faden im Gespräch, thematisiert logische Gliederung	6	5	4	3	2	1
Ausstrahlung von Sicherheit? Sicheres, selbstbewusstes, ruhiges Auftreten, ruhige verständliche Stimme und Körperhaltung	6	5	4	3	2	1
Offene Frage am Ende? Gibt Patient die Gelegenheit, weitere Fragen zu stellen	6	5	4	3	2	1
Block B: Inhalt der Aufklärung						
Einleitung: Welche Diagnose liegt vor? Warum muss operiert werden? Wann muss operiert werden? *Zusammenfassung der OP-Indikation:* Beschwerden der Patientin, bereits stattgehabte Komplikation und ERCP. *Dringlichkeit:* elektive Operation	6	5	4	3	2	1
Welches OP-Verfahren? *Wie wird operiert? Warum laparoskopisches Vorgehen? Was sind die Vorteile/Nachteile?*: Geringeres Bauchtraum, kleinere Wunden, schnellere Wiederherstellung, dafür aber etwas höheres Risiko von Nebenverletzungen im Bauchraum durch Instrumente	6	5	4	3	2	1

Mögliche Erweiterung des Eingriffs?	Genannt	Nicht genannt
Umsteigen auf offenes Verfahren mit Eröffnung Bauchhöhle	2	1
Operation an den Gallenwegen, Entfernung von Lebergewebe	2	1
Allgemeine OP-Risiken		
Lagerungsschäden	2	1
Thrombose/Embolie	2	1
Blutung	2	1
Infektion	2	1
Verletzung von Gefäßen oder Nerven	2	1
Narbenbildung	2	1
Verwachsungen/Darmverschluss	2	1
Nahtbruch/Narbenbruch	2	1
Operationsspezifische Komplikationen		
Verletzung von Nachbarorganen (Darm, Leber, Leberarterien)	2	1
Verletzung der Gallenwege (Leckage, chron. Leberzirrhose)	2	1
Hautemphysem	2	1
Schmerzen im Schulterbereich	2	1
Pneumothorax	2	1
Postoperative Verhaltenmaßnahmen		
Zügiger Kostaufbau und rasche Mobilisation	2	1
Langfristig Vermeidung fettreicher, opulenter Speisen	2	1
Auswertung		
Punktzahl Block A x 0,3		
Punktzahl Block B x 0,7		
Gesamtpunktzahl		

OSCE bestanden?	**NEIN**	**JA**
Positivbeispiel für Abschlussbesprechung?		**JA**
Anmerkungen		

Abbildung XII: Checkliste „Leistenhernie"

Leistenhernie: Video OSCE „chirurgische Aufklärung"							
Gruppe	**Datum/Uhrzeit**	**Prüfer**					
1. Studierende(r) (im Video zu sehen)	**2. Studierender**						
Offener Leistenhernienverschluss mit Netzeinlage	**Bewertung**						
Sehr gut = 6 .... Ungenügend = 1							
Block A: Gesprächstechnik, Interaktion mit Patienten							
Patient begrüßt? Gibt die Hand, redet Patienten mit Namen an	6	5	4	3	2	1	
Sich mit Namen und Funktion vorgestellt? Namen und Funktion genannt	6	5	4	3	2	1	
Respekt gezeigt? Wahrt Distanz, Höflichkeit	6	5	4	3	2	1	
Angemessene Sprache? Verständliche angemessene Sprache, keine Fachsprache	6	5	4	3	2	1	
Aufklärungsgespräch in logischer Reihenfolge? Roter Faden im Gespräch, thematisiert logische Gliederung	6	5	4	3	2	1	
Ausstrahlung von Sicherheit? Sicheres, selbstbewusstes, ruhiges Auftreten, ruhige verständliche Stimme und Körperhaltung	6	5	4	3	2	1	
Offene Frage am Ende? Gibt Patient die Gelegenheit, weitere Fragen zu stellen	6	5	4	3	2	1	
Block B: Inhalt der Aufklärung							
Einleitung: Welche Diagnose liegt vor? Warum muss operiert werden? Wann muss operiert werden? *Zusammenfassung der OP-Indikation:* Bruchpforte wird sich nicht von selbst verschließen und wird größer. Bruchband zwecklos. Bruchinhalt kann Einklemmen und dann Notfall-OP. *Dringlichkeit:* elektive Operation	6	5	4	3	2	1	
Welches OP-Verfahren? *Wie wird operiert? Warum offenes Verfahren mit Netzeinlage? Was sind die Vorteile/Nachteile?*: Sehr gute und dauerhafte Verstärkung der Bauchdecke/des Leistenkanals, geringe Rezidivrate. Operation in Lokalanästhesie grundsätzlich möglich.	6	5	4	3	2	1	

Mögliche Erweiterung des Eingriffs?	Genannt	Nicht genannt
Eröffnung der Bauchhöhle (über Leiste oder Medianlaparotomie)	2	1
Resektion von Darm oder anderen Organen im Bruchsack	2	1
Allgemeine OP-Risiken		
Lagerungsschäden	2	1
Thrombose/Embolie	2	1
Blutung	2	1
Infektion	2	1
Verletzung von Gefäßen oder Nerven	2	1
Narbenbildung	2	1
Verwachsungen/Darmverschluss	2	1
Nahtbruch/Narbenbruch	2	1
Operationsspezifische Komplikationen		
Verletzung/Einengung von Haut- und Muskelnerven, Gefäßen	2	1
Verletzung v. Nachbarorganen (Darm, Harnblase, innere Genitale)	2	1
Unverträglichkeit des Kunststoffnetzes	2	1
Bewegungseinschränkung durch das Netz, chron. Schmerzen	2	1
Rezidiv, Dislokation des Netzes	2	1
Postoperative Verhaltenmaßnahmen		
Zügiger Kostaufbau und rasche Mobilisation	2	1
Langfristig keine Einschränkungen in der Ernährung zu erwarten	2	1
Auswertung		
Punktzahl Block A x 0,3		
Punktzahl Block B x 0,7		
Gesamtpunktzahl		
OSCE bestanden?	**NEIN**	**JA**
Positivbeispiel für Abschlussbesprechung?		**JA**
Anmerkungen		

8.2 Einverständiserklärung

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Universitätsmedizin Göttingen, 37099 Göttingen
Allgemein- u. Viszeralchirurgie, PD Dr. S. König, Robert-Koch-Str. 40, 37075 Göttingen

Zentrum 8 Chirurgie
Klinik für Allgemein- und Viszeralchirurgie
Direktor: Univ.-Prof. Dr. med. H. Becker

37099 Göttingen **Briefpost**
Robert-Koch-Straße 40, 37075 Göttingen **Adresse**
0551 39-8977 **Telefon**

Dokumentation und Einverständniserklärung über Videoaufzeichnung

Das Filmmaterial geht in das Eigentum der UMG über und darf die UMG nicht verlassen. Es wird ausschließlich für die Zwecke der Lehre genutzt. Nach Aufnahme gelangt das Material in die Abt. Allgemein- und Viszeralchirurgie, wo es auf einem Speichermedium archiviert wird. Ansprechpartnerin dort ist Frau PD Dr. König Tel.: 0551-398977.

Dateiname des Videotestats = Gruppennummer und beide Nachnamen:

.....

ggf. Laufzeiten des auszuwertenden Videoabschnittes:

von min/sec .....bis .....

Student(in) 1:
Ich erkläre mich einverstanden

.....
Name in leserlicher Druckschrift

.....
Datum, Unterschrift

Auf dem Video zu sehen ist / sind:

○ Student(in) 1

○ Student(in) 2

Student(in) 2:
Ich erkläre mich einverstanden

.....
Name in leserlicher Druckschrift

.....
Datum, Unterschrift

Abbildung XIII: Einverständniserklärung

9. Literaturverzeichnis

ÄApprO 2002: Approbationsordnung für Ärzte vom 27. Juni 2002. Bundesgesetzblatt Jahrgang 2002 Teil I, 44, ausgegeben zu Bonn am 03. Juli 2002, 2405 – 2435

Al-Wardy N M (2010): Assessment Methods in Undergraduate Medical Education. Sultan Qaboos Univ Med J 10, 203-209

Baribeau D A, Mukovozov I, Sabljic T, Eva K E, Delottinville C B (2012): Using an objective structured video exam to identify differential understanding of aspects of communication skills. Med Teach 34, e242 - e250

Barzansky B, Etzel S I (2003): Educational Programs in US Medical Schools 2002-2003. JAMA 290, 1190 – 1196

Barthel Y, Götze H, Schwarz R, Schön M: Qualitätserkundungsstudie: Die Interdisziplinäre Station Chirurgie A 4.1 des Universitätsklinikums Leipzig. Abschlussbericht. Leipzig 2005

BGBL 2002: Bekanntmachung der Neufassung des Bürgerlichen Gesetzbuchs vom 2. Januar 2002. Ausgegeben zu Bonn am 8. Januar 2002. Bundesgesetzblatt Teil I, 2, 42 – 346

BGBL 2013: Gesetz zur Verbesserung der Rechte von Patientinnen und Patienten vom 20. Februar 2013. Ausgegeben zu Bonn am 20. Februar 2013. Bundesgesetzblatt Teil I, 9, 277 – 282

Brannick M T, Erol-Korkmaz H T, Prewett M (2011): A systematic review of the reliability of objective structured clinical examination scores. Med Educ 45, 1181 – 1189

Breitkreutz R, Dutiné M, Scheiermann P, Hempel D, Kujumdshiev S, Ackermann H, Seeger F H, Seibel A, Walcher F, Hirche T O (2013): Thorax, Trachea, and Lung Ultrasonography in Emergency and Critical Care Medicine: Assessment ofan Objective Structured Training Concept. Emerg Med Int 2013, 312758

Brown G, Manogue M (2001): AMEE Medical Education Guide No. 22: Refreshing lecturing: a guide for lecturers. Med Teach 23, 231 – 244

Brydges R, Manzone J, Shanks D, Hatala R, Hamstra S J, Zendejas B, Cool D A (2015): Self-regulated learning in simulation-based training: a systematic review and meta-analysis. Med Educ 49, 368 – 378

Casey P M, Goepfert A R, Espey E L, Hammoud M M, Kaczmarczyk J M, Katz N T, Neutens J J, Nuthalapaty F S, Peskin E (2009): To the point: reviews in medical education – the Objective Structured Clinical Examination. Am J Obstet Gynecol 200, 35 – 34

Cate O T, Durning S (2007): Peer teaching in medical education: twelve reasons to move from theory to pratice. Med Teach 29, 591 – 599

Chen A C, Lee M S, Chen W J, Lee S T (2013): Assessment in Orthopedic Training – An Analysis of Rating Consistency by Using an Objective Structured Examination Video. J Surg Educ 70, 189 – 192

Chipman J G, Beilman G J, Schmitz C C, Seatter S C (2007): Development and Pilot Testing of an OSCE for Difficult Conversations in Surgical Intensive Care. J Surg Educ 64, 79 – 87

Cho K, MacArthur C (2011): Learning by Reviewing. J Educ Psychol 103, 73 – 84

Cleland J A, Abe K, Rethans J-J (2009): The use of simulated patients in medical education: AMEE Guide No. 42. Med Teach 31, 477 – 486

Collins J P, Harden R M (1998): AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. Med Teach 20, 508 – 521

Cox E P (1980): The Optimal Number of Response Alternatives for a Scale: A Review. J Mark Res 17, 407 – 422

Cronbach L J (1951): Coefficient alpha and the internal structure of tests. Psychometrika 16, 297 – 334

Dillschneider J, Theuer D, Mieth M, Büchler M W (2012): Zum neuen Patientenrechtegesetz, Bedeutung für den Chirurgen. Chirurg 83, 661 – 666

Etrillard S: Gesprächsrhetorik. 2. Auflage; Business Village, Göttingen 2009

Geisler L: Begegnung im Gespräch. 3., erweiterte Auflage; Pharma Verlag, Frankfurt 1992

Gómez S S, Ostos E M C, Solano J M M, Salado T F H (2013): An electronic portfolio for quantitative assessment of surgical skills in undergraduate medical education. BMC Med Educ 13, 65

Govindan V K (2008): Enhancing communication skills using an OSCE and peer review. Med Educ 42, 535 – 536

Harden R McG, Stevenson M, Downie W W, Wilson G M (1975): Assessment of Clinical Competence using Objective Structured Examination. Br Med J 1975, 447 – 451

Härtl A, Bachmann C, Blum K, Höfer S, Peters T, Preusche I, Raski B, Rüttermann S, Wagner-Menghin M, Wünsch A, Kiessling C, GMA-Ausschuss Kommunikative und Soziale Kompetenzen (2015): Wunsch und Wirklichkeit – eine Umfrage im deutschsprachigen Raum zum Lehren und Prüfen kommunikativer Kompetenzen im Medizinstudium. GMS Z Med Ausbild 32, Doc 56

Hawkins S C, Osborne A, Schonfield S J, Pournaras D J Chester J F (2012): Improving the accuracy of self-assessment of practical clinical skills using video feedback – The importance of including benchmarks. Med Teach 34, 279 – 284

Hoffman M, Wilkinson J E , Xu J, Wiecha J (2014): The perceived effects of faculty presence vs. absence on small-group learning and group dynamics: a quasi-experimental study. BMC Med Educ 14, 258

Humphris G M, Kaney S (2008): The Objective Structured Video Exam for assessment of communication skills. Med Educ 34, 939 – 945

Hulsmann R L, Ros W J, Winnubst J A, Bensing J M (1999): Teaching clinically experienced physicians communication skills. A review of evaluation studies. Med Educ 33, 655 – 668

Janssen J, Laatz W: Statistische Datenanalyse mit SPSS – Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests. 8. Auflage; Springer Verlag, Berlin, Heidelberg 2013

Kadmon M, Bender M J, Adili F, Arbab D, Heinemann M K, Hofmann H S, König S, Küper M A, Obertacke U, Rennekampff H-O, Rolle U, Rücker M, Sader R, Tingart M, Tolksdorf M M, Tronnier V, Will B, Wlacher F (2013): Kompetenzorientierung in der medizinischen Ausbildung. Chirurg 84, 277 – 285

Kalbitz M, Liener U, Kornmann M, Gebhard F (2010): Studentische Evaluation einer objektiven, strukturierten klinischen Prüfungsmethode (OSCE) im Fach Chirurgie und Orthopädie. Unfallchirurg 113, 726 – 733

Kaplan S H, Greenfield S, Ware J E (1989): Assessing the Effects of Physician-Patient Interactions on the Outcomes of Chronic Disease. Med Care 27, 110 – 127

Kastner M: Work-Life Balance für Extremjobber. In: Kaiser, Ringlstetter (Hrsg.): Work-Life Balance: Erfolgsversprechende Konzepte und Instrumente für Extremjobber. Springer-Verlag, Berlin, Heidelberg 2010, 1 – 25

Kelava A, Moosbrugger H: Testtheorie und Fragebogenkonstruktion. 2., aktualisierte und überarbeitete Auflage; Springer Medizin Verlag, Berlin, Heidelberg 2012

Kelly M, Murphy A (2004): An evaluation of the cost of designing, delivering and assessing an undergraduate communication skills module. Med Teach 26, 610 – 614

Khan K Z, Gaunt K, Ramachandran S, Pushkar P (2012): The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration. Med Teach 35, e1447 – e1463

Kiehl C, Simmenroth-Nayda A, Goerlich Y, Entwistle A, Schiekirka S, Ghadimi B M, Raupach T, Koenig S (2014): Standardized and quality-assured video-recorded examination in undergraduate education: informed consent prior to surgery. J Surg Res 191, 64 – 73

Kissin E Y, Grayson P C, Cannella A C, DeMarco P J, Evangelisto A, Goyal J, Al Haj R, Higgs J, Malone D G, Nishio M J, Tabechian D, Kaeley G S (2014): Musculoskeletal Ultrasound Objective Structured Clinical Examination: An Assessment of the Test. Arthritis Care Res (Hoboken) 66, 2 – 6

Koch A: Studentische Tutoren in einer „objective structured clinical examination“ (OSCE): Evaluation ihrer Bewertungsleistungen. Med. Diss. Göttingen 2008

Kuckartz U, Rädiker S, Ebert T, Schehl J: Statistik: Eine verständliche Einführung. 2., überarbeitete Auflage; Springer Fachmedien Verlag, Wiesbaden 2013

Laidlaw A, Salisbury H, Doherty E M, Wiskin C (2014): National survey of clinical communication assessment in medical education in the United Kingdom (UK). BMC Med Educ 14, 10

Langewitz W (2012): Zur Erlernbarkeit der Arzt-Patienten-Kommunikation in der Medizinischen Ausbildung. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 55, 1175 – 1182

Lernzielkatalog klinischer Studienabschnitt (2008): Der Göttinger Lernzielkatalog für den klinischen Studienabschnitt, http://www.med.uni-goettingen.de/de/media/G1-2_lehre/lernziel-katalog.pdf, letzter Aufruf: 16.05.2014. Website nach Umstellung des Lernzielkatalogs nicht mehr verfügbar.

Levine H G, McGuire C H (1970): The validity and reliability of oral examinations in assessing cognitive skills in medicine. J Educ Meas 7, 63 – 74

Lienert G A, Raatz U: Testaufbau und Testanalyse. 6. Auflage; Psychologie Verlags Union, Weinheim 1998

Maguire P, Pitceathly C (2002): Key communication skills and how to acquire them. BMJ 325, 697 – 700

Maloney S, Paynter S, Storr M, Morgan P (2013): Implementing student self-video of performance. Clin Teach 10, 323 – 327

Maloney S, Storr M, Morgan P, Ilic D (2013): The effect of student self-video of performance on clinical skill competency: a randomised controlled trial. Adv Health Sci Educ Theory Pract 18, 81 – 89

Matell M S, Jacobi J (1972): Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. J Appl Psychol 56, 506 – 509

McLaughlin K, Ainsline M, Coderre S, Wright B, Violato C (2009): The effect of differential rater function over time (DRIFT) on objective structured clinical examination ratings. Med Educ 43, 989 – 992

McNaughton N, Ravitz P, Wadell A, Hodges B D (2008): Psychiatric Education and Simulation: A Review of the Literature. Can J Psychiatry 53, 85 – 93

Miles J, Shevlin M: Applying Regression & Correlation – A Guide for Students and Researchers. 1. Auflage; SAGE Publications Ltd, London 2001

Miller G E (1990): The assessment of clinical skills/competence/performance. Acad Med 65, 63 – 67

Möltner A, Schellberg D, Jünger J (2006): Grundlegende quantitative Analysen medizinischer Prüfungen. GMS Z Med Ausbild 23, Doc 53

Nachreiner F, Rädiker B, Janßen D, Schomann C: Untersuchungen zum Zusammenhang zwischen der Dauer der Arbeitszeit und gesundheitlichen Beeinträchtigungen, Ergebnisse einer Machbarkeitsstudie. Gesellschaft für Arbeits-, Wirtschafts- und Organisationspsychologische Forschung e.V., Oldenburg 2005

O'Shea E (2003): Self-directed learning in nurse education: a review to literature. J Adv Nurs 43, 62 – 70

Pabst R (1995): Medical Education and Reform Initiatives in Germany. Acad Med 70, 1006 – 1011

Patrício M F, Julião M, Fareleira F, Carneiro A V (2013): Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med Teach 35, 503 – 514

Pell G, Fuller R, Horner M, Roberts T (2010): How to measure the quality of the OSCE: A review to metrics – AMEE guide no. 49. Med Teach 32, 802 – 811

Pflichtmodule klinischer Studienabschnitt (2014): Pflichtmodule im klinischen Studienabschnitt Humanmedizin M6.1 Operative Medizin, http://www.med.uni-goettingen.de/de/content/studium/1068_1233.html, letzter Aufruf: 16.05.2014. Website nach Umstellung des Lernzielkatalogs nicht mehr verfügbar.

Projektion Lernzielkatalog auf klinische Module (2008): Abbildung des Göttinger Lernzielkatalogs auf die klinischen Module, http://www.med.uni-goettingen.de/de/media/studium/stundenplaene_klinik/lehre_projektion_des_lernzielkataloges_auf_die_module_blockpraktika_und_pj.pdf, letzter Aufruf: 04.11.2013. Website nach der Umstellung des Lernzielkatalogs nicht mehr verfügbar.

Rau T, Fegert J, Liebhardt H (2011): Wie hoch liegen die Personalkosten für die Durchführung einer OSCE? Eine Kostenaufstellung nach betriebswirtschaftlichen Gesichtspunkten. GMS Z Med Ausbild 28, Doc13

Ringel N, Bürmann B M, Fellmer-Drueg E, Roos M, Herzog W, Nikendei Ch, Wischmann T, Weiss C, Eicher Ch, Engeser P, Schultz J-H, Jünger J (2015): Integriertes Peer Teaching klinischer und kommunikativer Kompetenzen – Wie bereiten wir studentische Tutoren darauf vor?. Psychother Psychosom Med Psychol ohne Angabe einer Ausgabe, online publiziert am 20.03.2015, http://dx.doi.org/10.1055/s-0034-1398549

Scheffé H: The analysis of Variance. John Wiley & Sons, New York 1999

Schmitt D C (2009): Kommunkation und Compliance – der Behandlungserfolg hängt von der Arzt-Patienten-Kommunikation ab. Breast Care 4, 128 – 129

Schmitt N (1996): Uses and Abuses of Coefficient Alpha. Psychol Assess 8, 350 – 353

Shah B, Miler R, Poles M, Zabar S, Gillespie C, Weinshel E, Chokhavatia S (2011): Informed Consent in the Older Adult: OSCEs for Assessing Fellows‘ ACGME and Geriatric Gastroenterology Competencies. Am J Gastroenterol 106, 1575 – 1579

Shrout P E, Fleiss J L (1979): Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull 86, 420 – 428

Srinivasan J (1999): Observing communication skills for informed consent: an examiner's experience. Ann R Coll Physicians Surg Can 32, 437-440

Tchorz K M, Binder S B, White M T, Lawhorne L W, Bentley D M, Delaney E A, Borchers J, Miller M, Barney L M, Dunn, M M, Rundell K W, Thambipillai T, Woods R J, Markert R J, Parikh P P, McCarthy M C (2013): Palliative and end-of-life care training during surgical clerkship. J Surg Res 185, 97 – 101

Theuer D, Verres R, Martin E und Büchler M W: Der „Gute Arzt“ – Über einen ethisch begründeten ärztlichen Umgang mit chirurgischen Patienten. In: Hax P-M, Hax-Schoppenhorst (Hrsg.): Kommunikation mit Patienten in der Chirurgie. Praxisempfehlungen für Ärzte aller operativen Fächer. W. Kohlhammer, Stuttgart 2011, 17 – 36

Tudiver F, Rose D, Banks B, Pfortmiller D (2009): Reliability and validity of an evidence-based medicine OSCE station. Fam Med 41, 89 – 91

Turner J L, Dankoski M E (2008): Objective Structured Clinical Exams: A Critical Review. Fam Med 40, 574 – 578

Van der Vleuten C P M (1996): The Assessment of Professional Competence: Developments, Research and Practical Implications. Adv Health Sci Educ Theory Pract 1, 41 – 67

Vivekananda-Schmidt P, Lewis M, Coady D, Morley C, Kay L, Walker D, Hassell A B (2007): Exploring the Use of Videotaped Objective Structured Clinical Examination in the Assessment of Joint Examination Skills of Medical Students. Arthritis Rheum 57, 869 – 876

Vyas R, Supe A (2008): Multiple choice questions: A literature review on the optimal number of options. Natl Med J India 21, 130 – 133

Wallace P (1997): Following the Threads of an Innovation: The History of Standardized Patients in Medical Education. Caduceus 13, 5 – 28

Wirtz M, Caspar F: Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Hogrefe Verlag, Göttingen 2002

Woodward-Kron R, Fraser C, Pill J, Flynn E (2015): How we developed Doctors Speak Up: an evidence-based language and communication skills open access resource for International Medical Graduates. Med Teach 37, 31 – 33

Yudkowsky R, Alseidi A, Cintron J (2004): Beyond fulfilling the core competencies: An objective structured clinical examination to assess communication and interpersonal skills in a surgical residency. Curr Surg 61, 499 – 503

Zylka Menhorn V (2013): Reden ist Gold, aber… Dtsch Ärztebl 110, A743

Danksagung

Einen besonderen Dank möchte ich an meine Betreuerin Frau Prof. Dr. med. Sarah König für Ihre engelsgleiche Geduld mit mir richten. Weiterer Dank geht an die Klinik für Allgemein-, Viszeral- und Kinderchirurgie – insbesondere Frau Koch – für die Unterstützung bei der Organisation und an das STÄPS für den technischen und räumlichen Background.