

**Aus der Klinik für Allgemeine, Unfall-, und Wiederherstellungschirurgie
der Ludwig-Maximilians-Universität München**

Vorstand: Prof. Dr. Wolfgang Böcker



**Evaluation chirurgischer Kompetenz am Beispiel einer
Mixed-Reality-basierten Trainingssimulation für perkutane Vertebroplastie (PV)**

Dissertation zum Erwerb des Doktorgrades der Zahnmedizin

an der Medizinischen Fakultät

der Ludwig-Maximilians-Universität zu München

vorgelegt von

Florinda Bierling

aus München

2021

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

Berichterstatter:

Prof. Dr. med. Ekkehard Euler

Mitberichterstatter:

Prof. Dr. Andreas Fottner

Priv. Doz. Dr. Volker Braunstein

Mitbetreuung durch den

promovierten Mitarbeiter:

Dr. med. Anna Maria von der Heide

Dekan:

Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung:

26.03.2021

Eidesstattliche Versicherung

Bierling, Florinda

Name, Vorname

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

**Evaluation chirurgischer Kompetenz am Beispiel einer
Mixed-Reality-basierten Trainingssimulation für perkutane Vertebroplastie (PV)**

selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht habe und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorliegende Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 04.04.2021

Florinda Bierling

Ort, Datum

Unterschrift Doktorandin

Inhaltsverzeichnis

1	Einleitung	1
1.1	Theoretische Grundlagen	1
1.2	Perkutane Vertebroplastie	1
1.3	Simulationsbasiertes Training in der Medizin	4
1.4	Definition des Virtual-Reality-Begriffes	6
1.5	Forschungsvorhaben	8
1.6	Gütekriterien wissenschaftlicher Messinstrumente	9
1.6.1	Objektivität	10
1.6.2	Reliabilität	11
1.6.3	Validität	11
1.7	Hypothese und Zielsetzung der Arbeit	12
2	Material und Methoden	14
2.1	Aufbau der PV-Trainingssimulation	16
2.2	Konzeption des BGDV-Fragebogens	20
2.2.1	Gegenüberstellung mit dem Prototyp des TSA-Bogens	21
2.2.2	Gliederung, Struktur und Inhalt des BGDV-Fragebogens	23
2.3	Konzeption des PV-Scores	28
2.4	Datenakquise	36
2.4.1	PV-Trainingssimulation	36
2.4.2	BGDV-Fragebogen	36
2.4.3	PV-Score	38
2.4.4	Validierungsmethoden Kriteriums- und Konstruktvalidität	38
2.5	Datenspeicherung und Statistik	40
3	Ergebnisse	43
3.1	Ergebnisse des BGDV-Fragebogens	43
3.1.1	quantitative Teilaufgaben	43
3.1.2	qualitative Teilaufgaben	48
3.1.3	simulationsbasierte Teilaufgaben	50
3.1.4	Abbildung des PV-Scores	54

3.2	Ergebnisse des PV-Scores	57
3.3	Ergebnisse der Validierungsmethoden	61
3.3.1	Kriteriumsvalidität	61
3.3.2	Konstruktvalidität	63
4	Diskussion	68
4.1	Teilnehmerzahl und Zeitraum	68
4.2	Interpretation der Ergebnisse	69
4.2.1	Interpretation der Ergebnisse des BGDV-Fragebogens.....	69
4.2.1.1	quantitative Teilaufgaben	69
4.2.1.2	qualitative Teilaufgaben.....	76
4.2.1.3	simulationsbasierte Teilaufgaben	78
4.2.2	Interpretation der Ergebnisse des PV-Scores.....	79
4.2.3	Interpretation der Ergebnisse der Validierungsmethoden	83
4.3	Limitationen und mögliche Modifikationen des PV-Scores der vorliegenden Arbeit	85
4.4	Schlussfolgerungen und Zukunftsausblicke.....	91
5	Zusammenfassung	96
6	Literaturverzeichnis	99
7	Anhang.....	106
7.1	BGDV-Fragebogen mit Fragebogenanhang.....	106
7.2	TSA-Bogen	111
7.3	textbezogener Anhang, weitere theoretische Grundlagen.....	115
7.4	Abkürzungsverzeichnis	118
7.5	Abbildungsverzeichnis	120
7.5.1	Tabellen	120
7.5.2	Abbildungen.....	121
8	Danksagung.....	123

1 Einleitung

1.1 Theoretische Grundlagen

Die vorliegende Arbeit konnte durch ein Zusammenwirken von Vertretern¹ der Informatik und der Medizin, im Rahmen der interdisziplinären Forschungsgruppe NARVIS (Navigation and Augmented Reality Visualization System) entstehen.

Den informatischen Anteil stellen Wissenschaftler des Lehrstuhls für Informatikanwendungen in der Medizin & Augmented Reality unter der Leitung von Prof. Dr. N. Navab der Technischen Universität (TU) München, während der medizinische Anteil durch ein Ärzteteam der Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie der Ludwig-Maximilians-Universität (LMU) repräsentiert und von Prof. Dr. E. Euler geleitet wird. (1) (2) Diese Kooperation wird durch Ingenieure vervollständigt und verfolgt das Ziel, durch die Entwicklung medizintechnischer Technologien und Innovationen eine patientensicherere und effektivere Diagnostik und Behandlung zu ermöglichen.

1.2 Perkutane Vertebroplastie

Im Zentrum der Thematik dieser Arbeit steht die perkutane Vertebroplastie (PV).

Diese minimal-invasive, bildgesteuerte Behandlungsmethode wurde erstmals 1984 von dem Neuroradiologen Deramond und dem Neurochirurgen Galibert in Frankreich eingesetzt, damals zur Behandlung eines schmerzhaften, aggressiven Wirbelkörperhämangioms der Halswirbelsäule. (3) Mittlerweile liegen die Hauptindikationen dieses Verfahrens jedoch in traumatisch oder pathologisch (osteoporotisch oder tumorbedingt) ausgelösten Frakturen oder fraktur- oder metastasenbedingten schmerzhaften Wirbelkörperprozessen. (4) Die Prävalenz Osteoporose-assoziiierter Wirbelkörperfrakturen liegt in Deutschland bei rund 250.000/Jahr. (5)

Die häufigste Indikation einer PV ist akuter und subakuter Schmerz bei Patienten mit einer Wirbelkörper-Kompressionsfraktur, bei denen eine konservative vier- bis sechswöchige Therapie aus Schmerz- und Physiotherapie nicht angeschlagen hat. Weltweit sind jährlich 1,4 Millionen Menschen von Wirbelkörper-Kompressionsfrakturen betroffen. Die Wahrscheinlichkeit, im Laufe des Lebens von einer Wirbelkörper-Kompressionsfraktur betroffen zu sein, liegt für das weibliche Geschlecht bei 16% und für das männliche bei 5%, wobei Osteoporose den größten ätiologischen Faktor darstellt. (6) Ziel des Eingriffes ist, neben der Stabilisierung des Wirbelkörpers, eine Schmerzreduktion für den Patienten. Diagnostische Bildgebung im Sinne von radiologischer Basisuntersuchung mit

¹ VertreterInnen: In der vorliegenden Arbeit wird aus Gründen der besseren Lesbarkeit im Folgenden ausschließlich die männliche Form verwendet. Sie bezieht sich auf Personen beiderlei Geschlechts. Experten/Innen, Teilnehmer/Innen, Chirurg/Innen, etc.

konventionellen Röntgenaufnahmen, in Kombination mit Schnittbilduntersuchungen zur weiteren Therapieplanung sind Voraussetzung für die Durchführung einer PV. Abhängig von der Kooperationsfähigkeit des Patienten, sowie von Anzahl und Lokalisation der betroffenen Wirbelkörper, wird die PV in Lokalanästhesie, ggf. mit Sedierung, oder in Allgemeinanästhesie durchgeführt. (4) (6)

In Bauchlage des Patienten wird der betroffene Wirbelkörper unter fluoroskopischer Überwachung bilateral über einen transpedikulären² Zugang mit einem Trokar punktiert. Über diesen minimalinvasiven Zugang wird im weiteren Verlauf Polymethylmetacrylat (PMMA)-Knochenzement³ in den Wirbelkörper injiziert. Dabei wird durch eine gute Auffüllung des Wirbelkörpers eine Stabilisierung desselben und durch die zyto-, neuro- und thermotoxischen Eigenschaften des Zements eine Schmerzreduktion beim Patienten angestrebt. Die Verletzung von z.B. nervalen Strukturen durch den Trokar oder Leckagen des Zements in Nachbarstrukturen oder -gefäße sind dabei zu vermeiden, bzw. zu minimieren.

Die Komplikationsrate gilt bei regelrechter Durchführung der PV als niedrig. Laut der amerikanischen „Society of Interventional Radiology“, sollte die Rate aller Komplikationen mit passageren und permanenten Folgen bei einer Osteoporose-bedingten Behandlung unter einem Grenzwert von 1% liegen. Bei einer durch einen malignen Tumor ausgelösten PV beträgt der Schwellenwert für passagere neurologische Defizite 10%. (4) (7)

Unter den Komplikationsrisiken zählt die Zementleckage als häufigste. Die Häufigkeit dieser meist asymptomatisch verlaufenden Leckagen bei wird mit Werten von 30% bis hin zu Höchstwerten von bis zu 90% angegeben, der Durchschnittswert liegt bei 41%. (4) (8) (9)

Zuallermeist handelt es sich dabei um inapparent verbleibende Leckagen in kleine Venen oder Intervertebrälräume. Sehr selten kann das Austreten von Zement jedoch auch klinische Auswirkungen zeigen. Die häufigste Erscheinungsform klinisch apparenter Zementleckage (in rund 1%, bzw. 5% bei der Behandlung von Osteoporose bzw. von Tumoren) ist die vorübergehende thermische Irritation der Spinalnervenzwurzeln in perineuralen venösen Gefäßen am Foramen intervertebrale durch die Polymerisationstemperatur des Zements. (4)

Eine größere Zementleckage kann zu Spinalnervenzwurzelskompressionen führen. Bei Auftreten im

² Es wird an dieser Stelle das Vorgehen an einem Lendenwirbelkörper (LWK) beschrieben, da die PV-Trainingssimulation der vorliegenden Arbeit einen LWK2 behandelt. Alternative Zugänge, bzw. unilaterales Vorgehen sind daher nicht in den Ausführungen enthalten.

³ Es liegen neuere Komposite sowie Kalziumphosphat-Zemente vor, dennoch wird PMMA weiterhin am häufigsten in der Behandlung von osteoporose- und metastasenbedingten Wirbelkörperfrakturen angewendet. (6)

Spinalkanal kann sie durch Myelonkompression oder Durchblutungsstörungen des Rückenmarks in der Folge eine (ggf. bleibende) Querschnittslähmung bewirken. Blasen-Mastdarm-Inkontinenz, Impotenz, Gefühlsstörungen oder chronische Schmerzen können weitere mögliche klinische Folgen darstellen. Kommt es zum Auftreten dieser seltenen Komplikationen, muss ggf. eine weitere Operation (OP) veranlasst werden. In seltenen Fällen gerät der Knochenzement in Venen oder Arterien, von wo er durch Mitschwimmen im Blutstrom in die Lunge oder andere Organe (z.B. über Niere-, Darm- oder Beingefäße) transportiert werden kann. Hier kann es durch Zementembolien zu entsprechenden Durchblutungsstörungen oder Kreislaufproblemen kommen, welche ggf. Organverlust oder den Tod herbeiführen können.

Neben den Risiken durch Zementleckage kann die Punktion des Wirbelkörpers durch den Trokar zu Blutungen, Infektionen oder zur Schädigung des Rückenmarks führen. Das Auftreten dieser Komplikationen kann ebenso das Einleiten einer weiteren OP erforderlich machen. (4) (10) ⁴

Die transpedikuläre Punktion des Wirbelkörpers ist neben der PV ebenso der erste operative Schritt bei der dorsalen transpedikulären Schraubenplatzierung zur Montage eines stabilisierenden Fixateur interne, der z.B. bei Wirbelsäulen-Frakturen zum Einsatz kommt.

Vor einigen Jahren war für chirurgische Tätigkeiten noch das Trainieren der Eingriffe an Kadavern oder auch am Lebeltier verbreitet. (11) (12) (13) (14) Derzeitig bestehen konventionelle Übungsmöglichkeiten der PV für in der Ausbildung befindliche Wirbelsäulenchirurgen in der Teilnahme an Operationen, wo sie erfahrenen Kollegen assistieren (clinical observation). Dabei werden zunächst das nötige Wissen sowie die dazugehörigen praktischen Fähigkeiten erlernt, welche später unter Anleitung eines Facharztes in selbst durchgeführten Operationen angewendet werden können (hands-on training). (8) (14) Erweiternde Übungsalternativen stehen in Form von elektronischen und computerbasierten Lernmitteln zur Verfügung. (15) Der zusammenfassende Leitspruch dieser konventionellen Trainingsmethoden, der auf den Chirurgen William Stewart Halsted (1890) zurückzuführen ist, lautet: „See one, do one, teach one“. (11) (16)

⁴ Weitere patientenspezifische Risiken (z.B. Allergien oder Blutdruck-Abfall durch den Zement) werden an dieser Stelle nicht erläutert, da sie für die vorliegende Arbeit nicht von Bedeutung waren.

1.3 Simulationsbasiertes Training in der Medizin

Simulatortraining (simulation-based training, SBT) stellt wertvolle Ergänzungsmöglichkeiten – bzw. sogar einen möglichen zukünftigen Ersatz – der konventionellen Trainingsmethoden dar.

Nachdem Kritik an den konventionellen Trainingsmethoden - hauptsächlich im Sinne der Patientensicherheit - laut wurde, nehmen mittlerweile Operationssimulatoren im Rahmen alternativer Trainingsmöglichkeiten eine wichtige Rolle in der medizinischen Ausbildung ein. Diese werden in einigen Teilgebieten der Medizin bereits routinemäßig zu Trainingszwecken oder als Assessment-Tool medizinischer Fähigkeit eingesetzt. Vorzugsweise in den Bereichen der Arthroskopie, Endoskopie, Gefäßchirurgie, Orthopädie, Ophthalmologie, Neurochirurgie, sowie Wirbelsäulenchirurgie, wo vermehrt minimal-invasive Eingriffe durchgeführt werden, kam es zu einer verstärkten Anwendung des SBT. In der im Zentrum der vorliegenden Arbeit stehenden Wirbelsäulenchirurgie wächst die Bedeutung des SBT stetig: „Simulators in spine surgery are increasingly introduced since many procedures involve high risks.“ (17) (vgl. Pfandler et al., S. 8, 2018) Die tatsächliche gegenwärtige Verbreitung der Anwendung des SBT in der Wirbelsäulenchirurgie ist jedoch nicht genau zu benennen. (18 S. 1352)

SBT kann dazu beitragen, die lange und beschwerliche Lernkurve, die minimal-invasive Eingriffe nach sich ziehen, zu reduzieren. (18) (19) Ein weiterer Vorteil liegt in der Risikofreiheit der Methode, da es bei einem Fehlverhalten des behandelnden Chirurgen nicht zur Gefährdung eines realen Patienten kommt. (20) Überdies ist die Simulationssoftware in der Lage, Daten zu speichern und in Bezug auf verschiedene Parameter wie beispielsweise Zeit, zurückgelegte Wegstrecke des Instrumentes, Fehlerscores oder Ökonomie der Bewegungen, auszuwerten. Somit kann auf einen erfahrenen Beobachter verzichtet werden, der die Leistung des trainierenden Chirurgen überwachen muss, wie dies bei konventionellen Trainingsmethoden nötig ist. Da es sich um ein simuliertes Prozedere handelt, bei der der Druck und die Anspannung einer realen OP weitgehend wegfallen, hat der Chirurg postoperativ die Möglichkeit, in entspannter Atmosphäre mithilfe der Simulationsdaten eine Analyse seiner Behandlungsabläufe vorzunehmen. Die Übertragbarkeit der im SBT erworbenen Fähigkeiten in reale Operationsumgebungen, sowie deren Anwendbarkeit und Nutzen sind bestätigt worden. (14) Die Verfügbarkeit der Simulatordaten erleichtert zudem die Möglichkeit, weltweit Expertenrat einzuholen und kann somit eine vernetzende Funktion einnehmen. (21)

Studien zufolge kann SBT das Sicherheitsprofil von Anfängern vergrößern, indem es die Spalte zwischen sicherer Aneignung chirurgischer Fertigkeiten und effektiver operativer Performance überbrückt. (22) Doch nicht nur Chirurgen in der Ausbildung können durch SBT im Sinne von

Zugewinn an Sicherheit, besserer Orientierung und Präzision der Behandlungsabläufe profitieren. Auch für bereits erfahrene Wirbelsäulenchirurgen bietet SBT Blickwinkel, die an der Wirbelsäule des realen Patienten nicht möglich sind und somit Vorteile durch Vertiefung und Erweiterung des bisherigen Kenntnisstandes. Es besteht zudem die Möglichkeit, Teilschritte einer OP mehrfach durchzuführen, um bestimmte Fähigkeiten gezielt zu trainieren. 3D-gedruckte, auf CT-Datensätzen realer Patienten beruhende Wirbelkörpermodelle, lassen sich kostengünstig nach Bedarf reproduzieren. Bei genügend Zeit bis zum endgültigen Eingriff besteht sogar die Möglichkeit, die OP anhand der patientenspezifischen Daten im Vorhinein einzuüben. (12)

Zudem kann durch die Möglichkeit, eigens ausgewählte 3D-gedruckte Wirbelkörpermodelle zu bearbeiten, die Behandlung anatomischer Besonderheiten wie spezifischer Pathologien, Deformitäten oder Frakturen gezielt trainiert werden. Dies stellt beispielsweise einen großen Vorteil zum konventionellen Training am Kadaver dar, an dem meist keine relevanten Pathologien vorhanden sind. (14) (13) (12)

Die zu beobachtende Tendenz von konventionellen Trainingsmethoden hin zu SBT rückt ein kompetenzbasiertes und problemorientiertes Lernen in den Vordergrund, bei dem das erwünschte Outcome des Trainings fokussiert wird. Im traditionell zeit-, prozess- und struktur-basierten Lernen der konventionellen Trainingsmethoden hingegen orientiert sich die Gestaltung des Trainings an spezifischen Inhalten, welche über eine vordefinierte Zeitspanne hinweg ausgeführt werden.

Tabelle 1 zeigt die grundlegenden Unterschiede zwischen den Bestandteilen konventioneller und alternativer Trainingsmethoden auf. (23) (24) Das folgende Kapitel liefert einen Einblick in die Integrationsformen der Virtualität im SBT.

A Comparison of the Elements of Structure- and Process-based vs. Competency-based Educational Programs		
Variable	Educational Program	
	Structure- and Process-based	Competency-based
Driving force for curriculum	Content-knowledge acquisition	Outcome-knowledge application
Driving force for process	Teacher	Learner
Path of learning	Hierarchical (teacher → student)	Non-hierarchical (teacher ↔ student)
Responsibility for content	Teacher	Student and teacher
Goal of educational encounter	Knowledge acquisition	Knowledge application
Typical assessment tool	Single subjective measure	Multiple objective measures
Assessment tool	Proxy	Authentic (mimics real tasks of profession)
Setting for evaluation	Removed	„In the trenches“ (direct observation)
Evaluation	Norm-referenced	Criterion-referenced
Timing of assessment	Emphasis on summative	Emphasis on formative
Program completion	Fixed time	Variable time

Tabelle 1: Vergleich von struktur- und prozessbasierten (≙ konventionellen) mit kompetenzbasierten (≙ alternativen/simulationsbasierten) Trainingsprogrammen (Quelle: in Anlehnung an Carraccio et al., 2002, (22 S. 362))

1.4 Definition des Virtual-Reality-Begriffes

Der Einsatz von Virtual-, Mixed- oder Augmented-Reality-Bestandteilen (VR, MR, AR) im Gesundheitswesen lässt sich als Medical VR bezeichnen. Unter diese Kategorie fällt auch die Trainingssimulation der vorliegenden Arbeit, an der sich eine PV in realistischer Operationsumgebung mit Elementen der MR durchführen lässt. (s. 2.1)

Man kann die drei Begriffe der VR, MR und AR als Zwischenstufen zwischen den beiden Polen „reine Realität“ und „reine Virtualität“ auffassen.

Die VR generiert dem Benutzer reale Umwelteindrücke als computergenerierte, dreidimensionale Welten. Dabei können verschiedene Sinneseindrücke, wie beispielsweise visuelle oder auditive, kombiniert einwirken, oder auch nur ein einzelner Sinneseindruck betroffen sein. Diese digital erzeugten Eindrücke simulieren einem Benutzer, der in Echtzeit mit ihnen interagiert, dass er sich in der realen Welt befinde. (18) (25) Die Fokussierung der Interaktion auf dem Pol der „reinen Virtualität“ bedingt in Ermangelung von Realität die Tatsache, dass VR nicht in realen

Operationen angewendet werden kann. Für Behandler stellt es eine Schwierigkeit dar, komplett in die virtuell generierte Situation einzutauchen. (26)

AR unterscheidet sich von VR dahingehend, dass der Fokus der Interaktion in der realen Welt liegt. AR erweitert sozusagen die Realität durch Integration zusätzlicher Informationen oder graphischer Elemente in die Umgebung des Benutzers in Echtzeit. Der praktische Nutzen von AR für chirurgische Anwendungen kann dahingehend problematisch sein, dass die meisten AR-basierten Lösungen komplizierte sowie sperrige Navigationssysteme und Geräte erfordern. (26)

MR-basierte Umgebungen ermöglichen dem Benutzer, sowohl die reale Außenwelt, als auch digitale, virtuelle Elemente, die über Displays präsentiert werden, wahrzunehmen.

Diese Technologie generiert dem Benutzer die Illusion, dass digitale virtuelle Elemente und physische reale im selben Bereich existierten. Praktischen Nutzen im chirurgischen Umfeld erfährt MR im Rahmen des chirurgischen Trainings, präoperativer Planung, sowie intraoperativer Orientierungshilfe. Im Vergleich mit traditionellen Methoden präsentiert sich die MR-Technologie als Kosten-Nutzen-effektiv. (26)

Ein weiterer wichtiger Begriff im Rahmen dieser medizinisch-informatischen Schnittstelle ist die Bezeichnung CAS (Computer Assisted Surgery) - bzw. spezifischer: Computer Assisted Orthopedic Surgery (CAOS). Darunter fasst man eine große Bandbreite an Systemen zusammen, bei der standardisierte Operationsmethoden (Schnittpunkt Medizin) mit modernen Technologien der Bildgebung und Navigation (Schnittpunkt Informatik) verknüpft werden. Zu Deutsch als Computerassistierte Chirurgie bekannt, bietet die CAS seit den 1990er Jahren in verschiedenen chirurgischen Fachdisziplinen weit gefächerte Möglichkeiten, angefangen von der virtuellen Operationsplanung, über Navigation und Assistenz bis hin zu ferngesteuerter Telechirurgie. (27)

1.5 Forschungsvorhaben

Limitierend bei dem Feedback durch analysierte Daten der Software einer Trainingssimulation wirkt sich momentan noch die Tatsache aus, dass zwar ein Überblick über die Fähigkeiten des Chirurgen in Bezug auf die einzelnen Parameter gewährt wird, es jedoch schwierig ist, daraus ein Resümée über die Gesamtpformance des Chirurgen zu ziehen. (21)

Diese Tatsache kombiniert mit der generellen Tendenz von konventionellen Trainingsmethoden hin zu SBT bewirkt einen erhöhten Bedarf an der Einführung wissenschaftlicher Messinstrumente, die es ermöglichen, chirurgische Kompetenz durch objektive Gesamtergebnisse möglichst nachvollzieh- und vergleichbar auszudrücken. Auch der Bedarf an einer Informationsverdichtung wächst, da nicht nur in der Medizin seit Jahrzehnten eine beständige Zunahme der Flut an verfügbarer Information zu beobachten ist. Verschiedene Studien kamen zu dem Schluss, dass ein Bedarf an wissenschaftlichen Messinstrumenten - speziell für die Bewertung chirurgischer Kompetenz in MR-basierten Operationsumgebungen wie der vorliegenden PV-Trainingssimulation - besteht. (12)

Dieser Bedarf an Quantifizierung, Objektivierung und Operationalisierung medizinischer Information kann anhand der Konzeption wissenschaftlicher Messinstrumente wie Fragebögen, klinischen Scores, Skalen oder Indexen gelingen. (28)

Das Forschungsvorhaben der vorliegenden Arbeit greift diesen Bedarf auf.

Es beinhaltet in einem experimentellen, prospektiven Teil die Konzeption sowie Durchführung eines Fragebogens (BGDV-Fragebogen) über relevante Bestandteile einer PV. Anhand dessen Auswertung erfolgt die Konzeption eines Scores (PV-Score) für die Bewertung chirurgischer Kompetenz. Der statistische, retrospektive Studienanteil beschreibt die beispielhafte Anwendung des konstruierten PV-Scores auf die ausgewerteten Daten einer MR-basierten PV-Trainingssimulation. Die klinische Studie, welche Teil des Forschungsvorhabens, jedoch nicht mehr Teil der vorliegenden Arbeit ist, besteht aus einem lang angelegten Prozess, welcher durch die klinische Anwendung des PV-Scores evaluieren soll, ob ein routinemäßiger Einsatz und eine mögliche Kommerzialisierung dieses Bewertungstools realisierbar sind.

Im Rahmen der Systemabgrenzung der Forschungsfrage wurde die vorliegende Studie als Pilotprojekt definiert. Gemäß der Natur eines Pilotprojektes sollten die Rahmenbedingungen des Forschungsvorhabens einerseits klar formuliert werden, um ausreichend Anhaltspunkte für relevante Forschungsinhalte zu bieten. So wurde das Forschungsvorhaben auf einen Zeitraum von zwei Jahren angelegt, in welchen es sich auf ein begrenztes Teilnehmerkollektiv, sowie ein

begrenztes Forschungsgebiet⁵ bezieht. Innerhalb dieser definierten Forschungsumstände soll das Pilotprojekt der Evaluation dienen, ob sich ein derartiges Vorgehen auch in einem größeren Rahmen resp. zukünftiger Forschung als sinnvoll erweisen kann.

Andererseits wurde innerhalb der genannten Rahmenbedingungen eine flexible Gestaltung der Forschungsfrage und ein exploratives Vorgehen befürwortet und praktiziert, um genügend Spielraum zu lassen, auf etwaige Veränderungen im Laufe der fortschreitenden Forschungsarbeit eingehen zu können.

1.6 Gütekriterien wissenschaftlicher Messinstrumente

Das Einhalten bestimmter Anforderungen in Form sogenannter Gütekriterien bei der Konstruktion wissenschaftlicher (im vorliegenden Fall: medizinischer) Messinstrumente ist von fundamentaler Wichtigkeit. Indem sie überprüfen, dass die Konzeption von Fragebögen und Scores, sowie daraus getroffene Schlussfolgerungen nachvollziehbar und verlässlich bleiben, dienen sie der Aufrechterhaltung medizinischer Standards. Die Nichterfüllung dieser Gütekriterien kann ein Defizit wissenschaftlich überprüfter Grundlagen, bzw. notwendiger Kontrolluntersuchungen signalisieren. (29) (30) In der vorliegenden Arbeit wurde im ersten Schritt ein Fragebogen zum Generieren eines Expertenkonsensus konzipiert, während im zweiten Schritt die Konzeption eines Scores auf der Basis dieses Expertenkonsensus erfolgte. Da es sich dabei um verschiedene Messinstrumente handelt, sind die geforderten Ausprägungsformen der Gütekriterien nicht bei beiden gleich zu ermitteln. In den entsprechenden Textabschnitten der vorliegenden Arbeit wird auf die jeweiligen Gütekriterien eingegangen resp. werden Möglichkeiten zu deren empirischer Bestimmung aufgezeigt.

Im Folgenden werden die sog. Hauptgütekriterien Objektivität, Reliabilität sowie Validität erläutert, welche in einer aufeinander aufbauenden Beziehung zueinander stehen. Objektivität stellt die Voraussetzung für Reliabilität dar, welche ihrerseits die Voraussetzung für Validität ist. Jedes der Kriterien lässt sich in Teilaspekte untergliedern, welche im Folgenden vorgestellt werden. (31) (32)

⁵ Das Forschungsgebiet begrenzte sich auf die Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie der LMU Campus Innenstadt und Campus Großhadern.

1.6.1 Objektivität

Dieses - auch als „Anwenderunabhängigkeit“ bezeichnete - Gütekriterium beschreibt das Ausmaß, in dem das Untersuchungsergebnis unabhängig ist von jeglichen Einflüssen außerhalb der untersuchten Person. Die Objektivität eines Messinstrumentes fordert, dass die Messergebnisse nur von dem zu messenden Merkmal abhängig sind und nicht vom Beobachter, resp. Befrager oder Situationsvariablen beeinflusst werden. (31) (32)

Ein triviales Beispiel, um die Objektivität besser verständlich zu machen: beim Erfassen der Körpergröße einer Person sollte die Messung unabhängig von der Person sein, welche die Datenerhebung vornimmt. Die perfekte Objektivität eines Messinstrumentes ist gegeben, wenn die Messungen verschiedener Beobachter identisch sind.

Typischerweise lässt sich die Objektivität nochmals in drei Aspekte aufteilen.

- **Durchführungsobjektivität:** Sie betrifft die Konstanz der Untersuchungsbedingungen und lässt sich durch eine maximale Standardisierung der Untersuchungssituation gewährleisten. Dies gelingt i.d.R. bei selbstauszufüllenden Messinstrumenten besser als bei persönlichen, mündlichen und telefonischen Verfahren. Wenn es durch Störfaktoren in einer Untersuchungsumgebung nicht gelingt, alle befragten Personen unter vergleichbaren Bedingungen zu testen, kann die Durchführungsobjektivität beeinträchtigt sein. Störfaktoren können dabei durch Interviewereffekte, Reihenfolgeeffekte der Items eines Messinstrumentes, Anfälligkeit der Itembeantwortungen für momentane individuelle Stimmungen oder Unterbrechungen bei der Fragebogenbearbeitung sein.
- **Auswertungsobjektivität:** Sie besteht, wenn die Untersuchungsergebnisse unabhängig vom Testauswerter sind. Defizite im Bereich der Auswertungsobjektivität können im Besonderen bei der Analyse offener Antworten auftreten, oder auch durch Fehler bei der Dateneingabe bedingt sein, d.h. beim Abtippen oder Einscannen eines Fragebogens.
- **Interpretationsobjektivität:** Sie liegt vor, wenn die Interpretation der Untersuchungsergebnisse unabhängig von der Person ist, die diese vornimmt. Wenn die Untersuchungsergebnisse eines Messinstrumentes von verschiedenen Diagnostikern in gleicher Weise interpretiert werden, liegt demnach eine hohe Interpretationsobjektivität vor. Aus diesem Grund sind klare Interpretationshinweise, sowie eine genaue Beschreibung des zu erfassenden Konstrukts in der Dokumentation eines Messinstrumentes wichtig. Norm-, Bezugs- oder Vergleichswerte sorgen beispielsweise für die Gewährleistung der Interpretationsobjektivität. (31) (32) (33)

1.6.2 Reliabilität

Mit Reliabilität wird die Genauigkeit bezeichnet, mit der ein Messinstrument ein Merkmal messen kann und lässt sich auch als Replizierbarkeit oder Zuverlässigkeit seiner Messergebnisse beschreiben. Auch an dieser Stelle sei zum besseren Verständnis ein triviales Beispiel eingefügt: bei der Erfassung der Körpergröße einer Person ist die Reliabilität eines Messinstruments daran erkennbar, dass die wiederholte Messung der Person immer wieder die gleiche Zentimeterlänge ergibt. (31)

Um Reliabilität zu überprüfen, müsste man ein Messverfahren zu einem Zeitpunkt mit einer Person mehrmals durchführen und anschließend die dabei gewonnenen Ergebnisse miteinander korrelieren. Da praktisch bei mehrmaliger Messung in einem eng begrenzten Zeitraum keine unabhängige Beantwortung der Items möglich ist, ist der Idealfall der Reliabilität schwer zu erreichen. Daher hat man sich vier verschiedene Methoden zur Bestimmung der Reliabilität zueigen gemacht: die Retest-, Paralleltest- und die Testhalbierungsreliabilität, sowie Konsistenzanalysen. Wann welches Verfahren zur Bestimmung der Reliabilität am sinnvollsten ist, hängt von den Gegebenheiten der jeweiligen Messumstände ab. Für die Güte eines Fragebogens stellt die Reliabilität in der Regel ein stärkeres Kriterium dar als die Objektivität. (33) (32)

1.6.3 Validität

Diese bezeichnet den Grad der Genauigkeit, mit der ein Messverfahren genau dasjenige Konstrukt erfasst, das es messen soll. Der Nachweis, dass ein Verfahren genau das zu messende Merkmal erfasst, garantiert die Validität des Verfahrens. Dabei kann es durchaus vorkommen, dass ein Verfahren objektiv und reliabel, aber nicht valide ist. Zum besseren Verständnis dient folgendes Beispiel: ein Thermometer zeigt objektiv und reliabel einen Wert an (d.h. unabhängig davon, welche Person es abliest, zeigt es immer wieder zuverlässig denselben Wert an), jedoch nicht die Raumtemperatur in Grad Celsius, für die es vorgesehen wurde, sondern beispielsweise einen Wert in Kilogramm.

Somit sind Messergebnisse nur bei einem validen Test interpretierbar. Dies führt dazu, dass die Validität als das wichtigste Testgütekriterium bezeichnet wird. (31) (32)

Dem Validitätsnachweis dienen verschiedene, sich ergänzende Validierungskonzepte, wobei im Normalfall zwischen drei Arten unterschieden wird:

- **Kontent-/Inhaltsvalidität:** Sie bezeichnet die inhaltliche Analyse eines Messverfahrens, um festzustellen, ob dessen Itempool das zu messende Konstrukt auch tatsächlich umfassend genug repräsentiert. Daher setzt die Inhaltsvalidität im Normalfall bereits in der

Anfangsphase der Konstruktion eines Messinstrumentes an. Um die Inhaltsvalidität eines Verfahrens zu überprüfen, wird z.B. untersucht, wie die Auswahl der Items vorgenommen wurde, ob also beispielsweise ein Expertengremium zur Überprüfung ihrer Gültigkeit eingesetzt wurde (s. 2.3 Konzeption des PV-Scores). (32)

- Kriteriumsvalidität: Diese stellt fest, ob das zu messende Konstrukt mit sog. Außenkriterien übereinstimmt, welche dasselbe Konstrukt messen. Diese Außenkriterien bezeichnen praktisch relevante, andere Kriterien - wobei es eine Schwierigkeit darstellen kann, für die jeweilige Situation angemessene Außenkriterien zu finden (s. 2.4.4 Validierungsmethoden Kriteriums- und Konstruktvalidität). (31) (32)
- Konstruktvalidität: Sie dient der Überprüfung, inwieweit das Messinstrument das zu erfassende Konstrukt misst. Das Konstrukt (im vorliegenden Fall: die chirurgische Kompetenz) ist dabei als Gedankenkonstrukt zu verstehen, welches aus Überlegungen und Erfahrungen abgeleitet wurde, um ein beobachtetes Verhalten zu erklären.

Es existieren zahlreiche Methoden, um die Konstruktvalidität eines Messinstrumentes zu prüfen. Eine verbreitete Methode besteht im Vergleich zwischen dem neu konzipierten Verfahren mit einem bereits etablierten Messinstrument, welches das gleiche, oder zumindest ein stark verwandtes Konstrukt erfasst (s. 2.4.4 Validierungsmethoden Kriteriums- und Konstruktvalidität). (21) (32)

1.7 Hypothese und Zielsetzung der Arbeit

Ziel des vorliegenden Pilotprojektes ist die Operationalisierung des Konstruktes „chirurgische Kompetenz“ kraft eines wissenschaftlichen Messinstrumentes. Der Kompetenzbegriff soll sich dabei auf die technischen Fähigkeiten eines Chirurgen bei der Durchführung einer PV beziehen.

Der experimentelle, prospektive Studienanteil soll über die Konzeption eines Fragebogens zu dem angestrebten Messinstrument in Form eines Scores führen.

Folgende Gesichtspunkte sollen bei der Konzeption des Scores besondere Berücksichtigung finden:

- Integration leistungskritischer und relevanter Items einer PV
- Formulierung eines nachvollziehbaren und repräsentativen Score-Algorithmus
- Realisierung einer diskriminativen Resultatformulierung
- Praktikabilität (Benutzerfreundlichkeit und Bedienbarkeit) des Scores

Die Ziele der im Rahmen des retrospektiven Studienanteils praktizierten Anwendung des konzipierten Scores auf Daten eines begrenzten Teilnehmerkollektivs - in Form der Simulationsdaten der Teilnehmer einer MR-basierten PV-Trainingssimulation – liegen in folgenden Punkten:

- Realisierung einer Klassifizierung der Teilnehmer in „bestanden“ vs. „nicht bestanden“ im Sinne der diskriminativen Resultatformulierung
- Eruierung von möglichen Einflüssen, resp. Fehlerquellen auf das Score-Ergebnis

Im Rahmen des statistischen Studienanteils sollen die Score-Ergebnisse der Teilnehmer einem Vergleich mit den jeweiligen Berufserfahrungen der Teilnehmer, sowie mit den Ergebnissen eines vergleichbaren, bereits etablierten Evaluierungstools unterzogen werden. Die NARVIS-Forschungsgruppe stellte für diese Validierungsmethoden folgende Nullhypothesen H0 auf:

Bei der praktischen Anwendung des Scores auf die Simulationsdaten der Teilnehmer der PV-Trainingssimulation wird...

... kein Zusammenhang zwischen dem Score-Ergebnis und der Berufserfahrung eines Teilnehmers erwartet.

... kein Zusammenhang zwischen dem Score-Ergebnis eines Teilnehmers und dem Ergebnis eines vergleichbaren, bereits etablierten Evaluierungstools erwartet.

Optimierungen und alternative Herangehensweisen sollen im Laufe des Pilotprojektes möglich sein, ohne die Rahmenbedingungen zu überschreiten oder die dargestellten Ziele aus den Augen zu verlieren.

Mit der Konzeption des Scores ist die Hoffnung verbunden, durch Erkenntnisse im Bereich der „Operationalisierung chirurgischer Kompetenz, resp. Leistung“ einen Beitrag zur Verbreitung des SBT und im weiteren Sinne zur Qualitätssicherung und Patientenschutz liefern zu können.

2 Material und Methoden

Zur besseren Orientierung kann der Material- und Methoden-Teil nach zeitlichen und inhaltlichen Gesichtspunkten in die folgende Gliederung unterteilt werden:

- a) experimenteller, prospektiver Teil
 - a1) Aufbau der PV-Trainingssimulation (2.1)
 - a2) Resultat: Konzeption und Durchführung des BGDV-Fragebogens sowie Konzeption des PV-Scores (2.2 und 2.3)
- b) statistischer, retrospektiver Teil
 - b1) Anwendung des PV-Scores auf die Daten der PV-Trainingssimulation (2.4.3)
 - b2) Validierungsmethoden Kriterien- und Konstruktvalidität (2.4.4)

Abbildung 1 zeigt einen Überblick über den Aufbau und Ablauf des vorliegenden Pilotprojekts im Sinne einer Versuchsdesign-Grafik.

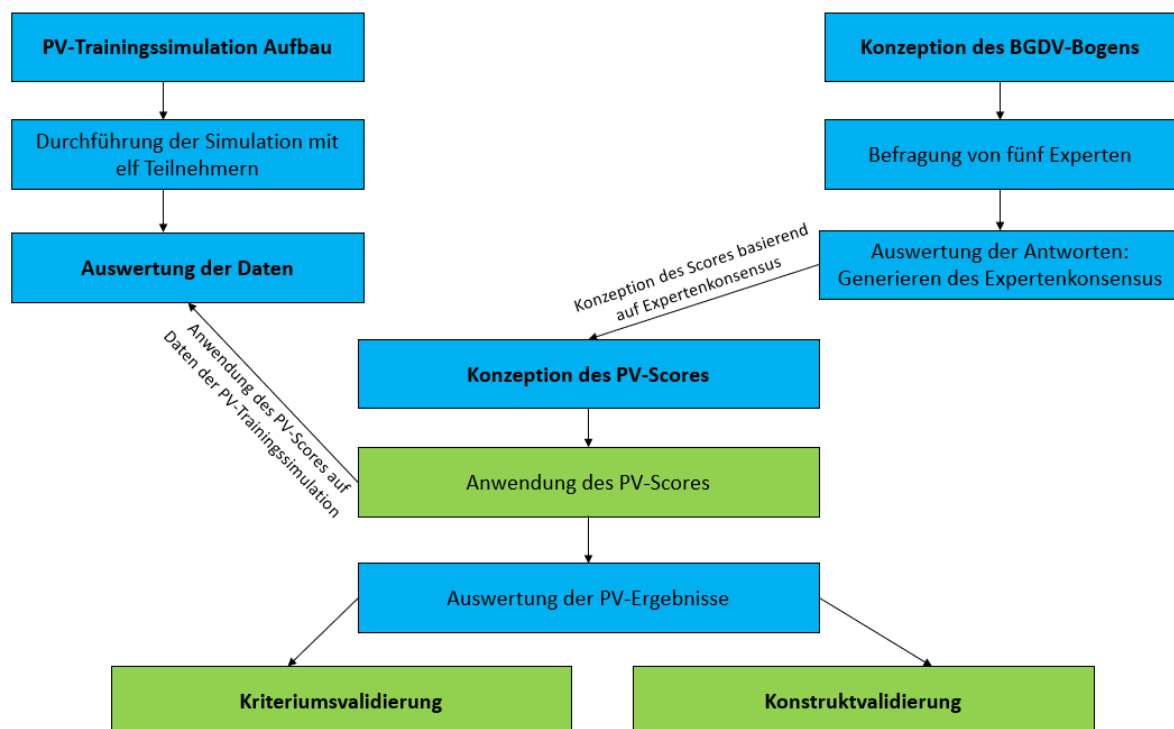


Abbildung 1: Versuchsdesign-Grafik, blau=prospektive Anteile, grün=retrospektive Anteile, eigene Darstellung

Im Folgenden wird das Vorgehen chronologisch beschrieben – beginnend vom Aufbau der PV-Trainingssimulation und der Konzeption des BGDV-Fragebogens und PV-Scores, über die Datenakquise bis hin zur abschließenden praktischen Anwendung des PV-Scores auf die Daten der PV-Trainingssimulation. Der Vollständigkeit halber sei angemerkt, dass die Schritte nicht immer streng zeitlich voneinander getrennt abliefen. So fand z.B. die Datenakquise, bzw. Auswertung des BGDV-Fragebogens stellenweise parallel zur Konzeption des PV-Scores statt, da diese aufbauend auf den ausgewerteten Ergebnissen des BGDV-Fragebogens erfolgte.

Das Vorgehen, welches im experimentellen, prospektiven Teil praktiziert wurde, beinhaltet eine Kombination verschiedener Herangehens- und Sichtweisen aus diversen Fachgebieten. Diese lässt sich unter dem Stichwort „Methodenpluralismus“ zusammenfassen und ist als solcher „als reflektierte und angepasste Auswahl zur Erkenntnis führender Methoden“ zu verstehen. (34) Dieser Methodenpluralismus stellt einen Aspekt sogenannter „Mixed-methods-Studien“ (MMS) dar. MMS bezeichnet einen Forschungsansatz, bei dem quantitative und qualitative Elemente kombiniert werden, so wie es im vorliegenden Studiendesign der Fall war: die Techniken reichten von Experteneinschätzungen durch Interviews und Fragebögen, Protokollen während einer Live-OP und Video- und Audioaufnahmen, über Literaturrecherche, bis hin zur Anwendung digitaler Simulationsdaten. (14)

Die Kombination der qualitativen und quantitativen Anteile dient dabei dem Zweck, einen möglichst breit gefächerten Erkenntnisgewinn zu erhalten. Sie kann sowohl innerhalb einer Untersuchung praktiziert werden, als auch in mehreren, aufeinander bezogenen Untersuchungen stattfinden. Letzteres wird als „Integration“ bezeichnet. Bei dieser - im Rahmen der MMS gerne praktizierten - Methode, dient eine der Studien als Grundlage für die Entwicklung der anderen. (34) (35)

Das Muster der „Integration“ wurde im vorliegenden Studiendesign angewandt: die Konzeption des PV-Scores erfolgte auf Basis der ausgewerteten Antworten des vorgeschalteten BGDV-Fragebogens. Einblicke in die im Laufe des Pilotprojekts angewandten „Mixed methods“ dieser pluralistischen Herangehensweise sollen dem Leser in den folgenden Kapiteln vermittelt werden.

2.1 Aufbau der PV-Trainingssimulation

Die dieser Arbeit zugrunde liegende PV-Trainingssimulation wurde vom Forschungsteam NARVIS, sowie Mitgliedern des ATMEOS-Teams (Assessment and Training of Medical Experts based on Objective Standards) im Sinne einer MR-basierten PV-Trainingssimulation entwickelt. Die realistisch und maßstabsgetreue konzipierte Operationsumgebung (full-scale team simulation environment, s. Abb. 6) diente der vorliegenden Arbeit als Single-User-Study, d.h. ausschließlich zur Analyse der operativen Leistung des Chirurgen. Das simulierte Setup beinhaltete jedoch den kompletten intraoperativen Behandlungsablauf einer PV für das gesamte OP-Team und diente verschiedenen Forschungsprojekten als Grundlage. (36) (37) (38) (12)

Das OP-Team bestand neben dem behandelnden Chirurgen aus einem Anästhesisten, einer OP-Schwester (sog. sterile Schwester) und einem Springer (sog. unsterile Schwester) (s. Abb. 2).

Die Rollen des OP-Teams wurden für die Studie von Personen mit medizinischem Hintergrund übernommen, welche sich während der gesamten Durchführung der OP an ein vordefiniertes Skript hielten. Die PV-Trainingssimulation beinhaltete ein synthetisches Patientenmodell (sog. Mannequin), einen stillgelegten C-Bogen, sowie einen Trokar, welche mit Tracking-Markern ausgestattet waren. Mithilfe dieser Tracking-Marker ließen sich ihre räumlichen Beziehungen mittels eines optischen Trackingsystems⁶ nachvollziehen. Aus den Informationen des Trackingsystems ließen sich so im Nachhinein verschiedene Metriken ableiten und untersuchen (s. Abb. 3).



Abbildung 2: OP-Team (anonymisiert): behandelnder Chirurg, sterile und unsterile Schwester, Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU

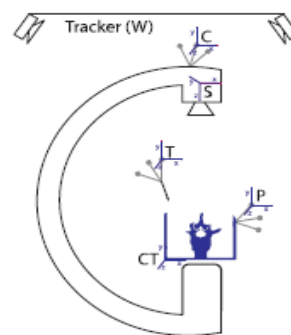


Abbildung 3: Schematisierte Abb. des Trackingsystems (=W), C-Bogen (=C), Patientenmodell (=P), virtuelle Kamera (=S), Instrument/Trokar (=T), Quelle: (11 S. 1337)

⁶ ARTTRACK2 4-camera optical outside-in

Im 3D-Druckverfahren wurde aus Computertomographieschnittbildern (CT-Bildern) einer 70-jährigen, 1,65 m großen, 97 kg schweren, adipösen Patientin ein synthetisches Knochenmodell eines Lendenwirbelsäulenabschnittes mit Osteoporose-bedingter Wirbelkörperfraktur mit Höhenminderung gewonnen. (s. Abb. 4 (b)) Dazu wurde im patientenspezifischen CT-Datensatz eine Segmentierung der Wirbelsäule vorgenommen und eine vierwandige Box darum herum generiert. Auf deren Oberfläche befanden sich 20 künstliche Orientierungsmarkierungen (landmark holes), die der Registrierung des gedruckten Knochenmodells im CT-Datensatz dienten.

Für die Segmentierung des Wirbelkörpers wurde ein Oberflächennetz (surface mesh) kreiert, welches dann der Anatomie entsprechend eingeebnet und per 3D-Druckverfahren in Polylactide-basierten Fasern (PLA) gedruckt wurde. Dabei wurden für eine möglichst realitätsnahe Darstellung der unterschiedlichen Dichte und Porosität des spongiösen und des kortikalen Knochenanteils zwei verschiedene Materialien verwendet.⁷

Bei dem verwendeten 3D-Drucker⁸ handelte es sich um einen Doppelsextruder⁹, welcher das Knochenmodell im Fused Filament Fabrication (FFF) – Verfahren¹⁰ herstellte. (13)

Diesem gedruckten Knochenmodell wurde eine optische „Tracking-Zielscheibe“ (tracking target) starr angefügt. Das so ausgestattete Knochenmodell wurde in das Mannequin eingebracht und im Folgenden mit rot gefärbtem Gelwachs und künstlicher Haut bedeckt, was eine weitgehend realistische Nachahmung des Schnitt- und Nahtverhaltens ermöglichen sollte (s. Abb. 4(d) und Abb. 5). (12)

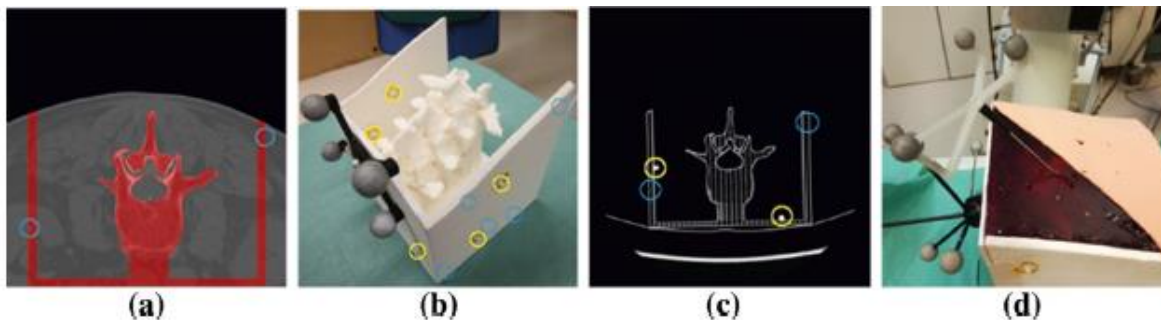


Abbildung 4: künstliche Orientierungsmarkierungen (blau und gelb): (a) im CT-Datensatz der Patientin, (b) im 3D- Knochenmodell, (c) im CT des 3D-Knochenmodells, (d) 3D-Knochenmodell in Gußform mit rotem Gelwachs, Quelle: (12 S. 1338)

⁷ EasyFil PLA (standard PLA, tensile modulus: 3.3 GPa, tensile strength: 110 MPa) und EasyWood PLA (PLA compound gravimetrically filled with 40% grinded wood, tensile modulus: 1.9 GPa, tensile strength: 71 MPa), beide von Formfutura BV, Nijmegen, Netherlands

⁸ Ultimaker2+ 3D Drucker, mit Zerstäuber-Durchmesser von 0,35mm und einer Schichtdicke von 0,15mm

⁹ Doppelsextruder: ermöglicht Drucken mit mehreren Farben und Materialien gleichzeitig

¹⁰ FFF: Fertigungsverfahren aus dem Bereich des Rapid Prototyping, bei dem schmelzfähiges, faserförmiges Plastik aufgeschmolzen und schichtweise auf eine Druckplatte aufgetragen wird. (83)



Abbildung 5: mit Kunsthaut überzogener OP-Situs des Mannequins, Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU



Abbildung 6: OP-Umgebung mit Instrumententisch, C-Bogen und Mannequin, Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU

Das synthetische Knochenmodell der Patientin wies einen Osteoporose-bedingt frakturierten Lendenwirbelkörper 2 (LWK2) auf. Es handelte sich bei der Fraktur des LWK2 um eine Deckplatten- Impressionsfraktur mit geringer Keilwirbelbildung und ohne Beteiligung der Wirbelkörper-Hinterkante. Epidemiologisch gesehen ist diese Fraktur besonders häufig, da über 50% aller Wirbelkörperfrakturen auf einem Kompressionsmechanismus basieren und der thorakolumbale Übergang (Thorakalwirbel Th10 – Lendenwirbel LW2) mit 67% die häufigste Frakturlokalisation darstellt. (39) (37)

Es konnten strahlenfrei simulierte Fluoroskopie-Röntgenbilder mittels eines funktionslosen mobilen einarmigen C-Bogens angefertigt werden, welche sowohl auf der Basis von CT-Datensätzen der realen Patientin, als auch auf der getrackten räumlichen Beziehung zwischen dem C-Bogen, dem Mannequin und dem Trokar beruhten (s. Abb. 7).

Die Methodik hinter diesem Generieren strahlenfreier Fluoroskopiebilder mittels CT-Datensätzen nennt sich Digitally Reconstructed Radiographs (DRR). (12)

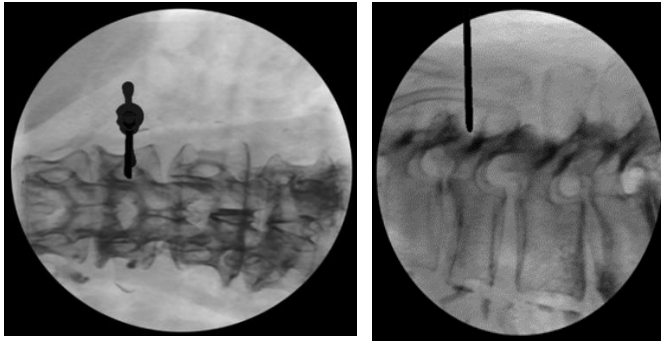


Abbildung 7: AP (anterior-posterior) und sag. (sagittal) Röntgen-Aufnahme des LWK2 mit Trokar, Quelle: (35 S. 276)



Abbildung 8: Instrumente (Tisch der sterilen OP-Schwester), Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU

Ein Injektionssystem ermöglichte ein präzises Messen der Knochenzement-Injektion. (37) Standardinstrumente der PV wie Skalpelle, Trokar, Hammer, Nahtmaterial und Haken, waren ebenfalls Teil der Simulation (s. Abb. 8).

Die Besonderheit im Aufbau dieser Operationsumgebung besteht darin, dass patientenbasierte, 3D-gedruckte Anatomie, in direkter Interaktion mit simulierter Bildgebung durch Fußpedal-Betätigung eines physisch präsenten, realen, aber nicht „strahlenden“ C-Bogens verwendet wurde.¹¹ Dieses Gefüge stellt eine Seltenheit im Vergleich zu anderen Simulationsumgebungen dar. Die Integration des realen C-Bogens stellt einen wichtigen Trainingsbestandteil für die Übertragbarkeit in reale Operationsumgebungen dar. Die realistische manuelle Interaktion mit dem 3D-gedruckten Patientenmodell und den realen Operationsinstrumenten ist eine wichtige Komponente im Trainieren der Hand-Augen-Koordination sowie in der mentalen Fähigkeit eines Chirurgen, die projizierten Röntgenbilder (2D) mit dem operativen Situs der realen Patienten-anatomie (3D) zu verbinden. (12)

¹¹ Die Darstellung des C-Bogens und seiner Betätigung reicht in Simulationsumgebungen in verschiedenen Realitätsgraden von komplett virtueller Realisierung über das Verwenden eines realen C-Bogens. (12)

2.2 Konzeption des BGDV-Fragebogens

Ein Team aus Wirbelsäulenchirurgen, Informatikern, Pädagogen, sowie Psychologen erarbeitete im Zeitraum von August bis November 2018 in den Räumlichkeiten des NARVIS-Forschungslabors einen Fragebogen, der den Titel „Fragebogen über die **Bewertung der Güte der Durchführung** einer Vertebroplastie zur Erstellung eines ‚Pedikel-Zugangs-Scores‘ “ (BGDV-Fragebogen)¹² trägt.

Zum Zeitpunkt der Konzeption des Fragebogens wurde der geplante Score noch unter der Bezeichnung „Pedikel-Zugangs-Score“ geführt. Im Laufe der fortschreitenden Forschungsarbeit und weiterführenden Zwischenergebnisse kristallisierte sich zunehmend heraus, dass diese Bezeichnung zu eng gefasst war, da nicht nur der Zugang zum Pedikel, sondern auch die weiterführenden Schritte einer PV in den Blickpunkt des Interesses gerückt werden sollten.

Bei der Konzeption des BGDV-Fragebogens ging es zunächst um die Eruierung von Teilaufgaben (TA) mit möglichst präzisiertem Informationsgehalt über die relevantesten Fähigkeiten und Anforderungen bei der Durchführung einer PV. Die angewandten Methoden zur Detektion dieser TA lassen sich der Bandbreite der Kognitiven Aufgaben-Analyse (Cognitive Task Analysis, CTA) zuordnen. Die CTA umfasst verschiedene Techniken, bzw. Forschungsmethoden. Beispielsweise wird durch Interview- und Beobachtungsstrategien ein Abbild der Aufgabe skizziert (\triangleq im vorliegenden Fall: Durchführung einer PV), indem aufgabenspezifische kritische Merkmale identifiziert werden. Gemeinsames Ziel dieser Methoden ist es, implizites und explizites Wissen, sowie kognitive Fähigkeiten zu erfassen, die Experten beim Ausführen komplexer Aufgaben (\triangleq Durchführung einer PV) aus ihrem Fachgebiet anwenden und das u.U. über das in der Literatur angegebene Wissen hinausgeht. (17) (40)

Aus diesem breit gefächerten Methodenangebot der CTA wurden offen gestaltete Interviews mit Wirbelsäulenchirurgen praktiziert, um durch deren Fachwissen auf relevante Bestandteile einer PV zu stoßen. Da medizinische Experten bei der Beschreibung einer OP ihres Fachgebietes oftmals dazu tendieren, wichtige detaillierte Informationen zu verbergen, die in Form des prozeduralen Gedächtnisses unbewusst und automatisiert ablaufen, wurde zusätzlich die Beobachtungsstrategie aus dem Methodenangebot der CTA in Form einer in situ-Observation einer PV im OP-Saal angewendet. In Kombination mit der Konsultation einschlägiger Fachliteratur (Artikel über Vertebroplastie in der Fachliteratur) konnten dadurch zusätzliche - etwaige in Interviews verborgene - Detailinformationen schriftlich dokumentiert werden. Mittels der in situ-Observation konnte durch Generieren eines tieferen Verständnisses für die Vorgänge einer PV die Verbindung

¹² Der BGDV-Fragebogen wird im Anhang unter 7.1 abgebildet.

zwischen den aus Interviews und Literatur gewonnenen theoretischen Informationen und den praktisch-klinischen Informationen hergestellt werden. (17)

2.2.1 Gegenüberstellung mit dem Prototyp des TSA-Bogens

Neben den genannten Interviews, Literaturkonsultation und PV-Observation wurde bei der Formulierung der TA des BGDV-Fragebogens an Bestandteile des Bewertungsbogens „Vertebroplasty: Technical Skills Assessment Scale for Surgeons“¹³ angeknüpft. In der vorliegenden Arbeit wird dieser abkürzend als „TSA-Bogen“ bezeichnet. Er war im Rahmen der „Simulated Operating Room Environment Study“ ebenfalls von Mitarbeitern des NARVIS-Teams konzipiert worden. Er bewertet das chirurgische Outcome (surgical outcome) eines Chirurgen bei der Durchführung einer PV. (37) Im Rahmen dieser Studie wurde zudem die chirurgische Performance (surgical performance) des Chirurgen mittels des etablierten Bewertungsschemas OSATS (Objective Structured Assessment of Technical Skill)¹⁴ evaluiert, welches an die Durchführung einer PV angepasst wurde. Die Kombination des Urteils aus TSA-Bogen (chirurgisches Outcome) und OSATS (chirurgische Performance) führte in dieser Studie zur Taxierung der chirurgischen Fähigkeit resp. Kompetenz eines Behandlers (= technische Fähigkeiten/Technical Skills, TS). Da das aus TSA-Bogen und OSATS bestehende Evaluierungstool (sog. TSA/OSATS-Score) dieselben elf Teilnehmer der PV-Trainingssimulation bewertete wie der PV-Score der vorliegenden Arbeit, wird der TSA/OSATS-Score unter 2.4.4 als bereits etabliertes, vergleichbares Evaluierungstool als Validierungsinstrument für den PV-Score herangezogen.

Neben den TS beurteilten zwei ausgebildete Beobachter (HF-Experten, Human Factor) in der „Simulated Operating Room Environment Study“ auch die Non-technical-skills (NTS) der Teilnehmer mithilfe des Ansatzes des OTAS (Observational Teamwork Assessment for Surgery). Hierbei werden nicht-technische Fähigkeiten (NTS) eines Chirurgen wie Kommunikation, Koordination, Kooperationsverhalten und Unterstützung des OP-Teams, Führungsverhalten und Situationsbewusstsein beurteilt. (41) Die NTS finden in der vorliegenden Arbeit keine Beachtung.

¹³ Der TSA-Bogen wird im Anhang unter 7.2.1 abgebildet. Außerdem ist er in einer leicht abgewandelten Version auffindbar unter: <http://links.lww.com/BRS/B445> (Appendix 1) (37)

¹⁴ Das OSATS-Bewertungsschema ermöglicht eine „objektive strukturierte Überprüfung technischer Fertigkeiten“ (82). Aus urheberrechtlichen Gründen wird die Version des OSATS-Bewertungsschemas, welche für den TSA/OSATS-Score verwendet wurde, in der vorliegenden Arbeit nicht abgebildet. Weitere Erläuterungen zum OSATS-Bewertungsschema sind in den Quellen (82) und (37) zu finden. Unter dem Link <http://cirugiaminimamenteinvasiva.es/SurgTTT-eLearning/index.php/de/mod8-3/assess-the-resident-progression/663-page-6-objective-structured-assessment-of-technical-skills-osats-2> (Zugriff: 03.04.2020) ist eine Abbildung der Grundstruktur des OSATS-Bewertungsschemas zu finden.

Der TSA-Bogen lässt sich als Prototyp des darauf aufbauenden BGDV-Fragebogens verstehen, da die Inhalte der beiden Fragebögen grundsätzlich übereinstimmen, die TA des BGDV-Fragebogens aber eine Präzisierung, bzw. Erweiterung darstellen.

Tabelle 2 visualisiert dieses Verhältnis der beiden Fragebögen in Form einer Gegenüberstellung. Dabei erscheinen die TA des TSA-Bogens **fett** gedruckt, entsprechende TA des BGDV-Fragebogens *kursiv*, L=links, R=rechts, RS=Risikostuktur, (I)EP=(idealer) Eintrittspunkt, Ant. WKW=anteriore Wirbelkörperwand und WK=Wirbelkörper.

TA	TSA-Bogen	BGDV-Fragebogen
TS_1/2_L/R <i>1a)</i>	Frage nach betroffenen No-Go-Areas/ RS: bivariates Outcome (Punktion RS: ja/ nein)	Präzisierung durch Kritizitätsranking der möglichen betroffenen RS, sowie Einordnung in Kritizitätskategorien
TS_3_L/R <i>1g)</i>	Klassifizierung des „EP“ in die Kategorien: angemessen/ einseitig nicht angemessen/ beidseitig nicht angemessen	Präzisierung des „IEP“ durch frei formulierbare Definition
TS_4_L/R <i>1 i)/j)</i>	Klassifizierung der „Anzahl der Platzierungsversuche des Trokars beim Eintrittsvorgang“: bivariates Outcome (Anzahl angemessen/ zu oft)	Präzisierung durch Differenzierung des „Platzierungsversuches in Bezug auf Trokarkontakt mit Knochen bzw. Hautperforation“ sowie Angabe der „noch akzeptablen Versuche“
TS_6_L/R <i>1m)/n) l)</i>	Klassifizierung der „Lage des Trokars innerhalb des Pedikels und des WK“ in die Kategorien: ideal/ nicht ideal, aber unkritisch/ kritisch	Präzisierung der „idealen Lage des Trokars innerhalb des WK“ durch Angabe des minimal einzuhaltenden Abstandes der Trokarspitze von Deck-/Bodenplatte/ Ant. WKW; Einzeichnen der „idealen Nadelposition“ in PC-Simulation
TS_10_L/R <i>1l)</i>	Klassifizierung der „Tiefe der Trokarspitze bei der Zementinjektion“ in die Kategorien: nicht tief genug/ angemessen/ zu tief	Präzisierung der „idealen Tiefe der Trokarspitze in der Endlage für die Zementinjektion“ durch frei formulier- bare Definition
TS_11/12/13 <i>2c)</i>	Klassifizierung der „RS bei Zement- leakage“: bivariates Outcome (Leakage in RS: kritisch/ unkritisch)	Präzisierung durch Kritizitätsranking der RS bei Zementleakage sowie Einordnung in Kritizitätskategorien
TS_21 <i>2a)</i>	Klassifizierung der „injizierten Zementmenge in den WK“ in die Kategorien: nicht genug/ angemessen/ zu viel	Präzisierung der Klassifizierung der „idealen Zementmenge“ durch genaue Milliliter (ml) -Angaben
TS_23 <i>3a)</i>	Klassifizierung der „passenden Ein- stellung des C-Bogens“ in die Kategorien: ja/ teilweise/ nein	Präzisierung der „passenden Einstellung des C-Bogens“ durch frei formulierbare Definition

TS_27 4a)	Klassifizierung der „Dauer der Gesamtprozedur“ in die Kategorien: zu schnell/ angemessen/ zu langsam	Präzisierung der „idealen Gesamtdauer“ durch Angabe des idealen Zeitfensters in Minuten (Min.)
---------------------	--	--

Tabelle 2: Gegenüberstellung der TA des TSA-Bogens mit den entsprechend präzisierenden TA des BGDV-Fragebogens, eigene Darstellung

2.2.2 Gliederung, Struktur und Inhalt des BGDV-Fragebogens

Der BGDV-Fragebogen lässt sich nach verschiedenen Gesichtspunkten in die folgenden Gliederungsebenen einteilen:

- a) Inhaltliche Komponenten einer PV (*inhaltlich*)**
- b) Art der Bearbeitung:** schriftlich bzw. simulationsbasiert (*medial*)
- c) Art des Dateninformationsgehaltes (*formal*):** durch offene/geschlossene Formulierung der TA (*Antwortformat*) Erfassen weicher/qualitativer, bzw. harter/quantitativer Merkmale (*Gehalt*)
- d) Verwendungszweck der TA:** zum direkten oder indirekten Einbau in den PV-Score bzw. für erweiterten Erkenntnisgewinn (*funktional*)

Abbildung 9 zeigt einen Überblick über die verschiedenen Gliederungsebenen des BGDV-Fragebogens. Die Prozentzahlen in der Abbildung bezeichnen die prozentualen Häufigkeiten der TA des BGDV-Fragebogens, die zu der jeweiligen Komponente gehören, wobei sie sich immer auf den Gesamtwert von 24 TA (\triangleq 100%) beziehen.

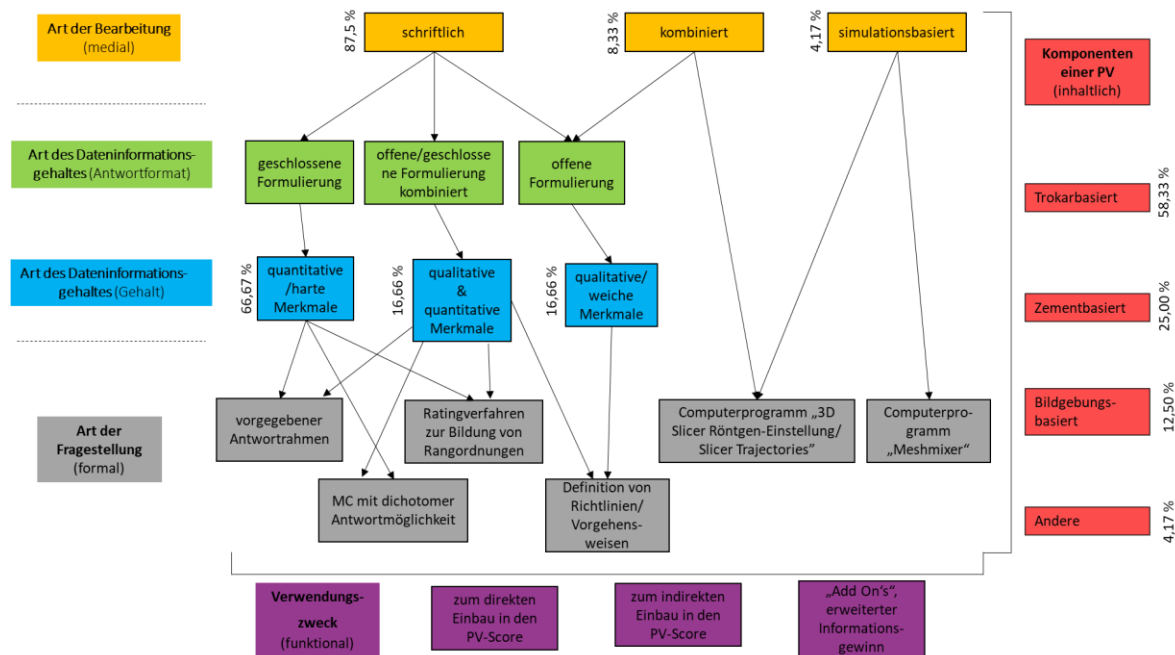


Abbildung 9: Überblick über die Gliederungsebenen der Teilaufgaben des BGDV-Fragebogens: Art der Bearbeitung, Art des Dateninformationsgehaltes, Art der Fragestellung, Verwendungszweck, inhaltliche Komponenten, eigene Darstellung

a) Inhaltliche Komponenten einer PV

Die Kategorisierung nach inhaltlichen Komponenten einer PV erfolgte mit dem Ziel, die Anwenderfreundlichkeit des BGDV-Fragebogens und damit die Compliance der Experten zu vergrößern. Durch eine überschaubare und einleuchtende Gliederung sollte eine positive Grundeinstellung der Experten gefördert, sowie deren Bereitschaft gesteigert werden, hilfreiche Informationen über den Ablauf einer PV darzulegen. Bei den Komponenten einer PV wurden inhaltlich die vier folgenden Kategorien differenziert:

1. Trokar-basiert: enthält Komponenten zum Platzierungsversuch, Eintrittspunkt und zur Endlage des Trokars, sowie mögliche Risikostrukturen bei einer Trokarpunktion
2. Zement-basiert: enthält Komponenten zur Zementmenge, sowie mögliche Risikostrukturen bei einer Zementleckage
3. Bildgebungs-basiert: enthält Komponenten zur Einstellung des C-Bogens, sowie zur Anzahl angefertigter Röntgenbilder
4. Anderes: enthält Komponenten zur benötigten Zeit für die Durchführung einer PV

Im Anschluss wurden die in den vorherigen Schritten durch Interviews, Literaturkonsultation, in-situ-Observation und Präzisierung des TSA-Bogens eruierten TA den jeweils passenden Kategorien zugeordnet.

Während die Kategorie „Trokarn-basiert“ mit 14 TA (58,33%) bestückt wurde, fielen sechs TA (25,00%) unter die Rubrik „Zement-basiert“, drei TA (12,50%) unter die Kategorie „Bildgebungs-basiert“ und eine TA (4,17%) zählte zur Kategorie „Anderes“. (s. Abb. 9)

Der Beitrag des beteiligten Psychologen- und Pädagogenteams bestand darin, die TA hinsichtlich ihrer Anwenderfreundlichkeit für die Experten zu beurteilen.

So kam es zu einer Feinjustierung, Umformulierung bzw. ggf. zur Streichung von TA, bei denen das Psychologen- und Pädagogenteam eine verminderte Compliance vonseiten der Anwender durch zu komplexe Formulierung oder komplizierte Aufgabenstellung prognostizierte.

b) Art der Bearbeitung

Von den insgesamt 24 zu bearbeitenden TA des BGDV-Fragebogens waren 21 (87,50%) schriftlich in Papierform auszufüllen, zwei stellten eine Kombination aus schriftlicher und simulationsbasierter Bearbeitung dar (8,33%) und eine (4,17%) fand gänzlich simulationsbasiert anhand eines Programmes am Computer statt. (s. Abb. 9) Die Konzeption der simulationsbasierten TA fiel unter den Zuständigkeitsbereich der Informatiker, welche den durch Interviews, Fachliteratur und in situ-Observation gewonnenen Informationsbedarf in virtuelle TA überführten.

c) Art des Dateninformationsgehaltes

Wie Collins und KollegInnen herausfanden, ist es ein typisches Vorgehen im Rahmen der MMS, sowohl qualitative als auch quantitative Befragungen an der gleichen Untersuchungsgruppe durchzuführen. (34 S. 20) Dieses Vorgehen wurde auch bei der Konzipierung des BGDV-Fragebogens praktiziert: ein Teil der schriftlich in Papierform auszufüllenden TA wurde als geschlossene Fragen formuliert, ein Teil als offene Fragen mit Freitextantworten – sodass durch diese Kombination verschiedener Formulierungen sowohl qualitative (weiche) als auch quantitative (harte) Merkmale, bzw. Daten registriert werden konnten.

Bei weichen (qualitativen) Daten handelt es sich um Informationen, welche nicht oder nur schwer mess-, bzw. quantifizierbar sind. Sie erfassen subjektive Bewertungen, bzw. Beschreibungen von Handlungen und müssen demnach interpretiert werden. Im BGDV-Fragebogen sind sie in Form von frei formulierbaren TA enthalten, beispielsweise durch die Definition bestimmter Ebenen oder Vorgehensweisen in der PV. (42) Diese Merkmale bieten den Vorteil, durch ihre offene Fragestellung

Zugewinn von Neuinformation liefern zu können, welche durch eine geschlossene Formulierung mit vorgegebenen Antwortmöglichkeiten u.U. übersehen, bzw. nicht erfasst werden würden. (34) (35) Harte (quantitative) Daten enthalten Informationen in numerischer Form, welche exakt mess- und quantifizierbar sind. Somit können harte Daten objektiv nachgeprüft werden und bedürfen nicht zwingend einer Interpretation. (42) In der vorliegenden Arbeit liegen sie in Form von geschlossen formulierten TA vor. Teilweise werden harte Daten im BGDV-Fragebogen durch Multiple-Choice- (MC-) Fragen mit dichotomer Antwortmöglichkeit („ja“ oder „nein“) erfasst. Anders formulierte TA mit hartem Datengehalt geben einen Rahmen zu den möglichen Antwortmöglichkeiten vor.

Wieder andere erfordern im Sinne von Ratingverfahren die Bildung von Rangordnungen. (s. Abb. 9) Der Vorteil der geschlossen formulierten TA liegt in einem präzisen und konkreten Informationsgewinn. Die Information ist direkt in der Antwort enthalten, ohne dass eine Interpretation, Zusammenfassung oder Kürzung vonnöten ist. Ein Nachteil kann sein, dass Erkenntnisgewinn durch die eingeschränkte Antwortfreiheit „nur“ im Rahmen der vorgegebenen Antwortmöglichkeiten stattfindet.

Zusammengefasst kann festgehalten werden, „dass die quantitativen Methoden zum Messen und die qualitativen Methoden zum Verstehen eingesetzt werden“, wobei die quantitativen Anteile als die Basis eines Messinstruments zu verstehen sind, welche zuverlässig die benötigte Information liefern sollen, während qualitative Anteile im Sinne von „Add-On’s“ zusätzliche Nebeninformationen zur Erkenntniserweiterung liefern können. (34) Beim Verhältnis von qualitativen zu quantitativen Fragen eines Messinstruments liegt der Schwerpunkt meist auf den quantitativen. Nur in seltenen Ausnahmen liegt eine höhere Gewichtung der qualitativen Anteile vor. (34) Auch im BGDV-Fragebogen ist dieser Fokus auf quantitativen Anteilen erkennbar: 16 (66,67%) Fragen waren ausschließlich geschlossen formuliert, vier (16,66%) enthielten offen und geschlossen formulierte Anteile und vier (16,66%) waren rein offen formuliert. (s. Abb. 9)

Im Hinblick auf die Auswertungsobjektivität des BGDV-Fragebogens wurde die Tatsache berücksichtigt, dass Defizite insbesondere bei der Auswertung offener Antworten durch unterschiedliche Codierungsmöglichkeiten etwaiger unterschiedlicher Auswerter auftreten könnten. Diese Form der Objektivität ist umso anfälliger für Beeinträchtigungen, je mehr der Auswertende die unmittelbaren Antworten des Befragten in Zahlenwerte transformieren muss. (32) Da die offen formulierbaren TA des BGDV-Fragebogens nicht mit der Absicht konstruiert wurden, konkreten quantitativen Dateninhalt zu generieren, sondern vielmehr zusätzlicher Hintergrundinformation dienten, tritt die Auswertungsobjektivität an dieser Stelle in den Hintergrund.

Das Auswerten der geschlossen formulierten TA gestaltet sich im Hinblick auf die Auswertungsobjektivität einfacher, da der Auswertende weniger Freiheit bei der Auswertung hat, da es sich bei den Angaben um konkret messbare Zahlenwerte handelt.

d) Verwendungszweck der TA

Bereits während der Konzipierung des BGDV-Fragebogens zeichnete sich ab, dass zwar ein Großteil der TA für den direkten oder indirekten Einbau in den PV-Score bestimmt war, einige Komponenten des BGDV-Fragebogens ferner aber vielmehr als sog. „Add-On's“ formuliert wurden, um Erkenntnisgewinne bzw. Anhaltspunkte für zukünftige Forschung zu liefern. Eine konkrete Einteilung der TA unter diesem Gesichtspunkt konnte zum Zeitpunkt der Konzeption des BGDV-Fragebogens aber noch nicht vorgenommen werden, da die Auswertbarkeit bestimmter Daten der PV-Trainingssimulation noch nicht vorlag.

2.3 Konzeption des PV-Scores

In Anlehnung an die von Schmidt M.W. et al. vorgeschlagenen Richtlinien für die Entwicklung eines aussagefähigen Scoring-Systems beruhte die Konzeption des PV-Scores auf folgenden Schritten (s. Abb. 10) (21 S. 2094):

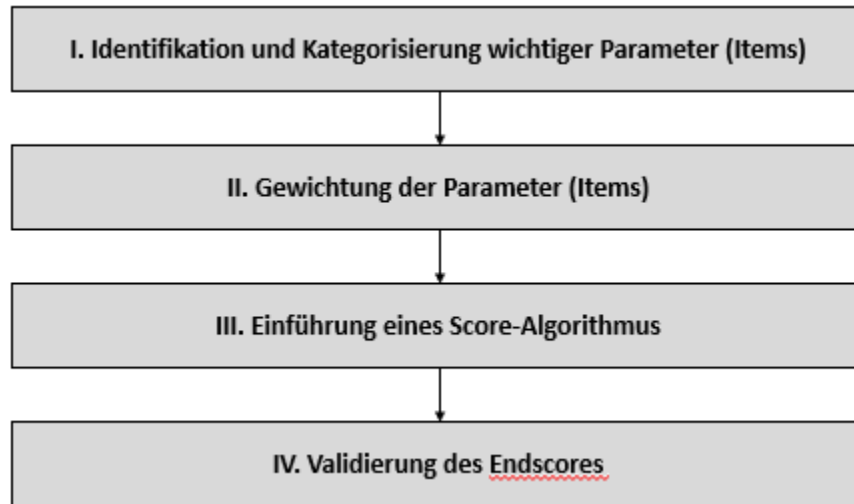


Abbildung 10: Schritte der Konzeption des PV-Scores, eigene Darstellung

Zu I. Identifikation und Kategorisierung wichtiger Parameter (Items)

Bereits bei der Konzipierung des BGDV-Fragebogens war unter 2.2.2 „a) Inhaltliche Komponenten einer PV“ darauf geachtet worden, nur TA aufzunehmen, die eine möglichst große Bandbreite der bei der Durchführung einer PV relevanten Fähigkeiten repräsentieren. Ein besonderes Augenmerk bei der Auswahl der relevanten TA lag stets darauf, dass sie im Hinblick auf die später angestrebte Bewertung der chirurgischen Kompetenz hilfreich sein sollten. Eine „Vorauswahl“ wichtiger Parameter war somit schon während der Fragebogenerstellung vorgenommen worden.

Im Zusammenhang mit der Konzeption des PV-Scores werden die Parameter im Folgenden als Items bezeichnet. Sie enthalten die Information der entsprechenden ausgewerteten TA des BGDV-Fragebogens in Form des Expertenkonsensus.¹⁵

Bei der schrittweisen Identifikation wichtiger Items des PV-Scores wurden – wie sich unter 2.2.2 „d) Verwendungszweck der TA“ bereits abzeichnete - nicht alle TA des BGDV-Fragebogens übernommen. Die Gesamtheit der frei formulierbaren und simulationsbasierten TA, sowie auch Teile der geschlossen formulierten TA, fanden keinen expliziten Eingang in den PV-Score.

¹⁵ Genauere Erläuterungen, wie die Auswertung der TA des BGDV-Fragebogens zum Expertenkonsensus führte, finden sich unter „2.5 Datenspeicherung und Statistik“.

Die Gründe dafür lagen teilweise darin, dass keine auswertbaren Daten der PV-Trainingssimulation für die jeweiligen TA vorlagen. Teilweise handelte es sich aber auch um TA, die eigens dafür konzipiert worden waren, implizit als Anhaltspunkte resp. Richtlinien für die Interpretation bzw. Gewichtung anderer TA sowie darüber hinaus als „Add-On's“ im Sinne von zusätzlichem Erkenntnisgewinn zu dienen.

Bei den TA, die explizit Eingang in den PV-Score fanden, handelte es sich ausschließlich um geschlossen formulierte TA mit quantitativem Datengehalt. Dies entspricht dem typischen Verhältnis im Umgang mit quantitativen und qualitativen Daten, wo es nur in Ausnahmefällen zu einer höheren Gewichtung der qualitativen im Vergleich zu den quantitativen Anteilen kommt. (34)

Die für den Einbau in den PV-Score als geeignet eingestuften TA des BGDV-Fragebogens wurden mitsamt des in ihnen enthaltenen Expertenkonsensus als Items des PV-Scores formuliert.

Im nächsten Schritt wurde eine Kategorisierung dieser Items vorgenommen.

War die Kategorisierung des BGDV-Fragebogens nach inhaltlichen Komponenten einer PV vorgenommen worden (1. Trokar-, 2. Zement-, 3. Bildgebungs-basiert, 4. Andere), so erfolgte die Kategorisierung der Items des PV-Scores nach inhaltlichen Komponenten chirurgischer Kompetenz. Bei der Auswahl geeigneter Kategorien wurde auf die Strukturierung bereits etablierter Scoring-Systeme zurückgegriffen, die ebenfalls das Konstrukt der chirurgischen Fähigkeit resp. Kompetenz erfassen.^{16 17}

In Anlehnung an die Kategorie-Einteilungen der beiden Scoring-Systeme RSA-Score (s. Fußnote 17) und Heidelberg-VR-Score (s. Fußnote 18) lauten die fünf Kategorien des PV-Scores:

1. Zeit, 2. Effizienz, 3. Sicherheit, 4. Geschicklichkeit, 5. Outcome

Diese stellen die Subdimensionen des Kompetenzbegriffes dar, die der PV-Score erfassen kann und die in ihrer Gesamtheit das zu messende Konstrukt der chirurgischen Kompetenz ergeben.

Es wurde sich für eine deutschsprachige Bezeichnung der Begriffe entschieden, da das Forschungsumfeld im Rahmen der vorliegenden Forschungsarbeit deutsch war.

Nachfolgend wurden die eruierten Items den fünf Kategorien zugeordnet. (s. Tabelle 3, S. 32)

¹⁶ Der Robotic-Skills-Assessment- (RSA-) Score ist für die standardisierte Bewertung chirurgischer Fähigkeit an VR-basierten chirurgischen Simulatoren geeignet und enthält die Kategorien „safety in operative field“, „economy of motion“, „bimanual dexterity“ und „critical error“. (22)

¹⁷ Der Heidelberg-VR-Score evaluiert chirurgische Fähigkeit an Laparoskopie-VR-Simulatoren und besteht aus den fünf Hauptkategorien „time“, „efficiency“, „safety“, „dexterity“ und „outcome“. (21)

Zu II. Gewichtung der Parameter (Items)

Unter Orientierung am Expertenkonsensus wurde für jedes Item eine Höchstpunktzahl festgelegt. Damit geschah die Skalierung des PV-Scores auf der Basis von datengestützter medizinischer Erfahrung. Die zu unterschiedliche Gestaltung der Codierung, resp. Bepunktung eines Scores kann dessen Handhabung beeinträchtigen und ihn fehleranfällig machen. (28) Daher wurde eine möglichst einheitliche Bepunktung angestrebt, wobei eine höhere Punktzahl definitionsgemäß mit einem schlechteren Ergebnis verbunden war. Die Punktwerte des PV-Scores sind also im Sinne von „Strafpunkten“ zu verstehen, die zunehmen, je mehr Fehler ein Teilnehmer in der Durchführung der PV begangen hat.

Beispielhaft sei dies an dem Item „Gesamtmenge an appliziertem Zement“ (3C.) erläutert.

Laut Expertenkonsensus (\triangleq ausgewerteten Antworten der TA 2a)) ¹⁸ wurde eine Gesamtmenge an appliziertem Zement bis 2,0 ml als „zu wenig“, von 2,1 bis 6,6 ml als „passend“ und ab 6,7 ml als „zu viel“ angesehen. Da außerdem laut Expertenkonsensus eine zu geringe Zementmenge positiver zu bewerten war als eine zu große (\triangleq ausgewerteten Antworten der TA 2b)), fiel die Gewichtung resp. Bepunktung dieses Items folgendermaßen aus:

bis 2,0 ml (zu wenig Zement):	1 Pkt.
2,1 bis 6,6 ml (passende Zementmenge):	0 Pkt.
ab 6,7ml (zu viel Zement):	2 Pkt.

Dieses Beispiel verdeutlicht den oben genannten impliziten Nutzen der TA 2b), welche zwar keinen direkten Einbau in den PV-Score erfuhr, aber zur Interpretation bzw. Gewichtung des Items 3C. herangezogen wird.

Um eine einheitliche Bepunktung zu erreichen, wurden - wenn möglich - ganzzahlige Punktwerte vergeben. Bei den Items 1A., 2A., 2B. musste jedoch von der ganzzahligen Bepunktung abgewichen werden, da sich bei Durchschau der Daten der Teilnehmer der PV-Trainingssimulation große Abweichungen vom jeweiligen Expertenkonsensus der TA ergaben. Die fortgeführte ganzzahlige Bepunktung hätte in diesen Items zu einer hohen Strafpunktezah jedes einzelnen der elf Teilnehmer geführt. Um eine Verfälschung der Resultate zu vermeiden und eine für die Daten der PV-Trainingssimulation angemessene Bepunktung zu erzielen, wurde in diesen Fällen eine Bepunktung mit Dezimalzahlen praktiziert. Beispielhaft sei dies an dem Item „Anzahl (Stück) angefertigter Röntgenbilder beim Platzierungsversuch bis zur Endlage des Trokars“ (2A.) erläutert.

¹⁸ Der Expertenkonsensus stellt die ausgewerteten Antworten des BGDV-Fragebogens dar und wird als solcher erst unter „3.1 Ergebnisse des BGDV-Fragebogens“ erläutert. An dieser Stelle wird an den Beispielen der ausgewerteten TA 2a)/b) (Item 3C.), sowie 1k)/3c) (Item 2A.) zum besseren Verständnis vorgegriffen.

Laut Expertenkonsensus (\triangleq ausgewerteten Antworten der TA 1k)) wurde eine Röntgenbilder-Anzahl von ≤ 11 Stück als angemessen erachtet. Da aber alle elf Teilnehmer eine weitaus höhere Röntgenbilder-Anzahl angefertigt hatten als der Expertenkonsensus vorgab, wurde aus den genannten Gründen eine entsprechend differenzierte Bepunktung angewendet:

bis zu 11 Rö-Bilder	: 0,0	Pkt.	bis 44 „“	: 0,6	Pkt.
bis 22 „“	: 0,2	Pkt.	bis 55 „“	: 0,8	Pkt.
bis 33 „“	: 0,4	Pkt.	ab 56 „“	: 1,0	Pkt.

Auch die Items 3A. und 4A. verwenden Dezimalzahlen, die in diesen Items jedoch nicht durch starke Abweichungen vom Expertenkonsensus begründet sind, sondern inhaltliche Gründe haben (zur Bepunktung der Items 3A. und 4A. s. 3.1.4 Abbildung des PV-Scores).

Da alle Items zuvor den unter I. genannten fünf Kategorien zugeordnet worden waren, konnte nun für jede Kategorie durch Addition der Höchstpunktzahlen der in ihr enthaltenen Items eine Höchstpunktzahl der jeweiligen Kategorie bestimmt werden.

Für die Kategorie „5. Outcome“ wurde keine Höchstpunktzahl bestimmt, da unter diese Kategorie Fehler fallen, die ein Erreichen des Ziels resp. die erfolgreiche Durchführung einer PV verhindern. Für derartige Fehler wurde der Ausdruck „Killerkriterien“ statuiert. Durch Konsultation einschlägiger Literatur sowie Rücksprachen mit Experten der Wirbelsäulenchirurgie wurden die Punktion des Rückenmarks, der Spinalnervenwurzeln und/oder der Aorta als Killerkriterien festgelegt.

- Killerkriterium Rückenmark: Der Verlauf einer iatrogen bedingten Myelopathie hängt von Einflussfaktoren wie dem Ausmaß der Punktion und der mikrovaskulären Versorgung des Rückenmarks ab. (43) Weitere Einzelheiten zu den möglichen Folgen einer Rückenmarksschädigung durch Trokarpunktion oder Zementleckage (z.B. Muskelschwäche, Lähmungserscheinungen, Empfindungsstörungen, mangelnde Blasen- und Darmkontrolle) finden sich in den Kapiteln 1.2, sowie 4.2.1 (TA 2c)).
- Killerkriterium Spinalnervenwurzeln: Eine Schädigung kann zu Empfindungsverminderungen/-verlusten/Missempfindungen oder Muskelschwäche führen, je nachdem, ob eine sensorische oder eine motorische Nervenwurzel betroffen ist. (44)
- Killerkriterium Aorta: Eine Schädigung kann zu arteriellen Blutungen oder Pseudoaneurysma führen. (45)

Durch Addition der Höchstpunktzahlen der Kategorien 1. bis 4. wurde die Gesamthöchstpunktzahl des PV-Scores ermittelt. Diese beträgt 18 Punkte und repräsentiert das Höchstmaß an Gesamtpunkten, wenn in jedem Item, bzw. jeder Kategorie das schlechteste Ergebnis erzielt, allerdings keines der definierten Killerkriterien erfüllt wird.

Tabelle 3 zeigt einen Überblick über die Zusammensetzung des PV-Scores.¹⁹ Dabei werden folgende Abkürzungen verwendet: PLV=Platzierungsversuch, TEL=Trokarendlage, RS=Risikostrukturen, DPL=Deckplatte, BPL=Bodenplatte, IEP= idealer Eintrittspunkt, KK=Killerkriterien, RM=Rückenmark, SpNW=Spinalnervenzurzel.

Kategorien	1. Zeit	2. Effizienz	3. Sicherheit	4. Geschicklichkeit	5. Outcome
Inhalt (Item)	1A.: Faktor Zeit beim PLV bis zur TEL, 1B.: Gesamtdauer der OP	2A.: Anzahl angefertigter RÖ beim PLV bis zur TEL, 2B.: Anzahl angefertigter RÖ in Gesamt-OP	3A.: Punktion einer RS, 3B.: Abstand der Nadel von DPL/BPL/Ant. WKW, 3C.: Gesamtmenge an appliziertem Zement	4A.: Abweichung vom IEP, 4B.: Ansatzpunkte der Nadel am Knochen beim PLV	KK: Punktion des RM, der SpNW und/oder der Aorta
Entspr. TA des BGDV-Fragebogens	1k), 4a)	1k)/3c), 3b)/3c)	1a)/b)/c/e), 1m)/n), 2a)/2b)	1f)/g)/h), 1i)	1a)/b)
Mögl. HPZ der Einzel-Items	2,1	2,1	4, 2, 2	2,2	Note 6
Mögl. HPZ der Kategorie	3	3	8	4	Note 6
Mögl. Gesamt- HPZ des PV-Scores *	18				

Tabelle 3: Inhalt der Kategorien des PV-Scores sowie entsprechende TA des BGDV-Fragebogens und mögliche Höchstpunktzahlen der Einzel-Items, der Kategorien (durch Addition der HPZ der Einzel-Items) und des PV-Scores (durch Addition der HPZ der Kategorien)

Zu * in Tabelle 3: Diese 18 Punkte stellen die Gesamthöchstpunktzahl des PV-Scores unter der Voraussetzung dar, dass keines der Killerkriterien erfüllt wurde und dass unter dem Item 3A.: „Punktion einer Risikostruktur“ nur *einmal* das schlechtest mögliche Ergebnis (vor Killerkriterien) erreicht wurde. Dem Expertenkonsensus zufolge entspricht dies unter 3A. einer Punktion der Vena cava, des Spinalkanals oder der Lunge. In der praktischen Anwendung der Endversion des PV-Scores auf die Daten der Teilnehmer wäre jedoch sozusagen

¹⁹ Die Endversion des PV-Scores wird nicht an dieser Stelle, sondern unter „3.1.4 Abbildung des PV-Scores“ dargestellt, da sie den Expertenkonsensus der neun Items 1A.-4B. enthält und somit den „Ergebnissen des BGDV-Fragebogens“ vorweggreifen würde.

eine „unendlich hohe“ Gesamthöchstpunktzahl möglich, da die etwaige Punktion *mehrerer* Risikostrukturen zu einer Addition der entsprechenden „Strafpunkte“ führt. Der Begriff der Gesamthöchstpunktzahl ist an dieser Stelle folglich im Sinne einer Obergrenze zu verstehen, anhand derer sich unter Punkt III. ein Score-Algorithmus generieren lässt.

zu III. Einführung eines Score-Algorithmus

Unter einem Algorithmus ist eine logische und systematische Regel, bzw. Vorgehensweise zu verstehen, die zur Lösung einer vorliegenden Situation, bzw. eines Problems angewendet wird. (46) Man kann einen Algorithmus mit einer Funktion vergleichen:

- (1.) Man gibt bestimmte Vorgaben, bzw. Daten ein (**Eingabe**) ...
- (2.) ... und der Algorithmus berechnet daran das Ergebnis (**Ausgabe**).

In einer Schulklasse wird der Algorithmus des Bewertungssystems einer Klassenarbeit vom Lehrer durch Aufstellen eines linearen Notenschlüssels festgelegt. Durch diesen Notenschlüssel wird jeder erreichbaren Punktzahl der Klassenarbeit (**Eingabe**) eine Schulnote zwischen eins und sechs zugewiesen (**Ausgabe**). Da jede Schulnote wiederum eine bestimmte Stufe an Leistung, bzw. Kompetenz kodiert (1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = ausreichend, 5 = mangelhaft, 6 = ungenügend), kann somit jedem Schüler durch seine erreichte Punktzahl eine objektive Bewertung seiner Leistung, bzw. Kompetenz zugeschrieben werden.

Bei der Generierung des Algorithmus eines Scores sollte es sich um eine standardmäßige, transparente und im Sinne der Anwenderfreundlichkeit möglichst einfach gestaltete Vorgehensweise handeln. (21) Nach Rücksprache mit den Pädagogen und Psychologen des Forschungsteams wurde daher ein möglichst einfach zu handhabender und praktikabler Algorithmus gewählt, welcher auf demselben Prinzip wie das Schulnotensystem basiert. Konträr zum Schulnotensystem, bei dem eine höhere Punktzahl mit einer besseren Schulnote einhergeht, ist beim PV-Score ein niedriger Punktwert ein Indikator für eine hohe chirurgische Kompetenz.

Um den Algorithmus des PV-Scores zu generieren, kam es im nächsten Schritt zum Verfahren des Standard-Setting. Dabei handelt es sich um die Festlegung von sog. Cut-Scores (Testtrenn-/ Schwellenwerten) auf einer kontinuierlichen Leistungstestskala, durch die sich benachbarte Kategorien abgrenzen lassen. Die Festlegung der Cut-Scores kennzeichnet eine äußerst kritische Verbindungskomponente zwischen dem evidenzbasierten und empirisch überprüfbaren Gesamtsystem (\triangleq Expertenkonsensus) einerseits und den konsequenzbezogenen, eher richtungsweisenden und praxisrelevanten Aspekten (\triangleq PV-Score) andererseits. (47)

Die unter II. ermittelte Gesamthöchstpunktzahl von 18 Punkten sollte definitionsgemäß die Obergrenze der Note 5 darstellen. Dies bedeutet, dass ein Teilnehmer, der über 18 Punkte ansammelt und/oder ein Killerkriterium erfüllt, mit der Note 6 bewertet wird. Das Festlegen dieses ersten Cut-Scores fand demnach in Orientierung an die Gesamthöchstpunktzahl statt.

Für die Identifikation der Cut-Scores für die Noten 1 bis 5 wurde die Gesamthöchstpunktzahl von 18 Punkten (\triangleq Obergrenze der Note 5) durch die verbleibenden möglichen fünf Notenwerte (Noten 1 bis 5) dividiert.

$$\rightarrow 18 \text{ Punkte} / 5 = 3,6 \text{ Punkte}$$

Demnach ist jeweils nach 3,6 Punkten (\triangleq Cut-Score) ein Notenwechsel indiziert.

Im Folgenden mussten die Schwellenwerte zwischen attestierter Kompetenz und nicht attestierter Kompetenz eines Teilnehmers klar dargestellt werden. (24) In Anlehnung an die bereits etablierte Cut-Score-Setzung des Schulnotensystems jenseits der Schulnote 4 (Schulnoten 1-4 = bestanden, Schulnoten 5 und 6 = nicht bestanden) fand eine Differenzierung des Kompetenzniveaus statt.

Teilnehmer inklusive „ausreichendem Kompetenzniveau“ (\triangleq Note 4) resultieren in einem Bestehen (pass) des Tests, Teilnehmer jenseits eines „ausreichenden Kompetenzniveaus“ (\triangleq Note 5 oder 6) in einem Nichtbestehen (fail) (s. Tabelle 4).²⁰

Damit lässt sich der PV-Score definitionsgemäß als diskriminativer Index bezeichnen. Dieser Begriff geht auf eine von Kirshner und Guyatt vorgeschlagene funktionelle Einteilung von Messinstrumenten zurück. Er bezeichnet einen Score, welcher eine Differenzierung von Individuen hinsichtlich eines Merkmals vornimmt, für das kein klassisches Referenzkriterium besteht.

Im vorliegenden Fall sind die Individuen Teilnehmer (Chirurgen), die anhand des Merkmals „chirurgische Kompetenz“ zu differenzieren sind. (28 S. 10)

Des Weiteren stellt der PV-Score eine Multi-Item-Skala da: das Konstrukt der Kompetenz wird anhand mehrerer Items erfasst und anschließend durch deren Aufsummierung bewertet. (32)

Neben dieser numerischen kam es durch die Postulierung der Killerkriterien zu einer weiteren Differenzierung auf inhaltlicher Ebene. Punktiert ein Teilnehmer eine (oder sogar mehrere) der drei Strukturen (Rückenmark, Spinalnervenzurkeln, Aorta), so führt dies automatisch zu einem Nichtbestehen in der Bewertung seiner chirurgischen Kompetenz durch den PV-Score. Somit ist ein Erreichen der Note 6 durch ein Überschreiten der Gesamtpunktzahl von 18 Punkten oder das Erfüllen eines Killerkriteriums möglich.

²⁰ Trotz der Praktikabilität der englischen Bezeichnungen „pass/fail“ werden im weiteren Verlauf die deutschen Bezeichnungen „bestanden/nicht bestanden“ weitergeführt, um die Einheitlichkeit des deutschen Sprachgebrauchs im Laufe der vorliegenden Forschungsarbeit fortzuführen.

Nachdem auf diese Weise der Algorithmus des PV-Scores konstituiert wurde, bekam jede Note eine Konsequenz zugewiesen, welche aus dem Resultat „bestanden“, bzw. „nicht bestanden“ abgeleitet wurde. Die resultierende Konsequenz der Noten 5 und 6 („optionales, bzw. obligates Absolvieren eines Trainingsprogrammes“) wird unter 4.4 näher beleuchtet, da die Ausgestaltung des vorschwebenden Trainingsprogrammes kein Bestandteil der vorliegenden Arbeit war und demnach Anknüpfungspunkte für zukünftige Forschung bietet.

Zusammenfassend lässt sich die Anwendung des PV-Scores folgendermaßen beschreiben: Durch **Eingabe** der Eingangsvariablen (\triangleq Daten eines Teilnehmers der PV-Trainingssimulation) und die nachfolgende Bepunktung sowie Aufaddierung der Eingangsvariablen gemäß des PV-Score-Algorithmus (s. 3.1.4 und Tabelle 4), erhält man den Gesamt-Scorewert des Teilnehmers (**Ausgabe**). Diese Art der Verknüpfung der Eingangsvariablen stellt den typischen Fall in der Formulierung von Score-Systemen dar. (28)

Punkte	0,0-3,6	3,7-7,2	7,3-10,8	10,9-14,4	14,5-18,0	> 18,1/KK
Note	1	2	3	4	5	6
Resultat	Bestanden	Bestanden	Bestanden	Bestanden	Nicht bestanden	Nicht bestanden
KOMP-Niveau	Sehr hoch	Hoch	Befr.	Ausr.	Mangelhaft	Ungen.
Konsequenz	Keine	Keine	Keine	Keine, TP angeraten	SW, zuvor TP optional	SW, zuvor TP obligat

Tabelle 4: Algorithmus des PV-Scores in Form eines Notenschlüssels, mit zugehörigem Resultat, Einstufung des chirurgischen Kompetenzniveaus, sowie daraus resultierender Konsequenz für den Teilnehmer. KOMP=Kompetenz, befr.=befriedigend, ausr.=ausreichend, ungen.=ungenügend

Durch die eindeutige Zuordnung der Punktzahl eines Teilnehmers zu der entsprechenden Note samt daraus hervorgehendem Resultat, Kompetenzniveau und Konsequenz wird der Interpretationsobjektivität (s. 1.6.1) Rechnung getragen. Die eindeutigen inhaltlichen Interpretationshinweise des PV-Scores lassen wenig Spielraum für alternative Interpretationen.

zu IV. Validierung des Endscores

Die Inhaltsvalidität (s. 1.6.3) der Items des PV-Scores wurde durch den Expertenkonsensus des vorgeschalteten BGDV-Fragebogens abgedeckt. Dieser stellte die Überprüfung der Gültigkeit der Item-Inhalte dar.

Unter 2.4.4 wird das Vorgehen der Überprüfung der Kriteriums-, sowie Konstruktvalidität (s. 1.6.3) des PV-Scores geschildert.

2.4 Datenakquise

2.4.1 PV-Trainingssimulation

Die unter 2.1 dargestellte PV-Trainingssimulation wurde im Simulationscenter des Klinikums für Allgemeine, Unfall- und Wiederherstellungschirurgie der LMU, Campus Innenstadt, durchgeführt.

Für die Teilnahme wurden Chirurgen aus zwei Universitäts-Kliniken rekrutiert, indem Einladungen im Schneeballverfahren an alle geeigneten Chirurgen der entsprechenden Fachabteilungen verteilt wurden. (37) Das Schneeballverfahren stellt ein willkürliches Auswahlverfahren dar, bei dem ein bereits gewonnener Teilnehmer einer Studie selbst weitere potenzielle Teilnehmer, die sich ggf. zur Teilnahme an der Studie bereit erklären könnten, rekrutiert. (48)

Elf chirurgische Vertreter (n=11), davon acht (72,7%) männlich und drei (27,3%) weiblich, nahmen im Zeitraum vom 02.02. bis 12.02.2018 an der PV-Trainingssimulation teil. Im weiteren Verlauf der vorliegenden Arbeit werden diese elf Chirurgen als Teilnehmer bezeichnet. Die Teilnehmer wiesen eine mittlere Arbeitserfahrung (Mittelwert, M) von 7,8 Jahren (M=7,8; Standardabweichung SD=9,38; Range 0-33) auf und hatten im Mittel 35,0 PVs (M=35,0; SD= 62,75, Range 0-200) durchgeführt.²¹ 54,5% der Teilnehmer gaben „Unfallchirurgie“ als Fachrichtung an, 27,3% „Orthopädie“ und 18,2% „Unfallchirurgie und Orthopädie“ in Kombination.

Der unter 2.1 dargestellte frakturierte LWK2 war von den Teilnehmern nach einer Demonstration und Erklärung der Simulationsumgebung, sowie einer kurzen Eingewöhnungszeit mit einer beidseitigen PV zu behandeln. Die OP jedes Teilnehmers wurde per Audio- und Videoaufnahmen aufgezeichnet, wobei das Team-Time-Out (TTO) den Startpunkt und das Zurücktreten des Chirurgen vom Behandlungssitus nach Nahtlegung den Endpunkt der Videoaufnahmen kennzeichnete.

Ein Votum der Ethikkommission für die Datenerhebung und -bearbeitung liegt vor (Nr. 773-15).

2.4.2 BGDV-Fragebogen

Es konnten fünf Wirbelsäulenchirurgen (n=5) – im Folgenden Experten genannt – sowohl vom Campus Innenstadt, als auch vom Campus Großhadern der Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie, zur Beantwortung des BGDV-Fragebogens herangezogen werden. Vier (80%) der Experten waren männlichen, eine (20%) weiblichen Geschlechts. Die mittlere Berufserfahrung betrug 17,4 Jahre (M=17,4; SD=10,03; Range 10-36).²²

Die Anzahl bereits wirbelsäulenchirurgisch durchgeführter Eingriffe betrug im Mittel 340 Stück

²¹ Tabelle 5 unter 2.4.4 zeigt die Einzelwerte, die zu diesen Mittelwerten führten unter „ERF_1“ und „ERF_2“.

²² Der Anhang des BGDV-Fragebogens, der diese Daten erfasste, ist unter 7.1 abgebildet.

(M=340; SD=198,49, Range 50-500). Die Anzahl bereits selbstständig durchgeführter PV's lag im Mittel bei 82 Stück (M=82, SD=110,89, Range 10-300), wobei einer der Teilnehmer zwar noch keine PV durchgeführt hatte, stattdessen jedoch 100 Kyphoplastien.²³

Um der Durchführungsobjektivität (s. 1.6.1) Rechnung zu tragen, wurde darauf geachtet, die Durchführung des BGDV-Fragebogens unter möglichst standardisierten Bedingungen durchzuführen. Sie erfolgte in Einzelterminen sowohl in den Seminarräumen des NARVIS-Forschungslabors, als auch in Arztzimmern des Klinikums Großhadern in Anwesenheit zweier an der Konzeption des Fragebogens beteiligter Mitglieder. Für die Einhaltung der standardisierten Bedingungen wurde auf eindeutige Anweisungen für die Bearbeitung des Fragebogens und deren Einhaltung geachtet. Die beiden Mitglieder sollten eventuelle Verständnisfragen vonseiten der Experten beantworten, sowie etwaige verbale Aussagen der Experten vermerken, welche u.U. implizites oder explizites Wissen zum Ablauf einer PV enthielten und im Rahmen der unter 2.2 erläuterten CTA dienlich sein könnten. Die Experten wurden angewiesen, zwischen den Einzelterminen keine Rücksprache untereinander zu halten, um ein möglichst reelles Gesamtbild ihrer individuellen Einstufung der Teilschritte einer PV zu generieren. Für die simulationsbasierten TA (s. Abbildung des BGDV-Fragebogens unter 7.1) stand ein Vertreter der Informatik zur Verfügung, der die Handhabung der virtuellen Simulationen demonstrierte. Die TA 1f) wurde mit dem Programm „Meshmixer“ bearbeitet. Hier konnten die Experten an einer 3D-simulierten, per Mausklick und Pfeiltasten beweglichen Abbildung des LWK2 beidseitig den idealen Eintrittspunkt des Trokars mit grüner Markierung, bzw. den akzeptablen Bereich mit orangener Markierung einzeichnen.

Die TA 1l) wurde mit dem Programm „3D Slicer Trajectories“ bearbeitet. Hier konnten die Experten - wiederum an einem 3D-simulierten, per Mausklick und Pfeiltasten beweglichen LWK2 - die ideale Nadelposition des Trokars in der Endlage einstellen. Indem der LWK2 transparent dargestellt wurde, ließ sich die Endlage des Trokars aus den verschiedenen Ebenen des Raumes betrachten.

Die TA 3a) wurde mit dem Programm „3D Slicer Röntgen Einstellung“ bearbeitet. Hier konnten die Experten an einem 3D-simulierten, per Mausklick und Pfeiltasten beweglichen Lendenwirbelsäulen-Abschnitt eine ideale Einstellung des Bildwandlers vornehmen.

Es wurde keine zeitliche Frist für die Beantwortung des BGDV-Fragebogens vorgegeben.

²³ Der Kyphoplastie-TM wurde in der Anzahl der PV's mit „0“ bewertet, da dies streng genommen der Anzahl seiner bereits durchgeführten Vertebroplastien entspricht. Da es sich bei der Kyphoplastie aber in weiten Teilen um dieselben Inhalte und ein ähnliches Prozedere wie bei der PV handelt, ist seine Einstufung als „Experte“ gerechtfertigt.

2.4.3 PV-Score

Nach Fertigstellung der Konzeption des PV-Scores konnten die Audio- bzw. Videodaten, sowie die von den Informatikern analysierten Simulatordaten der Teilnehmer der PV-Trainingssimulation für die praktische Anwendung des PV-Scores herangezogen werden. Die für die Auswertung verwendete Version des PV-Scores ist unter 3.1.4 abgebildet, da die Fertigstellung des Scores auf dem Expertenkonsensus des BGDV-Fragebogens beruhte und damit sinngemäß den Ergebnissen des BGDV-Fragebogens zuzuordnen ist. Mittels der Items 1A.-4B. des PV-Scores wurden die Daten der Teilnehmer am jeweiligen Expertenkonsensus gemessen. Mithilfe des unter 2.3 aufgezeigten Algorithmus des PV-Scores (\triangleq dem Notenschlüssel, s. Tabelle 4, S.35) resultierte jede Punktzahl eines Teilnehmers in einer Note samt zugehörigem Resultat, Kompetenzniveau und Konsequenz.

2.4.4 Validierungsmethoden Kriteriums- und Konstruktvalidität

Validität stellt das zentrale Kriterium im Assessment von Fähigkeiten (skill assessment) dar. (12)

Laut Downing ist Validität folgendes: „[...] the sine qua non of assessment, as without evidence of validity, assessments in medical education have little or no intrinsic meaning.“ (49)

Daher wurden anhand der ausgewerteten Ergebnisse des PV-Scores Korrelationsanalysen zur Überprüfung der Kriteriums- und Konstruktvalidität (s. 1.6.3) durchgeführt. Für die Überprüfung der Kriteriumsvalidität wurde mittels Spearman-Rangkorrelationen die Berufserfahrung (ERF) der Teilnehmer als Außenkriterium herangezogen, um zu beurteilen, ob diese mit dem Ergebnis des PV-Scores (in Form der ordinalskalierten Noten 1-6) der jeweiligen Teilnehmer korrelierte.

Dabei wurde die Berufserfahrung in drei Ausprägungsformen angegeben (s. Tabelle 5):

- Jahre seit Approbation (ERF_1)
- Anzahl bereits durchgeführter PV's (ERF_2)
- Jahre als selbstständiger, verantwortlicher Operateur von PV's (ERF_3)

Dabei wurde angenommen, dass eine höhere Anzahl an Jahren seit Approbation, bzw. als selbstständiger Operateur, bzw. eine größere Anzahl bereits durchgeführter PV's eine größere chirurgische Kompetenz bedeutet und folglich mit einer niedrigeren Note, resp. einem besseren Ergebnis des PV-Scores einhergehen müsste, sofern statistisch signifikante Zusammenhänge zwischen den genannten Szenarien bestehen.

TM	1	2	3	4	5	6	7	8	9	10	11
ERF_1	9,0	3,0	7,0	1,5	15,0	8,0	3,5	2,0	33,0	4,0	0,0
ERF_2	100	2 *	4	1	50	10	1	2	200	15	0
ERF_3	4	0	0	0	5	3	0	2	10	0	0

Tabelle 5: Berufserfahrung der elf Teilnehmer in drei Ausprägungsformen (ERF_1/2/3), *mit Oberarzt

Zur Überprüfung der Konstruktvalidität wurden die Ergebnisse des PV-Scores (\triangleq Evaluierungstool 1) der Teilnehmer einem Vergleich mit den Ergebnissen des unter 2.2.1 erläuterten TSA/OSATS-Scores (\triangleq Evaluierungstool 2) unterzogen. (37) Der TSA/OSATS-Score nahm damit die Rolle eines bereits etablierten, vergleichbaren Evaluierungstools ein. Da beide Scores für eine Beurteilung der chirurgischen Kompetenz bei der Durchführung einer PV konstruiert und auf dieselben Teilnehmer der PV-Trainingssimulation angewendet wurden, wurde angenommen, dass beide in Bezug auf einen Teilnehmer zum selben Resultat (bestanden bzw. nicht bestanden) führen müssten, gesetzt den Fall, dass eine statistisch signifikante Korrelation vorliegt. Die Korrelation wurde einmal anhand einer polytomen Formulierung des Ergebnisses des PV-Scores (Note 1-6) und einmal anhand einer bivariaten Ergebnisformulierung (1= bestanden, 2= nicht bestanden) getestet. Dazu wurden die Noten der Teilnehmer entsprechend umformuliert: die Noten 1-4 entsprechen der bivariaten Formulierung 1, die Noten 5-6 der bivariaten Formulierung 2.

Die Ergebnisformulierung des TSA/OSATS-Scores verwendet ausschließlich ein bivariates Outcome (1= bestanden, 2= nicht bestanden). Tabelle 6 zeigt die Ergebnisse der Teilnehmer 1-11 nach der Bewertung durch den TSA/OSATS-Score, sowie beide Varianten der Formulierung des PV-Score-Ergebnisses.

TM	1	2	3	4	5	6	7	8	9	10	11
ERG TSA/ OSATS	1	2	1	2	2	1	1	2	1	1	2
ERG PV- Score biv. (polytom)	1 (2)	1 (2)	1 (2)	2 (6)	1 (4)	1 (2)	1 (2)	2 (6)	1 (3)	1 (2)	2 (6)

Tabelle 6: ERG (Ergebnisse) der elf TM des TSA/OSATS-Scores in biv. (bivariater) Formulierung, bzw. des PV-Scores in biv. und polytomer Formulierung (erscheint kursiv, in Klammern), eigene Darstellung

Weitere Validierungsstrategien des PV-Scores sind denkbar, es sei dazu auf die Diskussion am Ende der vorliegenden Arbeit verwiesen.

2.5 Datenspeicherung und Statistik

Das Pilotprojekt bestand aus fünf partizipierenden Experten (n=5) am BGDV-Fragebogen, sowie elf Teilnehmern (n=11) an der PV-Trainingssimulation.

Die im prospektiven Studienteil erhobenen Daten der PV-Trainingssimulation wurden in Rohformaten auf Computern und Servern des Studienteams gespeichert und dann im folgenden Schritt systematisch aufbereitet. Mit den Audio- und Videoaufnahmen aller Operationen der PV-Trainingssimulation wurde ebenso verfahren.

Zusätzlich wurde die Anzahl der angefertigten Röntgenbilder und die benötigte Zeit beim Platzierungsversuch bis zur Endlage des Trokars, die benötigte Gesamtzeit der OP, sowie die Gesamtzahl der angefertigten Röntgenbilder durch ein anwesendes Mitglied der Forschungsgruppe dokumentiert. Die so erfassten Daten wurden anhand einer anschließenden Analyse der intraoperativ erstellten Videoaufnahmen kontrolliert. Auf Basis der Audio- und Videoaufnahmen, sowie durch Analyse des behandelten Knochenmodells und der Simulatordaten wurden schließlich die Metriken der Teilnehmer ausgewertet.

Die im prospektiven Studienteil erhobenen Daten des BGDV-Fragebogens (\triangleq Expertenkonsensus) sowie die des angewendeten PV-Scores (\triangleq Ergebnissen der Teilnehmer der PV-Trainingssimulation) wurden im Tabellenkalkulationsprogramm Microsoft Excel ²⁴ ausgewertet und gespeichert.

Die bearbeiteten BGDV-Fragebögen waren vollständig ausgefüllt und enthielten keine fehlenden oder nicht zuordenbaren Werte. Der Umgang mit fehlenden oder uneindeutigen Werten stellte daher an dieser Stelle keine Beeinträchtigungen der Auswertungsobjektivität (s. 1.6.1) dar.

Die Auswertung der TA mit quantitativem Datengehalt erfolgte ebenfalls mit Microsoft Excel mittels arithmetischer Mittelwertbildung (M), da die untersuchte Stichprobe (die Antworten der Experten) nicht mit Ausreißern verunreinigt war und damit der Mittelwert eine präzisere Kennzahl darstellte als der Medianwert.

Die Auswertung der TA des BGDV-Fragebogens mit qualitativem Datengehalt wurde mittels deskriptiver Statistik durchgeführt. Dabei handelt es sich um eine sogenannte beschreibende Variante der Statistik im Rahmen der Datenanalyse. Diese kann beispielsweise durch graphische Datenaufbereitung (z.B. Diagramme, Häufigkeitstabellen) oder beschreibende Darstellung des Datenmaterials geschehen. Oftmals dient sie der Gewinnung erster Eindrücke bzw. Ideen zur weiteren Analyse. (50)

²⁴ Microsoft Office 365 Business, Version 16.0.12430.20288

Gemäß der strukturierenden, qualitativen Inhaltsanalysemethode, welche für die theoriegeleitete Analyse von Textmaterial geeignet ist, wurden im ersten Schritt die in Form von Wörtern vorliegenden qualitativen Daten der frei formulierbaren TA mit Microsoft Excel nach inhaltlichen Übereinstimmungen untersucht. Dadurch konnten inhaltlich ähnliche Antworten in themenbezogene Kategorien zusammengefasst werden. Im Folgenden wurden anhand konkreter Textstellen Antworten mit Buchstaben (A-D) codiert, um den Datengehalt noch weiter zu konkretisieren. Teilweise fand eine Visualisierung der Kategorien bzw. Codierungen in Tabellen resp. Grafiken statt, um gemeinsame Grundaussagen im Sinne des Expertenkonsensus herauszufiltern. (51)

Abbildung 11 zeigt eine schemenhafte Darstellung dieses Ablaufes. Die im Verlauf des Auswertungsprozesses angefertigten Tabellen resp. Grafiken mit den Kategorisierungen resp. Codierungen der Antworten der Experten lassen sich primär der statistischen Vorarbeit zuordnen, welche dem Herausarbeiten der gemeinsamen Tendenzen der Antworten dienen. Da sie dem Leser keinen weiterführenden Informationsgehalt liefern, werden sie in der vorliegenden Arbeit nicht abgebildet. Stattdessen werden die durch die Inhaltsanalyse eruierten Ergebnisse der qualitativen TA in Form des Expertenkonsensus unter 3.1.2 in verschriftlichter Form dargestellt. (52)

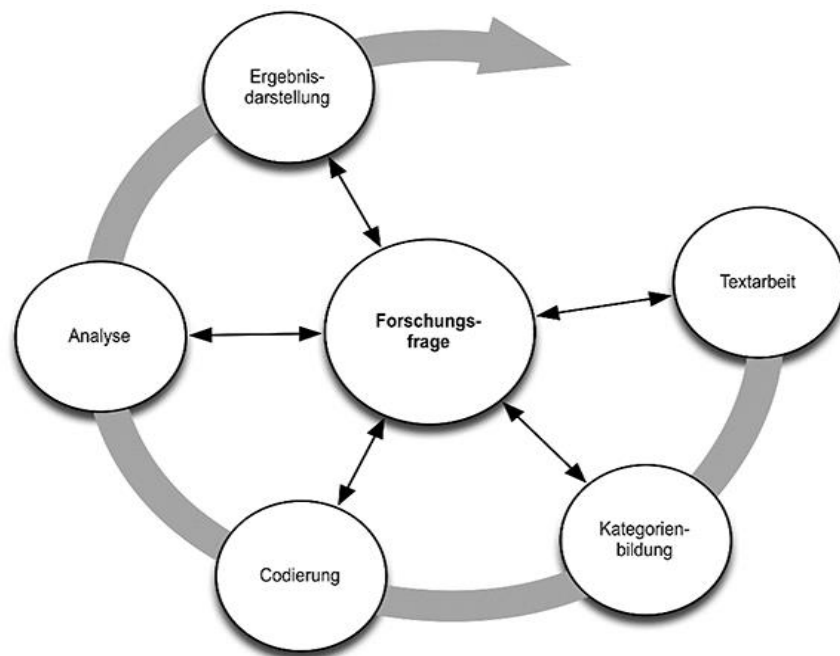


Abbildung 11: generelles Ablaufschema einer qualitativen Inhaltsanalyse, vgl. Kuckartz, 2016, S.45, Abb.4, (52)

Die Bearbeitung, Auswertung und Speicherung der simulationsbasierten Anteile des BGDV-Fragebogens erfolgte für TA 1f) mit dem Programm „Meshmixer“, für TA 1l) mit dem Programm „3D Slicer Trajectories“ und für TA 3a) mit dem Programm „3D Slicer Röntgen Einstellung“. Für genauere Erläuterungen zu diesen Programmen siehe 2.4.2.

Für die Visualisierungen der Expertenconsensus der TA des BGDV-Fragebogens wurde neben Microsoft Excel auch Microsoft Powerpoint ²⁵ verwendet.

Die durch praktische Anwendung des PV-Scores generierten Daten wurden mit Microsoft Excel anhand des vom PV-Score vorgegebenen Algorithmus ausgewertet. Mittel-, bzw. Durchschnittswerte der einzelnen Items sowie der Gesamtnoten wurden mit Microsoft Excel erhoben.

Die Zusammenhangsanalysen zur Validitätsüberprüfung wurden mit dem Statistikprogramm SPSS ²⁶ durchgeführt. Die nichtparametrischen Korrelationstests zur Kriteriumsvalidierung wurden mittels Spearman-Rangkorrelationen berechnet. Die nichtparametrischen Tests zur Konstruktvalidierung wurden im Rahmen des Gruppenvergleichs mittels Mann-Whitney-U-Test für ordinal verteilte, unabhängige Stichproben, sowie Chi-Quadrat-Test durchgeführt. Für die gesamte Arbeit wurde ein Signifikanzniveau von 5% gewählt.

Die statistische Auswertung erfolgte in Kooperation mit dem Institut für Arbeits-, Sozial- und Umweltmedizin der LMU.

²⁵ Version 2001

²⁶ „Statistical Package for the Social Sciences“, 24.0 (IBM Chicago Inc.)

3 Ergebnisse

3.1 Ergebnisse des BGDV-Fragebogens

Da die Dauer der Bearbeitung für die vorliegende Studie unerheblich war, gab es keine definierten Start- und Endpunkte der Bearbeitung. Die benötigte Zeit der Experten betrug im Mittel etwa 30 Minuten.

Im Folgenden werden die Ergebnisse zu den TA abgebildet. Zur Erinnerung sei an dieser Stelle nochmals ein Überblick über die Inhalte der vier Kategorien des BGDV-Fragebogens abgebildet (für einen ausführlicheren Überblick siehe 2.2.2).

1. Trokar-basiert: enthält TA zur Positionierung des Trokars
2. Zement-basiert: enthält TA zur Verteilung des eingebrachten Knochenzements
3. Bildgebungs-basiert: enthält TA zur Handhabung des C-Bogens
4. Anderes: enthält TA zum Zeitfenster für die Gesamtprozedur

Für eine Ansicht der genauen Formulierung der jeweiligen TA siehe wiederum 7.1.

Die Ergebnisse des BGDV-Fragebogens stellen, wie bereits thematisiert, den sog. Expertenkonsensus dar, auf dessen Basis die Bewertung der Items des PV-Scores erfolgte.

3.1.1 quantitative Teilaufgaben

1a)

Kritizitätskat.	Risikostruktur	Kritizität
a	Aorta	1,4
a	Rückenmark	2,4
a	Vena cava	3,4
a	Lunge	4,8
a	Spinalkanal	5,4
a	SpNW	5,6
b	Segmentgefäße	6,4
c	ant. WK-Wand	8,6
c	Bandscheiben	9,4
c	Deck-Bodenplatte	9,6
d	anderer WK	9,6
c	Facettengelenk	11,6
d	Muskulatur	12,8

Abbildung 12: Ranking der Kritizitätsangaben und Kritizitätskategorien bei Punktion einer Risikostruktur, (aufsteigend, 1,0=am kritischsten, 13,0=am unkritischsten), die Kritizitätskategorien a-d wurden empirisch festgelegt

Wie aus der Abbildung des BGDV-Fragebogens unter 7.1 ersichtlich wird, haben die Kritizitätskategorien a-d folgende Bedeutung:

a = Verletzung der Patientensicherheit, fügt dem Patienten Schaden zu
b = Gefährdung der Patientensicherheit
c = nicht optimal, aber keine Gefährdung der Patientensicherheit
d = unkritisch

Die Kritizitätsangaben der Risikostrukturen in Abbildung 12 (1,4-12,8) stimmen mit der Einordnung in die Kritizitätskategorien (a-d) weitgehend überein. Lediglich die „Punktion eines anderen Wirbelkörpers“, welche mit 9,6 Punkten auf gleicher Kritizitätsstufe wie die „Punktion der Deck-/Bodenplatte“ liegt, rangiert nicht in der entsprechenden Kritizitätskategorie c, sondern wird neben der Muskulatur (gemeint ist hier operationssitus-ferne Muskulatur) in die Gruppe d eingeordnet.

1b) Kritik/Anregungen zur Formulierung der vier Kritizitätskategorien (a-d):

Keine, 100% der Experten (5 von 5) erachten sie als sinnvoll.

1c) Angaben weiterer relevanter Risikostrukturen in Höhe LWK2:

Keine: 80% der Experten (4 von 5)

Nennung von Leber, Milz und abdominalen Organen: 20% der Experten (1 von 5).

1d) Angaben zur Kritizität des Nadel-Abstands zu anatomischen Strukturen:

Struktur	Dimension im behandelten WS-Abschnitt (mm)	unkritisch bis zu einem Abstand von ... mm zur genannten Struktur	bedingt kritisch bis zu einem Abstand von... mm zur genannten Struktur (bzw. ohne gravierende Auswirkungen aufs Patienten-Outcome)	Kritisch ab einem Abstand von ... mm zur genannten Struktur
Rückenmark	6,7	2,3	1,3	0,5
Spinalkanal	10,0	1,4	1	0,5
Aorta	15,9	4,5	2,5	0,8
V. cava	6,7	4,5	2,5	0,8
Spinalnervenzurwurzeln	3,7	3,0	1,6	0,6
Lunge	k.A.	4,3	2,3	0,5

Segment-Gefäß	4,3	4,3	2,5	1
Bandscheiben	19,0	2,7	2,3	1
Facettengelenk	5,5	1	0,3	0
Anderer Wirbelkörper	32,3	5	2,5	1

Tabelle 7: Kritizitätsabstufungen (in mm) im Umfeld einer potenziell punktierten Risikostruktur

Die Daten eines Experten hatten für die Kritizitätsabstufungen pro Risikostruktur jeweils den gleichen Betrag. Demnach handelte es sich dabei um Datenpunkte, die als Ausreißer zu bezeichnen sind und von der Auswertung ausgeschlossen wurden, da es sich um sinnfreie Angaben handelte, die das Ergebnis verfälscht hätten.

Diese TA konnte nicht in den PV-Score aufgenommen werden, da in der PV-Trainingssimulation nur ausgewertet werden konnte, ob eine Risikostruktur punktiert wurde (Ja oder Nein).

1e) Rating von Punktionsparametern (1= sehr wichtig bis 5= vollkommen unwichtig):

Art der RS:	1,0
Häufigkeit der Punktion der RS:	1,4
Strecke/Tiefe der Punktion der RS:	1,6

Das Expertenranking zeigt, dass die Art der Risikostruktur wichtiger als die Häufigkeit ihrer Punktion und diese wiederum wichtiger als die Tiefe der Punktion ist. Allen drei Punktionsparameter wird durch ihre Bepunktung zwischen 1 und 2 eine sehr hohe bis hohe Wichtigkeit eingeräumt.

1h) Einstufung der Abweichung (in mm) vom idealen Eintrittspunkt (EP), pro Seite:

ideal bis zu: 2,0	nicht ideal, aber akzeptabel bis zu: 4,0	nicht akzeptabel ab: > 4,0
----------------------	---	-------------------------------

1i) Einstufung der Knochenansatzpunkte (Anzahl) des Trokars beim Platzierungsversuch, pro Seite:

angemessen: ≤ 5	zu viel: ≥ 6
-----------------	--------------

1j) Einstufung der erneuten Hautperforationen (Anzahl) durch Zurückziehen des Trokars beim Platzierungsversuch, pro Seite:

angemessen: ≤ 2	zu viel: ≥ 3
----------------------	-------------------

Diese TA konnte aus Gründen der Nichtauswertbarkeit der entsprechenden Metrik der Teilnehmer der PV-Trainingssimulation nicht in den PV-Score aufgenommen werden.

1k) Einstufung der Röntgenbilder (Anzahl) und Zeit (Min.) beim Platzierungsversuch bis zur Endlage des Trokars, pro Seite:

Röntgenbilder, angemessen: ≤ 11	Röntgenbilder, zu viel: ≥ 12
Zeit, angemessen: $\leq 3,00$	Zeit, zu viel: $\geq 3,01$

1m)/n) Die ursprünglichen TA „Gibt es einen minimalen Abstand der Nadelspitze zur Deck-/Bodenplatte, der eingehalten werden sollte?“ und „Gibt es einen minimalen Abstand der Nadelspitze vom anterioren WK-Rand, der eingehalten werden sollte?“ wurden aus Gründen der Praktikabilität und Auswertbarkeit der Simulatordaten zusammenfassend umformuliert in:
„Gibt es einen minimalen Abstand der Nadelspitze zu irgendeiner der Wände des Wirbelkörpers, der eingehalten werden sollte?“

Einstufung des einzuhaltenden Abstandes (mm) der Trokarspitze zu einer der Wirbelkörperwände, pro Seite:

angemessen: $\geq 3,0$	Zu wenig: $<3,0$
------------------------	------------------

2a) Einstufung der Gesamtzementmenge (ml) für den behandelten LWK2:

zu wenig: $\leq 2,0$	passend: $2,1-6,6$	zu viel: $\geq 6,7$
----------------------	--------------------	---------------------

2b) (*quantitativer Anteil*) generelle Einstufung der injizierten Gesamtzementmenge:

100% der Experten (5 von 5) werten eine zu hohe Menge an injiziertem Zement negativer als eine zu geringe Menge.

2c) (*quantitativer Anteil*) 80% der Experten (4 von 5) geben an, dass sich etwas an der Kritizität einer der genannten Strukturen ändert, je nachdem, ob sie von Trokarpunktion oder Zementleckage betroffen wird. 20% (1 von 5) gibt an, dass sich dadurch nichts an der Kritizität einer Struktur ändert.

2d)

27

Kritizitätskat.	Risikostruktur	Kritizität
a	Subduralraum	1
b	intraspinale Venen	2,8
b	prävertebrale Venen	3,2
a	Segmentgefäße	4
c	Bandscheiben	4,4
d	prävertebraler Raum	4,8

Abbildung 13: Ranking der Kritizitätsangaben und -abstufungen, (aufsteigend, 1=am kritischsten, 4,8=am unkritischsten), die Kritizitätskategorien a-d wurden empirisch festgelegt

Nennung weiterer bei einer Zementleckage relevanter Risikostrukturen in Höhe des LWK2:

Keine (100%).

2e) Einstufung der idealerweise zu injizierenden Zementmenge (ml) vor Anfertigung eines Röntgen-Kontrollbildes:

0,98

Die TA 2c)/d)/e) konnten aus Gründen der Nichtauswertbarkeit der entsprechenden Metriken der Teilnehmer der PV-Trainingssimulation nicht in den PV-Score aufgenommen werden.

2f) (quantitativer Anteil) 100% der Experten (5 von 5) werten die Anfertigung eines Röntgenkontrollbildes nach einer zu großen injizierten Zementmenge negativer als ein nach einer zu geringen Zementmenge angefertigtes.

3b) Einstufung der während der Gesamtprozedur angefertigten Röntgenbilder (Anzahl), pro Seite:

angemessen: ≤ 25	zu viel: ≥ 26
-----------------------	--------------------

3c) (quantitativer Anteil) Einstufung der Menge angefertigter Röntgenbilder:

„Zuviel Fluoroskopiebilder negativer zu werten“: 20% (1 von 5)

„Zuw wenig Fluoroskopiebilder negativer zu werten“: 80% (4 von 5)

²⁷ Für eine Erläuterung der Kritizitätskategorien a-d siehe TA 1a) oder 7.1.

4a) Definieren eines geeigneten Zeitfensters (Min.) für die Durchführung einer PV:

optimal:	$\geq 18,0$ bis $\leq 41,0$	zu lang:	$\geq 41,01$
----------	-----------------------------	----------	--------------

3.1.2 qualitative Teilaufgaben

1g) Definition des „idealen Eintrittspunktes“ (IEP) in den Pedikel:

- Alle Experten orientieren sich für die Definition des IEP am Pedikel.
Dabei ist es nicht von Bedeutung, ob der C-Bogen streng sagittal oder in der Achse des Pedikels eingestellt wird.
- Ein Großteil der Experten orientiert sich beim idealen EP etwas dezentral der Pedikelachse in die Nähe der lateralen Pedikelwand.
- Aus der Aussage von einem Experten lässt sich noch eine weitere Information bzgl. des Durchmessers des IEP's ablesen: die Fläche, in der der IEP liegt, sollte maximal so groß wie der kleinste Pedikel-Innendurchmesser sein.

1l) (*qualitativer Anteil*) Definition einer Ebene, anhand derer eine Bewertung der Tiefe der Trokarspitze in der Endlage möglich ist:

- Die Nadelspitze sollte in der seitlichen Ansicht des Wirbelkörpers im ventralen Drittel liegen und einen Abstand von circa 0,5 bis 1,00 cm zur projizierten vorderen Wirbelkörper-Begrenzung haben.

2b) (*qualitativer Anteil*) Begründung für die negativere Einstufung von einer zu hohen Menge an injiziertem Zement im Vergleich zu einer zu geringen Menge:

- Die zunehmende Gefahr des Zementaustrittes bei einer zu hohen Zementmenge erhöht das Komplikationsrisiko.

2c) (*qualitativer Anteil*) Nennung der Strukturen, deren Kritizität sich ändert, je nachdem, ob sie von einer Punktion oder von einer Zementleckage betroffen wurden:

- Wird der Spinalkanal von einer Zementleckage betroffen, kann es zu Thermoschäden an den nervalen Strukturen kommen.
- Bei einer Zementleckage in Gefäße resp. Venen, deren Punktion durch den Trokar folgenlos bleibt, kann es zu venöser Embolie kommen.

- Die Punktion der Bandscheiben bleibt folgenlos, was sich bei Zementleckage in die Bandscheiben ändern kann.

2f) (qualitativer Anteil) Begründung für die negativere Einstufung der Anfertigung eines Röntgenkontrollbildes nach einer zu großen injizierten Zementmenge im Vergleich zu einer zu geringen.

- Ein zu spät angefertigtes Röntgenkontrollbild kann durch möglicherweise zu spät wahrgenommenen Zementaustritt gravierende Folgen haben. Dementsprechend ist ein Röntgenkontrollbild nach zu geringer Zementmenge einem nach zu großer Zementmenge vorzuziehen.

3a) (qualitativer Anteil) Definition der „passenden Einstellung“ des C-Bogens.

- sagittale Projektion: Der C-Bogen soll streng-seitlich eingestellt werden, sodass die Deck- und Bodenplatte des Wirbelkörpers orthograd getroffen werden und sich linienförmig darstellen.
- AP-Projektion: Der Wirbelkörper soll mittig positioniert sein, sodass die Processi spinosi mittig und zentral abgebildet werden.

Diese TA konnte aus Gründen der Nichtauswertbarkeit der entsprechenden Metrik der Teilnehmer der PV-Trainingssimulation nicht in den PV-Score aufgenommen werden.

3c) (qualitativer Anteil) Begründung für die negativere Einstufung einer zu geringen Menge an Fluoroskopiebildern im Vergleich zu einer zu hohen:

- Ein Zuwenig an angefertigten Fluoroskopiebildern ist negativer zu werten als ein Zuviel, da etwaige Komplikationen wie z.B. Zementaustritt womöglich zu spät erkannt werden.
- Eine höhere Strahlenbelastung wird zugunsten der Prozedur-Sicherheit toleriert.

3.1.3 simulationsbasierte Teilaufgaben

1f) Die Abbildungen 14, 15 und 16 zeigen die idealen, bzw. akzeptablen EP's in die Pedikel (IEP=grün markiert, akzeptabler EP=orange-gelb markiert).

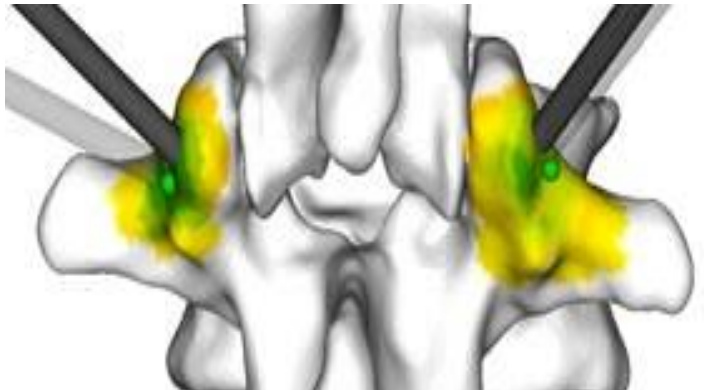


Abbildung 14: ideale und akzeptable Eintrittspunkte des Trokars in die Pedikel, beidseitig

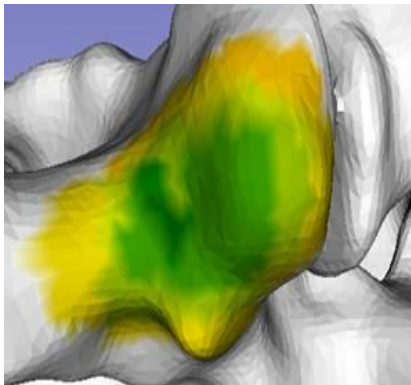


Abbildung 15: idealer und akzeptabler Eintrittspunkt des Trokars in den linken Pedikel

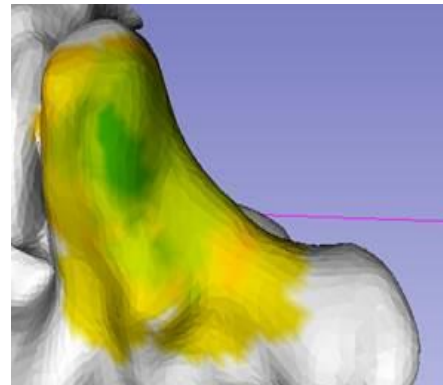


Abbildung 16: idealer und akzeptabler Eintrittspunkt des Trokars in den rechten Pedikel

1I) (simulationsbasierter Anteil)

Die Tabellen 8 und 9 zeigen die ideale Nadelposition beim bipedikulären Zugang für beide Pedikel, je zweimal in AP-Projektion (Proj.) (Ansicht von anterior, Ansicht von posterior), und zweimal in sagittaler Projektion (Ansicht von links, Ansicht von rechts).

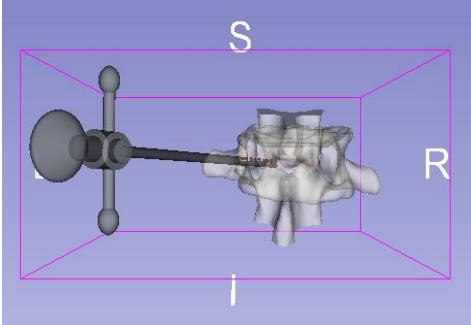
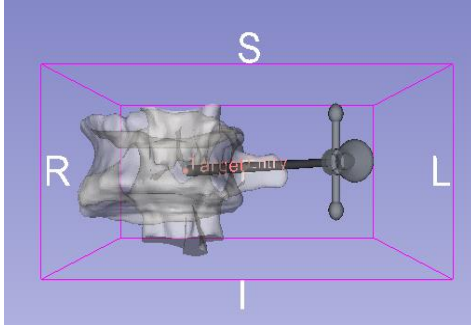
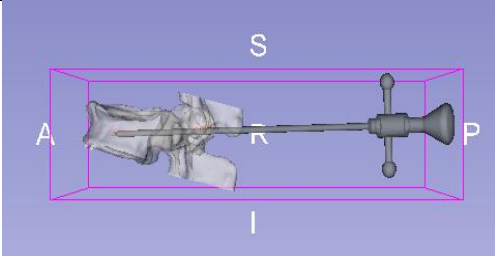
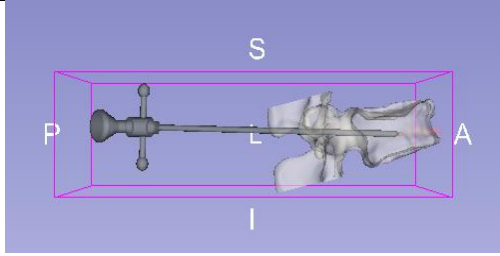
Linker Pedikel			
Proj.	Ansicht von posterior	Ansicht von anterior	
AP			
sag.	Ansicht von links	Ansicht von rechts	
			

Tabelle 8: ideale Nadelposition im linken Pedikel in AP und sagittaler Projektion

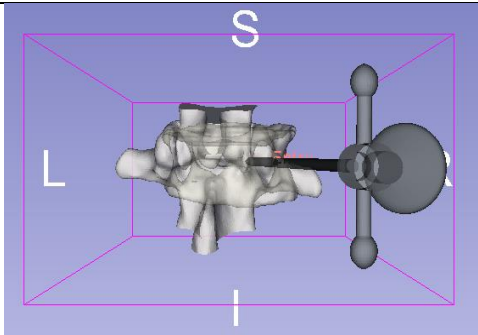
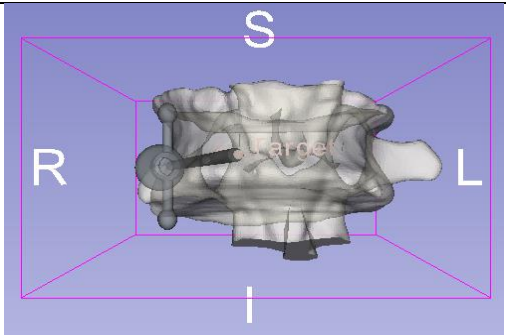
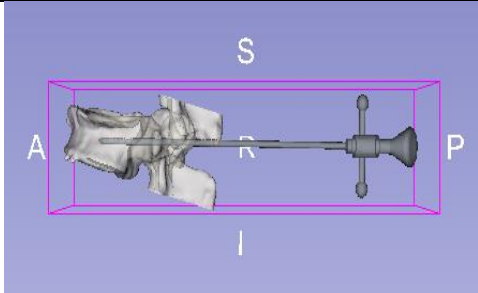
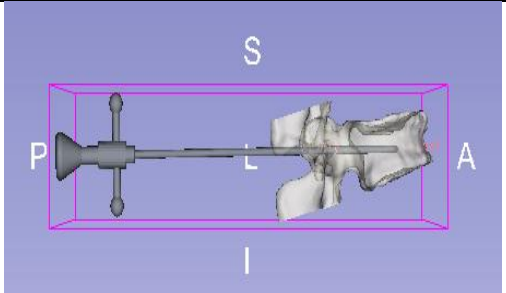
Rechter Pedikel		
Proj.	Ansicht von posterior	Ansicht von anterior
AP		
sag.		

Tabelle 9: ideale Nadelposition im rechten Pedikel in AP und sagittaler Projektion

3a) (*simulationsbasierter Anteil*) passende Einstellung des Bildwandlers, in AP und sagittal:

Diese TA wurde nicht im Sinne eines Expertenkonsensus ausgewertet. Daher zeigt die Tabelle 10 Ausschnitte der „passenden Einstellung des C-Bogens“, die die Experten in der Computersimulation eingestellt haben.



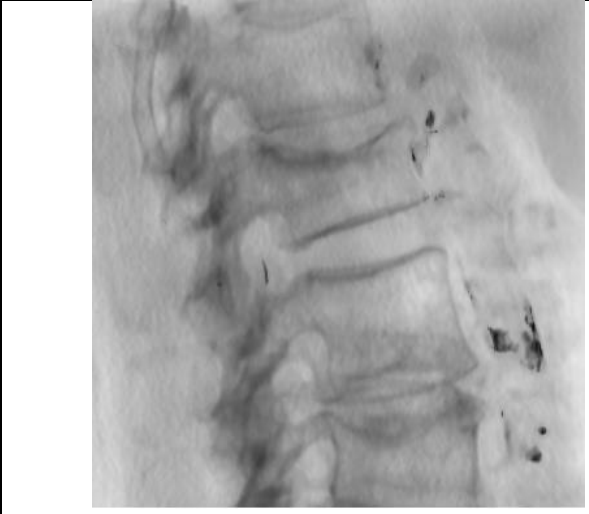
Proj.		
AP		
sag.	Ansicht von links	Ansicht von rechts
		

Tabelle 10: passende Einstellung des Bildwandlers, in AP und sagittaler Projektion

3.1.4 Abbildung des PV-Scores

Nachdem die Konzeption des PV-Scores auf Basis des Expertenkonsensus erfolgte und damit auf die Ergebnisse des BGDV-Fragebogens angewiesen war, wird an dieser Stelle die in der vorliegenden Arbeit entwickelte Version des PV-Scores abgebildet.

Perkutane-Vertebroplastie-Score (PV-SCORE) zur Bewertung chirurgischer Kompetenz anhand der Items 1A.- 4B.

Die Daten der PV-Durchführung eines Teilnehmers (TM) sind durch die folgenden Richtwerte der Items **1A.-4B.** mit den entsprechenden Punktwerten zu versehen und diese unter „**Pkt.ges.(TM)**“ zu vermerken. Bei Items (1A., 2A., 3B., 4A., 4B.) mit getrennter Betrachtung der rechten und linken Seite (*L, R getrennt*) ergibt sich der Gesamtpunktwert des Items „**Pkt.ges.(TM)**“ durch Addition von L und R. Die in den Items **1A.-4B.** erreichten Punkte „**Pkt.ges.(TM)**“ sind zu einer PV-Gesamtpunktzahl „**PV-Pkt.ges.(TM)**“ zu addieren. Die Note des PV-Scores, sowie Resultat, Kompetenzniveau und Konsequenz ergeben sich durch die Bewertung der „**PV-Pkt.ges.(TM)**“ durch den beigefügten Notenschlüssel.

TTO=Team-Time-Out Rö=Röntgen RS=Risikostruktur
 PLV=Platzierungsversuch EK=Expertenkonsensus KK=Killerkriterium
 EdT=Endlage des Trokars Pkt.=Punkte IEP= idealer Eintrittspunkt
 SW=Simulationswiederholung TP=Trainingsprogramm
 PLV bis EdT= [Hautinzision] bis [Rö-Kontrollbild vor Zementinjektion]
 Gesamt-OP= [Start TTO] bis [Zurücktreten des Chirurgen nach Nahtlegung]

1. Zeit

1A. benötigte Zeit (Min.) beim PLV bis zur EdT (*L, R getrennt*):

Min.	≤ 3,00	≤ 6,00	≤ 9,00	≤12,00	≤ 15,00	> 15,00
Pkt.	0	0,2	0,4	0,6	0,8	1
L						
R						
Pkt. ges. (TM)						

1B. benötigte Zeit (Min.) für die Durchführung der Gesamt-OP:

Min.	< 18,00 *	≤ 18,00 ≤ 41,00	≥ 41,01
Pkt.	0 (2)*	0	1
Pkt. ges. (TM)			

*bei einer Gesamtpunktzahl des Teilnehmers ≥ 10,8 Pkt. erhält ein Teilnehmer bei < 18 Min. nachträglich 2 Pkt.

2. Effizienz

2A. Anzahl (Stück) angefertigter Rö-Bilder beim PLV bis zur EdT (*L, R getrennt*):

Stück	≤ 11	≤ 22	≤ 33	≤ 44	≤ 55	> 56
Pkt.	0	0,2	0,4	0,6	0,8	1
L						
R						
Pkt. ges. (TM)						

2B. Anzahl (Stück) angefertigter Rö-Bilder während der Gesamt-OP:

Stück	≤ 50	≤ 75	≤ 100	≤ 125	≤ 150	> 150
Pkt.	0	0,2	0,4	0,6	0,8	1
Pkt. ges. (TM)						

3. Sicherheit

3A. Punktion einer der folgenden RS:

RS ^a	keine	Gruppe a*	Gruppe b**	Gruppe c***	Gruppe d****	KK*****
Pkt.	0	4	2,5	1	0,5	direkt Gesamtnote 6
Pkt. ges. (TM)						

* Gruppe a: V. cava, Spinalkanal, Lunge

** Gruppe b: Segmentgefäße

*** Gruppe c: Deck-/Bodenplatte, Bandscheiben, Facettengelenk, ant. WK-Wand

**** Gruppe d: anderer WK, Operationssitus-ferne Muskulatur

***** KK: Rückenmark, Spinalnervenzurzen, Aorta

^a bei Mehrfachpunktion einer Struktur oder Punktion mehrerer Strukturen sind die Punkte zu addieren

3B. Minimaler eingehaltener Abstand (mm) zu einer der Wirbelkörperwände (Deck-/Bodenplatte, bzw. ant. WK-Wand), (*L, R getrennt*):

mm	≥ 3,00	< 3,00
Pkt.	0	1
L		
R		
Pkt. ges. (TM)		

3C. Gesamtmenge (ml) an appliziertem Zement:

ml	≤ 2,0	≥ 2,1 ≤ 6,6	> 6,7
Pkt.	1	0	2
Pkt. ges. (TM)			

4. Geschicklichkeit

4A. Abweichung (mm) vom IEP (*L, R getrennt*):

mm	≤ 2,00	≥ 2,01 ≤ 4,00	> 4,01
Pkt.	0,5	0	1
L			
R			
Pkt. ges. (TM)			

4B. Anzahl (Stück) der Ansatzpunkte der Nadel am Knochen beim PLV (*L, R getrennt*):

Stück	≤ 5	≥ 6
Pkt.	0	1
L		
R		
Pkt. ges. (TM)		

Notenschlüssel:

PV-Pkt.ges.	0,0 - 3,6	3,7 - 7,2	7,3 - 10,8	10,9 - 14,4	14,5 - 18,0	> 18,0/KK
PV-Pkt.ges (TM)						
Note	1	2	3	4	5	6
Kompetenz-niveau	Sehr kompetent	Kompetent	Befriedigend kompetent	Ausreichend kompetent	Mangelhaft kompetent	Ungenügend kompetent
Resultat	Bestanden	Bestanden	Bestanden	Bestanden	Nicht bestanden	Nicht bestanden
Konsequenz	Keine	Keine	Keine	Keine, TP angeraten	SW, zuvor TP optional	SW, zuvor TP obligat

Abbildung 17: Perkutane-Vertebroplastie-Score (PV-Score)

3.2 Ergebnisse des PV-Scores

Die „Ergebnisse des PV-Scores“ bezeichnen die erzielten PV-Score-Resultate der Teilnehmer, welche sich durch Anwendung der Items 1A.-4B. des PV-Scores auf die ausgewerteten Daten der PV-Trainingssimulation ergaben. Die TM 3, 4, 8 und 11 begannen mit einer Instrumentierung am linken Pedikel, die übrigen Teilnehmer starteten rechts. Die Tabellen 11 bis 19 zeigen pro Item jeweils die entsprechenden Werte der TM1-11, sowie in der rechten Spalte den Durchschnittswert (M) der jeweiligen Metrik. Tabelle 20 zeigt die Auswertung der Punktzahlen der Teilnehmer, die jeweils in einer entsprechenden Note resultieren.

1A. Tabelle 11 zeigt die benötigte Zeit (Min.) beim PLV bis zur EdT (*L, R getrennt*).

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
PLV-EdT [min] L	4,45	20,62	10,78	10,21	11,64	4,79	13,87	20,19	19,73	11,88	13,8	12,82
PLV-EdT [min] R	10,14	10,28	6,34	4,22	15,74	4,76	20,83	10,99	8,08	17,94	25,9	12,29
Pkt. L	0,2	1	0,6	0,6	0,6	0,2	0,8	1,0	1	0,6	0,8	0,67
Pkt. R	0,6	0,6	0,4	0,2	1	0,2	1,0	0,6	0,4	1	1,0	0,64
Σ Pkt.	0,8	1,6	1	0,8	1,6	0,4	1,8	1,6	1,4	1,6	1,8	1,31

Tabelle 11: benötigte Zeit beim Platzierungsversuch

1B. Tabelle 12 zeigt die benötigte Zeit (Min.) für die Durchführung der Gesamt-OP.

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
Ges. dauer [min]	17,26	34,7	20,5	19,62	30,32	9,68	38,32	32,04	27,97	33,6	41,87	27,81
Σ Pkt.	0	0	0	0	0	0	0	0	0	0	1	0,1

Tabelle 12: Dauer der Gesamt-OP

2A. Tabelle 13 zeigt die Anzahl (Stück) angefertigter Röntgenbilder beim PLV bis zur EdT (*L, R getrennt*).

	TM1	TM 2	TM 3	TM 4	TM5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
Röntgen-Bilder [St.] L	22	75	36	29	52	37	57	60	94	32	33	47,91
Röntgen-Bilder [St.] R	34	17	50	22	47	14	49	32	65	31	89	40,91
Pkt. L	0,2	1,0	0,6	0,4	0,8	0,6	1,0	1,0	1,0	0,4	0,4	0,67
Pkt. R	0,6	0,2	0,8	0,2	0,8	0,2	0,8	0,4	1,0	0,4	1,0	0,58
Σ Pkt.	0,8	1,2	1,4	0,6	1,6	0,8	1,8	1,4	2,0	0,8	1,4	1,25

Tabelle 13: Anzahl angefertigter Röntgenbilder beim Platzierungsversuch

2B. Tabelle 14 zeigt die Anzahl (Stück) angefertigter Röntgenbilder während der Gesamt-OP.

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
Röntgen-Bilder [St.] ges.	97	136	168	93	163	75	144	124	200	102	160	132,91
Σ Pkt.	0,4	0,8	1,0	0,4	1,0	0,2	0,8	0,6	1,0	0,6	1,0	0,71

Tabelle 14: Anzahl angefertigter Röntgenbilder während Gesamt-OP

3A. Punktion einer der folgenden RS:

Anmerkung: Für die Auswertung lagen die Risikostrukturen als 3D Volumenkörper vor.

Davon wurden zwei „distance maps“ erstellt, von denen eine bzgl. der Außenhülle einer Risikostruktur nach innen ging (RS_innen) und eine bzgl. der Außenhülle nach außen (RS_außen). Dann wurde für jeden Punkt (Voxel) der Trokarspitze ein Abstand zur nächsten Kontur/Außenhülle der Struktur ausgerechnet.

Die Angabe „0“ in der Tabelle 15 bedeutet, dass die jeweilige Risikostruktur nicht punktiert wurde. Eine Angabe von „3,20“ wie bei TM 4 beim RM bedeutet, dass das RM 3,20 mm tief punktiert wurde (bezieht sich auf „RM_innen“ durch die nach innen führende distance map). Bei der Begutachtung einer etwaigen Punktion der anterioren Wirbelkörperwand (Ant. WKW) hingegen wird die nach

außen führende distance map benötigt. Die Angabe von „0,71mm“ wie bei TM 5 bedeutet demnach ein „Ausbrechen“ der Trokarspitze um 0,71mm (bezieht sich auf „Ant.WKW_außen“), welches in Bezug auf die Ant. WKW gleichbedeutend mit einer „Punktion der Ant.WKW“ gewertet wird.

In der Tabelle werden nur die Risikostrukturen abgebildet, die durch mindestens einen TM von einer Punktion betroffen waren. Durchschnittswerte (M) werden in dieser TA nicht angegeben.

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM11
RM	0	0	0	3,20	0	0	0	1,5	0	0	2,69
Ant. WKW	0	0	0	0	0,71	0	0	0	0	0	0
BS	0	0	0	0	0,5	0	0	0	0	0	0
Fac. gelenk	0	0,5	0	0	0	0	0	0	0	0	0,5
Σ Pkt.	0	1	0	Note 6	2	0	0	Note 6	0	0	Note 6

Tabelle 15: Punktion von Risikostrukturen

Die Note 6 ergibt sich bei TM 4, 8 und 11 durch das Erfüllen eines Killerkriteriums in Form der Punktion des Rückenmarks. Die zusätzliche Punktion des Facettengelenks durch TM 11 erfährt keine gesonderten Konsequenzen. Die Punktion des Facettengelenks, bzw. der anterioren Wirbelkörperwand und der Bandscheibe durch TM 2, bzw. TM 5 ergibt sich durch Einordnen dieser Risikostrukturen in Gruppe c des Items **3A.** des PV-Scores, welche eine Punktzahl von „1“ pro punktierter Risikostruktur vorsieht.

3B. Tabelle 16 zeigt die minimal eingehaltenen Abstände (mm) zu einer der Wirbelkörperwände (Deck-/Bodenplatte, bzw. Ant. WKW), (L, R getrennt):

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM11	M
[mm] L	9,51	6,44	4,42	11,07	0	7,97	6,02	5,79	5,72	3,5	7,35	6,16
[mm] R	7,12	6,75	3,5	5,02	1,58	7,45	4,36	4,5	0,5	3,5	2,45	4,25
Pkt. L	0	0	0	0	1	0	0	0	0	0	0	0,09
Pkt. R	0	0	0	0	1	0	0	0	1	0	0	1,18
Σ Pkt.	0	0	0	0	2	0	0	0	1	0	0	0,27

Tabelle 16: minimaler eingehaltener Abstand der Trokarspitze zu einer der Wirbelkörperwände

3C. Tabelle 17 zeigt die Gesamtmenge an appliziertem Zement. Die genauen Einzelwerte an appliziertem Zement eines Teilnehmers pro Seite (L und R getrennt) konnten nicht ermittelt werden.

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
[ml], ges.	6,06	5,6	6,04	3,86	2,18	7	3,22	4,9	10,66	5,06	7,58	5,65
Σ Pkt.	0	0	0	0	0	2	0	0	2	0	2	0,55

Tabelle 17: Gesamtmengen an appliziertem Zement

4A. Tabelle 18 zeigt die Abweichungen der Trokarspitze (mm) vom IEP (*L, R getrennt*).

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
[mm] L	4,4	5,75	2,71	8,93	11,24	8,29	10,41	3,87	5,09	8,97	2,76	6,58
[mm] R	4,32	1,62	8,38	5,69	6,97	2,3	6,65	10,62	10,06	7,91	9,65	6,74
Pkt. L	1	1	0,5	1	1	1	1	0,5	1	1	0,5	0,86
Pkt. R	1	0	1	1	1	0,5	1	1	1	1	1	0,86
Σ Pkt.	2	1	1,5	2	2	1,5	2	1,5	2	2	1,5	1,73

Tabelle 18: Abweichungen der Trokarspitze vom idealen Eintrittspunkt in den Pedikel

4B. Tabelle 19 zeigt die Anzahl (Stück) der Ansatzpunkte der Nadel am Knochen beim PLV (*L, R getrennt*).

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11	M
[St.] L	2	12	1	1	6	2	5	13	4	4	4	4,91
[St.] R	2	1	1	1	2	2	4	5	2	3	10	3
Pkt. L	0	1	0	0	1	0	0	1	0	0	0	0,27
Pkt. R	0	0	0	0	0	0	0	0	0	0	1	0,09
Σ Pkt.	0	1	0	0	1	0	0	1	0	0	1	0,36

Tabelle 19: Anzahl der Ansatzpunkte des Trokars am Knochen

Auswertung:

	TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7	TM 8	TM 9	TM 10	TM 11
Σ Pkt. Items	4	6,6	4,9	-	11,2	4,9	6,4	-	9,4	5	-
Note	2	2	2	6	4	2	2	6	3	2	6

Tabelle 20: Auswertung der Punktzahlen der Teilnehmer anhand des PV-Scores

Aus den Noten der Tabelle 20 ergibt sich ein Notendurchschnitt von 3,36. Die Verteilung der Noten präsentiert sich entsprechend der Abbildung 18:

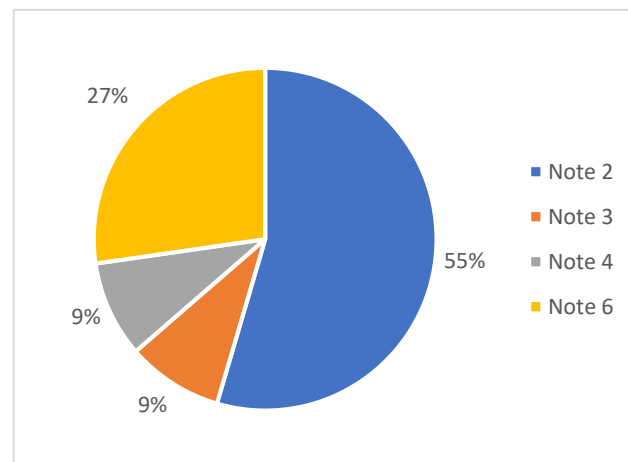


Abbildung 18: prozentuale Häufigkeiten der Noten des PV-Scores, eigene Darstellung

3.3 Ergebnisse der Validierungsmethoden

3.3.1 Kriteriumsvalidität

Der Spearman-Korrelationskoeffizient ρ (Rho) stellt ein Maß für die Effektstärke dar. Die Effektstärke ist ein Maß für die Stärke eines Phänomens (z.B. eines statistisch signifikanten Ereignisses), welche dazu dient, die praktische Relevanz resp. den Nutzen dieses Phänomens einordnen zu können. (53)

Nach Cohen signalisiert ein Korrelationskoeffizient von:

$\rho = .10$ eine geringe Korrelation

$\rho = .30$ eine mittlere Korrelation und

$\rho = .50$ eine hohe Korrelation. (54) (55)

Ein positives Vorzeichen lässt erkennen, dass es sich um eine gleichsinnige Beziehung der beiden Variablen handelt. (56) Demzufolge bedeutet ein negatives Vorzeichen eine gegenläufige Beziehung, welche im vorliegenden Fall zu erwarten wäre, falls es einen statistisch signifikanten Zusammenhang

zwischen der Note des PV-Scores und der Erfahrung der Teilnehmer geben sollte: höhere Werte der Erfahrung eines Teilnehmers gingen dementsprechend mit niedrigeren Werten der Note im PV-Score einher. Da es keine Annahme über die Richtung der Korrelation gab, d.h. ob sie negativ oder positiv sein würde, wurde ein zweiseitiger Signifikanztest (Sig. 2-seitig) gefordert.

Die Werte auf der Diagonalen der Tabelle 21 zeigen den Zusammenhang jeder Variable mit sich selbst. Diese Korrelation beträgt stets 1, da jede Variable perfekt mit sich selbst korreliert ist. (56) Tabelle 21 zeigt die Spearman-Korrelationskoeffizienten Rho (ρ) und die p-Werte (Signifikanz Sig. 2-seitig) für die Korrelation der Noten des PV-Scores mit den drei Ausprägungsformen der Erfahrung (ERF_1/2/3) der Teilnehmer. Mit keiner der drei Ausprägungsformen der Berufserfahrung ergibt sich statistische Signifikanz (p-Werte alle $> 0,05$). Darüber hinaus sind folgende Zusammenhänge erkennbar:

- ERF_1: $\rho = -0,43$; p-Wert = 0,19

Die Note des PV-Scores korreliert in mittlerem Maße mit der ERF_1 der Teilnehmer.

- ERF_2: $\rho = -0,34$; p-Wert = 0,30

Die Note des PV-Scores korreliert in mittlerem Maße mit der ERF_2 der Teilnehmer.

- ERF_3: $\rho = 0,10$; p-Wert = 0,76

Die Note des PV-Scores korreliert in geringem Maße mit der ERF_3 der Teilnehmer.

Bei sehr kleinen Stichproben sollte man tendenziell allerdings davon ausgehen, dass Korrelationen mit einem Korrelationskoeffizienten von 0,2 – 0,5 als gering einzustufen sind. (54) Da es sich in der vorliegenden (mit $n=11$) um eine derartige Stichprobe handelt, sind die mittleren Korrelationen von ERF_1 und ERF_2 daher unter Vorbehalt zu betrachten.

		Note PV-Score	ERF_1	ERF_2	ERF_3
Note PV-Score	ρ	1,00	-,43	-,34	,10
	Sig. (2-seitig)	.	,19	,30	,76
ERF_1	ρ	-,43	1,00	,92	,78
	Sig. (2-seitig)	,19	.	,00	,00
ERF_2	ρ	-0,34	,92	1,00	,78
	Sig. (2-seitig)	,30	,00	.	,00
ERF_3	ρ	,10	,78	,78	1,00
	Sig. (2-seitig)	,76	,00	,00	.

Tabelle 21: SPSS-Output der nichtparametrischen Spearman-Rangkorrelationen, Spearman-Korrelationskoeffizient ρ und p-Werte für die Noten des PV-Scores mit drei Ausprägungsformen der Erfahrung der Teilnehmer (ERF_1, ERF_2, ERF_3)

3.3.2 Konstruktvalidität

Die zentralen Tendenzen der Ergebnisse der beiden Evaluierungstools TSA/OSATS-, sowie PV-Score wurden überprüft.

Dazu fand mittels des Mann-Whitney-U-Tests eine Gegenüberstellung der Bewertungsgruppen TM mit „Note 1,00“ (\triangleq „bestanden“) bzw. TM mit „Note 2,00“ (\triangleq „nicht bestanden“) - bezogen auf die Bewertung durch den TSA/OSATS-Score - bezüglich deren Ergebnissen im PV-Score (\triangleq Note 1-6) statt.

Das Histogramm (Abbildung 19) zeigt die „absoluten Häufigkeiten“ aller erzielten „Ergebnisse im PV-Score“ (Noten 1-6), sowie das arithmetische Mittel M (\triangleq Notendurchschnitt des PV-Scores) je Bewertungsgruppe („Note 1,00“ oder „Note 2,00“, bezogen auf die TSA/OSATS-Bewertung).

Anhand des Histogramms ist zu erkennen, dass die absoluten Häufigkeiten der erzielten Noten des PV-Scores in der Bewertungsgruppe TM mit „Note 1,00“ sich im Bereich niedrigerer Notenwerte (\triangleq bessere Leistung, blaue Säulen) konzentrieren, während die absoluten Häufigkeiten der erzielten Noten des PV-Scores in der Bewertungsgruppe TM mit „Note 2,00“ eher im höheren Notenbereich zu finden sind (\triangleq schlechtere Leistung, orange Säulen).

Teilnehmer der Bewertungsgruppe „Note 1,00“ (die also durch TSA/OSATS als „bestanden“ gewertet werden) erhalten durch den PV-Score anders verteilte Noten (nämlich bessere Noten, s. blaue Säulen) als die „Note 2,00“-Teilnehmer (die also durch den TSA/OSATS als „nicht bestanden“ gewertet werden), deren PV-Noten hier dementsprechend schlechter ausfallen (s. orangene Säulen).

Dementsprechend beträgt der Mittelwert M der Noten des PV-Scores der Bewertungsgruppe „Note 1,00“ 2,17, während er bei der Bewertungsgruppe „Note 2,00“ mit 4,80 deutlich höher liegt. Der Vergleich des Mittelwertes der Noten des PV-Scores der Bewertungsgruppe „Note 1,00“ mit dem Mittelwert der Noten des PV-Scores der Bewertungsgruppe „Note 2,00“ zeigt ebenfalls, dass in der Gruppe „Note 1,00“ (durch den TSA/OSATS-Score) bessere PV-Score-Noten erzielt wurden ($M=2,17$) als in der Gruppe „Note 2,00“ ($M=4,80$).

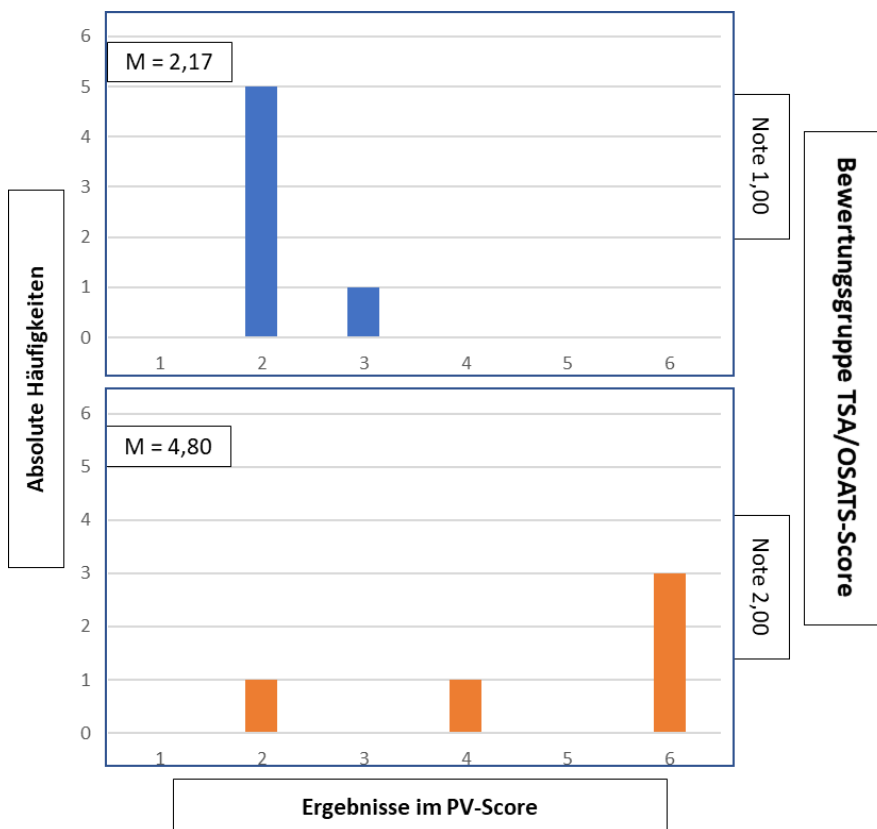


Abbildung 19: Histogramm der „Ergebnisse im PV-Score“ (Noten 1-6) je Bewertungsgruppe „Note 1,00, bzw. 2,00“ des TSA/OSATS-Scores: absolute Häufigkeiten und Mittelwerte der erzielten Noten des PV-Scores

Die Signifikanzmessung mittels Mann-Whitney-U-Tests für ordinal verteilte, unabhängige Stichproben zeigt Tabelle 22. Da die Daten viele gleiche Werte aufwiesen (sog. ties), wurde die asymptotische Signifikanz für die Beurteilung der zentralen Tendenzen herangezogen. Es ergab sich folgender Zusammenhang: die beiden Gruppen zeigen einen tendenziell signifikanten Unterschied ($p = 0,021$), d.h. dass die unterschiedliche Verteilung der Noten des PV-Scores bzgl. der Bewertungsgruppen TM mit „Note 1,00“, bzw. TM mit „Note 2,00“ des TSA/OSATS-Scores statistisch signifikant ist.

Ränge

	TSA/OSATS-Score*	N	Mittlerer Rang (MR)	Rangsumme (RS)
Note PV-Score	1,00	6	4,08	24,50
	2,00	5	8,30	41,50
	Gesamt	11		

Statistik für Mann-Whitney-U-Test*

	Note PV-Score
Mann-Whitney-U	3,500
Wilcoxon-W	24,500
Z	-2,315
Asymptotische Signifikanz (2-seitig)	,021
Exakte Signifikanz [2*(1-seitige Sig.)]	,030

Tabelle 22: SPSS-Output des Mann-Whitney-U-Tests, Verteilung der Ränge und Teststatistik, signifikanter Unterschied in den Noten des PV-Scores je Bewertungsgruppe, * Gruppenvariable: Bewertungsgruppen TM mit „Note 1,00“ bzw. TM mit „Note 2,00“ (bezogen auf die Bewertung durch den TSA/OSATS-Score)

Neben einer Überprüfung der zentralen Tendenzen der Ergebnisse der beiden Evaluierungstools mittels des Mann-Whitney-U-Tests wurde der Chi-Quadrat-Test angewendet. Die Kreuztabelle (Tabelle 23) zeigt einen Vergleich der Anzahl der „Noten des PV-Scores“ (als bivariates Outcome formuliert: 1,00 oder 2,00) mit der entsprechenden Anzahl der „Noten des TSA/OSATS-Scores“ (ebenfalls als bivariates Outcome formuliert: 1,00 oder 2,00). Der Chi-Quadrat-Test zeigt, dass der Chi-Quadrat-Wert hier 4,950 beträgt und die Freiheitsgrade (df) den Wert 1 haben.

Die Freiheitsgrade berechnen sich bei einer Kreuztabelle (mit n-Spalten und m-Zeilen) anhand der Anzahl der Kategorien der beiden Zufallsvariablen (Note_TSA/OSATS und Note_PV) folgendermaßen (53): $df = (n - 1) \times (m - 1)$

➔ Im vorliegenden Fall also: $df = (2-1) \times (2-1) = 1$

Die wichtigste Kennzahl in diesem Output ist der p-Wert. Dieser ist in der Spalte „Asymptotische Signifikanz (2-seitig)“ dargestellt ergibt einen statistisch signifikanten Zusammenhang ($p = 0,026$) zwischen den Verteilungen der Ergebnisse der beiden Evaluierungstools.

Mit diesen drei Kennwerten stellt sich das Ergebnis des Tests folgendermaßen dar (57): Der χ^2 -Test resultiert in einem Ergebnis von $\chi^2(1) = 4,950$, $p = 0,026$.

Abbildung 20 visualisiert die Zusammenhänge der Kreuztabelle (Tabelle 23): von den TM, die durch den PV-Score mit der Note 2 bewertet wurden, erhält keiner durch den TSA/OSATS-Score die Note 1 (keine blaue Säule). Von denen, die durch den PV-Score jedoch mit der Note 1 bewertet wurden, erfahren 6 TM (blaue Säule) auch durch die TSA/OSATS-Bewertung die Note 1.

Im „Idealfall“ der perfekten Korrelation zwischen den beiden Bewertungssystemen (TSA/OSATS- und PV-Score) wären von den TM mit „Note 1 durch den PV-Score“ dementsprechend null TM durch den TSA/OSATS-Score mit einer Note 2 bewertet worden (sprich keine orangene Säule in der Abbildung 20 bei „PV Note 1“), sondern alle TM ebenfalls mit einer Note 1. Das ist zwar nicht der Fall, dennoch ist der Zusammenhang zwischen den Bewertungssystemen statistisch signifikant ($p = 0,026$).

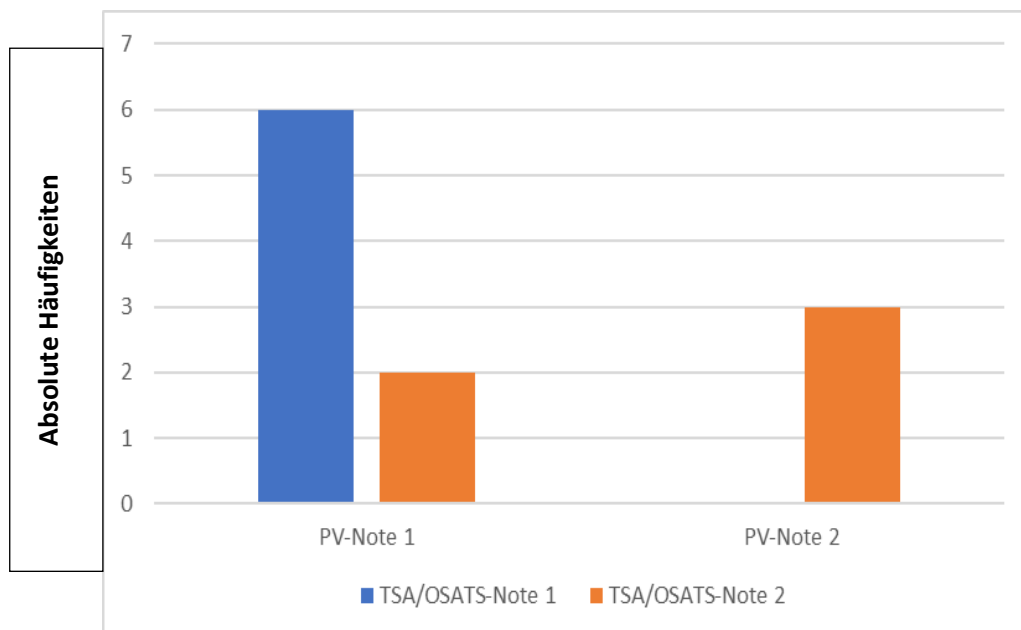


Abbildung 20: Histogramm der absoluten Häufigkeiten der Teilnehmer einer „TSA/OSATS-Note“ (bivariat, 1 oder 2) je „PV-Note“ (bivariat, 1 oder 2)

Kreuztabelle

Anzahl

		Note_PV		Gesamt
		1,00	2,00	
Note_TSA/OSATS	1,00	6	0	6
	2,00	2	3	5
Gesamt		8	3	11

Statistik für Chi-Quadrat-Test

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	4,950	1	,026		
Kontinuitätskorrektur	2,387	1	,122		
Likelihood-Quotient	6,161	1	,013		
Exakter Test nach Fisher				,061	,061
Zusammenhang linear-mit-linear	4,500	1	,034		
Anzahl der gültigen Fälle	11				

Tabelle 23: SPSS-Output des Chi-Quadrat-Tests, Kreuztabelle zu Note_TSA/OSATS mit Note_PV und Teststatistik

4 Diskussion

4.1 Teilnehmerzahl und Zeitraum

Das in der vorliegenden Arbeit für die Bearbeitung des BGDV-Fragebogens herangezogene Expertenkollektiv umfasste fünf Personen. Diese Anzahl von Experten entstammte der Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie der LMU. Eine Ausweitung resp. Präzisierung dieses limitierten Expertenkonsensus in zukünftiger Forschung könnte die angefangene Arbeit dieser Forschung fortführen. Indem die TA des BGDV-Fragebogens von Experten anderer Kliniken bzw. Institutionen bearbeitet würden, ließe sich aus der Auswertung der Angaben dieses vergrößerten Expertenkollektivs ein entsprechend angepasster und u.U. repräsentativerer Expertenkonsensus generieren, welcher dementsprechende Änderungen resp. Anpassungen der Formulierung der Items des PV-Scores nach sich ziehen würde.

Das Teilnehmerkollektiv der PV-Trainingssimulation dieser Arbeit umfasste elf Personen. Auch diese entstammten zwei Universitätskliniken der Stadt München. Ebenso könnte zukünftige Forschungsarbeit an dieser Stelle ansetzen, indem die PV-Trainingssimulation von einem vergrößerten Teilnehmerkollektiv anderer Kliniken, bzw. Institutionen vollzogen würde. Damit lägen umfassendere und möglicherweise heterogenere Datensätze vor, auf die der PV-Score angewendet werden könnte. Nur durch diese Ausweitung des Experten-/Teilnehmerkollektivs lässt sich hinreichend ermitteln, ob eine praktische Anwendung des PV-Scores (bzw. einer entsprechenden Abwandlung desselbigen) in der Zukunft praktikabel sein und die Umsetzung sich bewähren könnte. Der limitierte Forschungszeitraum der vorliegenden Arbeit ermöglichte lediglich die *einmalige* Durchführung *einer* PV eines Teilnehmers an der PV-Trainingssimulation. Zukünftige Forschungsansätze könnten Teilnehmer der PV-Trainingssimulation über einen längeren Zeitraum begleiten, bzw. mehrere Durchführungen von PV's eines Teilnehmers an der Trainingssimulation aufzeichnen. Dadurch ließe sich evaluieren, ob der PV-Score geeignet ist, Veränderungen der Leistung resp. Kompetenz eines Teilnehmers zu erfassen.

4.2 Interpretation der Ergebnisse

4.2.1 Interpretation der Ergebnisse des BGDV-Fragebogens

Generell sind Ungenauigkeiten im Rahmen des Datenakquise-, Auswertungs- und Interpretationsprozesses durch etwaige Messfehler der konzipierten Messinstrumente oder der Trainingssimulationssoftware nicht auszuschließen. Ebenso können unterschiedliche Urteilsanker der Experten, bzw. unterschiedliche Informationsbasis, sowie ein situationsspezifisch unterschiedliches Verhalten der Teilnehmer derartige Ungenauigkeiten bedingen.

Auch die Tatsache, dass der Fragebogenumfang beschränkt werden musste, um die Motivation der Experten zu erhalten, kann stellenweise zu Unklarheiten der Angaben geführt haben.

Während sich Messfehler der Messinstrumente anhand von zukünftiger Reliabilitätsüberprüfung (s. 1.6.2 und 4.3) in der Größenordnung abschätzen lassen können, gestaltet sich die Taxierung der anderen Einflussfaktoren tendenziell schwieriger. (54)

Auch an dieser Stelle sei zur Erinnerung nochmals ein Überblick über die Inhalte der vier Kategorien des BGDV-Fragebogens abgebildet (für einen ausführlicheren Überblick siehe 2.2.2, für eine genaue Formulierung der TA siehe 7.1).

1. Trokar-basiert: enthält TA zur Positionierung des Trokars
2. Zement-basiert: enthält TA zur Verteilung des eingebrachten Knochenzements
3. Bildgebungs-basiert: enthält TA zur Handhabung des C-Bogens
4. Anderes: enthält TA zum Zeitfenster für die Gesamtprozedur

4.2.1.1 quantitative Teilaufgaben

Die folgenden Anmerkungen beziehen sich auf das Ergebnis-Unterkapitel 3.1.1.

1a) Ranking der Kritizitätsangaben und Kritizitätskategorien:

Der Grund für die Einordnung der „Punktion eines anderen Wirbelkörpers“ in eine unkritischere Kategorie (d) als die Kritizitätsangabe (9,6) vorgibt, kann in einer unterschiedlichen Auffassung der Experten vom Verständnis des Punktionsausmaßes liegen. So könnte ein Experte bei der Beantwortung eine „minimale Punktion des falschen Wirbelkörpers“ im Kopf gehabt haben, die sofort korrigiert wurde. Dementsprechend würde dieser Experte die „Punktion eines anderen Wirbelkörpers“ in eine unkritischere Kategorie ordnen als ein Experte, der eine „weitreichende Punktion des falschen Wirbelkörpers“ mit späterer Korrektur visualisiert.

Eine konkretere und unmissverständlichere Formulierung der TA des BGDV-Fragebogens würde dazu beitragen, etwaige Differenzen im Verständnis vonseiten der beantwortenden Experten zu vermeiden. Auch eine präzisere Definition der Kritizitätskategorien wäre beispielsweise vorteilhaft

gewesen. Anstelle der vergleichsweise vagen Formulierung der Kategorien „a = Verletzung der Patientensicherheit, fügt dem Patienten Schaden zu“ und „b = Gefährdung der Patientensicherheit“ wäre ein genaues Definieren der klinischen Unterschiede zwischen den beiden Gruppen für ein konkreteres Anwenderverständnis hilfreich gewesen.

1b)/c) Kritik/Anregungen zur Formulierung der vier Kritizitätskategorien (a-d) und Angaben weiterer relevanter Risikostrukturen in Höhe LWK2:

Da die Kritizitätskategorien (a-d) von allen fünf Experten Zustimmung erfuhren, wurden sie beibehalten. Die von einem Experten zusätzlich genannten Risikostrukturen Leber, Milz und abdominelle Organe wurden nicht nachträglich in die Betrachtung einbezogen, da es als sehr unwahrscheinlich einzustufen ist, dass diese bei einer PV auf Höhe des LWK2 punktiert werden und der PV-Score in der vorliegenden Version vielmehr den Anspruch hat, einen Gesamtüberblick über die chirurgische Kompetenz eines Teilnehmers liefern zu können, als sämtliche Eventualitäten im Rahmen einer PV zu erfassen.

1d) Angaben zur Kritizität des Nadel-Abstands zu anatomischen Strukturen:

Der in dieser TA enthaltene Expertenkonsensus ist zu verifizieren. Es existieren in der Literatur keine Angaben zu kritischen Abständen zu den unter 3.1.1 in Tabelle 7 genannten Strukturen.

Es könnte diskutiert werden, dass die unter 1a) als kritisch eingestuften Risikostrukturen höhere Werte in den mm-Abständen der Kritizitätsbereiche erfahren müssten als die unkritischeren.

Die Nähe der Trokarspitze zur jeweiligen Struktur müsste demnach bereits ab einem größeren Abstand als kritisch zu bezeichnen sein, als bei unkritischeren Strukturen, bei denen auch ein geringerer Abstand der Trokarspitze noch als unkritisch zu bezeichnen sein könnte.

In diesen möglichen Zusammenhang müsste jedoch auch die Beachtung der anatomischen Größenunterschiede der Strukturen einfließen, welche ausschlaggebend dafür sein könnte, dass dieser mögliche Zusammenhang zwischen der Kritizität einer Struktur und dem kritischen Abstand der Trokarspitze von der Struktur sich nicht in einer solch linearen Form ausdrücken lässt wie die obige Erläuterung zugrunde legt.

Die Angaben eines Experten waren situationsbedingt unreflektiert und mussten daher ausgeklammert werden.

Die TA diene dazu, ein genaueres Verständnis über etwaige Kritizitätsabstufungen der Punktion im nahen Umfeld einer Risikostruktur zu generieren. Die Aufgabe hätte u.U. durch ansprechendere Gestaltung, z.B. Visualisierung der Risikostrukturen mit einzuzeichnenden Kritizitätsabständen,

mehr Akzeptanz vonseiten der Experten erfahren können.

1e) Rating von Punktionsparametern:

Aus dem Expertenkonsensus lässt sich generalisiert ableiten, dass die einfache Tatsache, *dass* eine Risikostruktur punktiert wurde, den größten Einfluss auf das Outcome hat. Die Tatsache, ob diese im Folgenden noch *mehrfach* punktiert wurde oder ob der Operateur mit dem Trokar die Struktur über *lange oder kurze Strecken hinweg* punktierte, scheint weniger wichtig zu sein. Diese Darstellung ist durch weiterführende Simulationstests zu verifizieren, bzw. widerlegen, indem beispielsweise die Auswirkungen einer Mehrfach- im Vergleich zu einer Einfachpunktion, sowie die Auswirkungen einer langstreckigen im Vergleich zu einer kurzstreckigen Punktion getestet werden.

1h) Einstufung der Abweichung (in mm) vom idealen Eintrittspunkt:

Themenspezifische Literaturrecherche zur Verifizierung des Expertenkonsensus über die Abweichung vom IEP bei der Durchführung einer PV war bis zum Zeitpunkt der Fertigstellung dieser Arbeit nicht zielführend. Allerdings konnten Ähnlichkeiten beim Vergleich der Angaben des Expertenkonsensus mit den Angaben der Klassifikation von Gertzbein und Robbins (G&R) festgestellt werden. Diese beziehen sich mit ihrer Klassifikation jedoch auf die Abweichung von Pedikelschrauben von ihrer Ideallage. Dennoch ist es möglich, dass den Experten die G&R-Klassifikation für Schraubenlagen geläufig war und die Ähnlichkeit der Werte des Expertenkonsensus daher rührt. Eine Verifizierung des Expertenkonsensus ist dadurch nicht möglich, die folgende Gegenüberstellung dient lediglich der Orientierung. (58)

Abweichung*	Bezeichnung durch Expertenkonsensus	Bezeichnung durch G&R-Klassifikation
0 mm	(ideal)	Grad A
0-2mm	ideal	Grad B
>2mm≤ 4mm	nicht ideal, aber akzeptabel	Grad C
≥ 4mm	nicht akzeptabel	Grad D
≥ 6mm	(nicht akzeptabel)	Grad E

Tabelle 24: vergleichende Gegenüberstellung der Angaben des Expertenkonsensus des PV-Scores zur „Abweichung des Trokars vom IEP“ mit der G&R-Klassifikation zur „Abweichung von Pedikelschraubenlagen“, *bezieht sich für Expertenkonsensus des BGDV-Fragebogens auf „Abweichung vom IEP bei einer PV“, bei der G&R-Klassifikation auf die „Abweichung von der lateralen Pedikelkortikalis“ ($\hat{=}$ Penetration) durch abweichende Schraubenlage

1i)/j)/k) Einstufung der Knochenansatzpunkte (Anzahl) des Trokars beim Platzierungsversuch/
Einstufung der erneuten Hautperforationen (Anzahl) durch Zurückziehen des Trokars beim
Platzierungsversuch/ Einstufung der Röntgenbilder (Anzahl) und Zeit (Min.) beim
Platzierungsversuch bis zur Endlage des Trokars:

Die Expertenkonsensus dieser TA bleiben zu verifizieren. Themenspezifische Literaturrecherche
lieferte zu diesen Aspekten keine weiterführenden Erkenntnisse, da die Beantwortung dieser TA
sehr stark von dem vorliegenden Fall abhing.

1m)/n) Einstufung des einzuhaltenden Abstandes (mm) der Trokarspitze zu einer der
Wirbelkörperwände:

Der Expertenkonsensus von „ ≥ 3 mm als minimal einzuhaltenden Abstand der Nadelspitze zu
irgendeiner Wand des Wirbelkörpers“ bleibt zu verifizieren. Themenspezifische Literaturrecherche
lieferte zu diesem Aspekt keine weiterführenden Erkenntnisse. Ebenso steht eine Überprüfung aus,
ob auch ein „zu großer Abstand der Nadelspitze von einer der Wände“ existiert und demnach mit
einem Strafpunkt im PV-Score zu bewerten wäre.

2a) Einstufung der Gesamtzementmenge (ml) für den behandelten LWK2:

Angaben zur idealen Zementmenge werden kontrovers betrachtet. Laut einer groß angelegten
prospektiven Studie mit 403 Patienten liegt keine signifikante Korrelation zwischen dem injizierten
Zementvolumen und dem klinischen Outcome, bzw. dem Ausmaß der Schmerzreduktion beim
Patienten vor. (59) (6) (60) Da patientenspezifische Größenunterschiede der Wirbelkörper durch
genaue ml-Angaben nicht erfasst werden können, können folgende Richtlinien im individuellen Fall
sinnvoller als die Angabe einer konkreten Zementmenge sein.

- Die zu injizierende Zementmenge eines Wirbelkörpers entspricht der Menge, die benötigt wird, um 50-70% des Residualvolumens des komprimierten Wirbelkörpers zu füllen.
- Die anterioren 2/3-3/4 des Wirbelkörpers sollten gleichmäßig gefüllt werden.
- Idealerweise durchdringt der Zement die vorderen 3/4 des Wirbelkörpers und mindestens 2/3 in der Transversalrichtung.
- Im AP-Strahlengang sollte der Zement die Mittellinie überschreiten, um beide Hälften des Wirbelkörpers zu stabilisieren.

Da diese Angaben jedoch aus den Simulatordaten schwierig auswertbar sind, wurde für die
Bewertung durch den PV-Score dennoch eine Bewertung durch den konkret formulierten
Expertenkonsensus der idealen Zementmenge (2,1-6,6 ml) vorgenommen.

Ein vergleichender Blick in die Literaturangaben zeigt durchschnittlich bei einer PV pro Wirbelkörper injizierte Zementmengen zwischen 2 und 15 ml und beträgt im Mittel etwa 7 ml. Dabei beziehen sich die höheren Werte tendenziell auf LWK's, während niedrigere Werte den Thorakalbereich betreffen. (61) (62) (63) Konkrete Postulierungen zur idealen Zementmenge werden aus den genannten Gründen tendenziell vermieden. (6) Andere Quellen werden konkreter und postulieren, dass meist unnötig zu viel Zement injiziert würde und dass 1,5–2ml vollkommen ausreichend seien. (59)

2b) generelle Einstufung der injizierten Gesamtzementmenge:

Der Expertenkonsensus bevorzugt eine zu niedrige Menge an injiziertem Zement im Vergleich zu einer zu hohen Menge. Für weitere Erläuterungen zu diesem Sachverhalt s. Unterkapitel 4.2.1.2, TA 2b).

2c) Ranking der Kritizitätsangaben und -abstufungen:

Laut Kritizitätsangabe der Segmentgefäße (4,0) müssten diese in die Kritizitätskategorie b zwischen prävertebralen Venen (3,2) und Bandscheiben (4,4) eingeordnet werden, statt in die Kritizitätskategorie a. In der Literatur sind Verläufe beschrieben, die durch Abschwemmen von Zement über die prävertebralen Venen und einer dadurch ausgelösten fulminanten Lungenembolie tödlich endeten. Da die (arteriellen) Segmentgefäße am thorakolumbalen Übergang gut kollateralisiert sind, kommen arterielle Embolien in das Rückenmark mit neurologischen Ausfällen kaum vor. Mit größerer Wahrscheinlichkeit ist eine revisionsbedürftige arterielle Blutung bei Punktion der Segmentgefäße zu erwarten.

- ➔ Diese Verläufe unterstützen die kritischere Einstufung der prävertebralen Venen (3,2) im Vergleich zu den beiden anderen genannten Komplikationen und sprechen für eine Einordnung der Segmentgefäße in die Kritizitätskategorie b. (9)

Zementleckage in den Spinalkanal tritt nur sehr selten auf und es gibt nur wenige Fallberichte. Dennoch stellt dies eine potenziell verheerende Komplikation dar, welche im Bereich des Rückenmarks oder der Spinalnerven zu Nervenreizungen, Lähmungen und selten auch zu kompletten Querschnittslähmungen führen kann. Die Einwirkungen des Zements reichen dabei von mechanischer Kompression, über thermische Schädigung während der exothermen Phase der Zementaushärtung, bis hin zu chemischer Irritation von Nervengewebe. (9) (59) (64) (65)

- ➔ Die hohe Kritizitätseinordnung der Experten von Subduralraum und intraspinalen Venen wird demnach von der Literatur gestützt.

Zement kann bei Eintreten in das extravertebrale Gefäßsystem potenziell lebensgefährliche Komplikation in Form von kardiopulmonalen Extravasationen nach sich ziehen. Ausgetretener Zement kann über den extraspinalen perivertebralen Venenplexus, die V. cava, sowie die Lungenarterien vaskuläre, kardiale und pulmonale Zementembolien auslösen.

➔ Die Einordnung der prävertebralen Venen in die zweitkritischste Gruppe b lässt sich damit belegen. Auch eine Einordnung in die kritischste Gruppe a wäre im Anbetracht der obigen Erläuterungen in Bezug auf die prävertebralen Venen diskutabel.

➔ Allerdings lassen sich Zementleckagen in verschiedene extravertebrale Kompartimente oft nicht vermeiden und bleiben meist klinisch inapparent. Die Einordnung der Experten des prävertebralen Raumes in Gruppe c könnte derartige Zementleckagen aufgreifen. (66)

Zementleckagen in die Bandscheiben bleiben im Normalfall unmittelbar symptomlos, können allerdings eine Degeneration der Bandscheibe vorantreiben, oder für Schmerzen durch die exotherme Hitzeentwicklung beim Aushärten des Zements sorgen. (6) Auch eine erhöhte Rate von Anschlussfrakturen durch Veränderungen in der Biomechanik des betroffenen Wirbelkörperabschnittes durch Einbringen des sehr steifen Zementes werden berichtet. (67) (68)

➔ Die Einordnung der Bandscheiben in Kritizitätskategorie c kann daher akzeptiert werden.

Der Expertenkonsensus spricht sich dafür aus, dass sich etwas an der Kritizität einer der genannten Strukturen ändert, je nachdem, ob sie von Trokarpunktion oder Zementleckage betroffen wird.

2d) Da keiner der Experten neben den genannten noch weitere bei einer Zementleckage relevante Risikostrukturen in Höhe des LWK2 nannte, wurden die genannten Risikostrukturen beibehalten.

Die Anmerkung eines Experten (TM2): „Die Einteilung a-d finde ich hier schwierig, da eine Leckage auch bei korrekter Pedikulierung auftritt und dann die Zementierung beendet werden muss.

Heißt, die Leckage ist oft nicht Fehler des Anwenders.“, wird von der Literatur gestützt. Häufig liegen bei berichteten Komplikationen durch Zementleckage keine technischen Fehler vonseiten des Operateurs vor. Zementleckage kann in solchen Fällen als unvermeidbare und unberechenbare Komplikation trotz perfekter Technik angesehen werden. (69)

Die Einteilung in die Kritizitätskategorien a-d wurde dennoch beibehalten, da die Bewertung durch den PV-Score keinen Unterschied macht, ob etwaige Komplikationen behandlerbedingt passieren oder durch andere Umstände bedingt sind. Zudem fand diese TA keinen direkten Einbau in den PV-Score.

2e) Einstufung der idealerweise zu injizierenden Zementmenge (ml) vor Anfertigung eines Röntgen-Kontrollbildes:

Literaturrecherche ergab eine angeratene zu injizierende Zementmenge von 0,1-0,2 ml vor Anfertigung eines Röntgen-Kontrollbildes und der darauf folgenden erneuten Zementinjektion. (63) Bei Einhaltung dieser Richtlinie würden allerdings bei einer Zementmenge von 3 ml bereits 15 bis 30 Röntgenkontrollen anfallen, was einen Konflikt mit der Strahlenhygiene darstellt.

Die Abweichung vom Expertenkonsensus, der ein Röntgen-Kontrollbild nach 0,98 ml vorschlägt, lässt sich durch zahlreiche Parameter der Zementinjektion erklären, die schlecht oder überhaupt nicht abschätzbar sind und somit generalisierte Richtlinien erschweren: darunter fallen Zementviskosität, Injektions- und intravertebraler Druck, aber auch Wirbelkörper-assoziierte Parameter wie die Ausdehnung der Frakturlinien. (66)

2f) Der Expertenkonsensus wertet die Anfertigung eines Röntgenkontrollbildes nach einer zu großen injizierten Zementmenge negativer als ein nach einer zu geringen Zementmenge angefertigtes.

Dies entspricht dem Expertenkonsensus der TA 2b), welcher eine zu große Menge an injiziertem Zement negativer bewertet als eine zu niedrige. Für weitere Erläuterungen zu diesem Sachverhalt s. Unterkapitel 4.2.1.2, TA 2b).

3b) Einstufung der während der Gesamtprozedur angefertigten Röntgenbilder (Anzahl):

Die Literaturrecherche ergab keine genauen Richtlinien zur Gesamtanzahl der während einer PV angefertigten Röntgenbilder. Vergleichende Angaben können durch den Report einer prospektiven Kohortenstudie des Suzhou Municipal Hospitals getroffen werden, bei der zwischen 2009 und 2012 143 Patienten mit osteoporotisch bedingten Wirbelkörperfrakturen durch eine PV behandelt wurden. Dabei wird von einer durchschnittlichen Anzahl von 16.00 ± 1.58 Fluoroskopiebildern pro PV durch einen einarmigen C-Bogen berichtet. (70)

Im Vergleich mit dem Expertenkonsensus liegt dieser mit „bis zu 25 Röntgenbilder pro Seite sind angemessen“ und einer dementsprechenden Angabe von „bis zu 50 Röntgenbilder sind in der Gesamtprozedur angemessen“ weitaus höher. Für weitere Ausführungen zur Anzahl der Röntgenbilder s. 4.2.2, Item 2A./2B.

3c) Einstufung der Menge angefertigter Röntgenbilder:

Die Tatsache, dass einer der Experten (20%) eine potenziell zu geringe Anzahl an Fluoroskopiebildern einer potenziell zu hohen Anzahl vorzog, lässt sich darauf zurückführen, dass sein Hauptaugenmerk dabei auf der erhöhten Strahlenbelastung anstelle der Prozedursicherheit lag.

Der Expertenkonsensus, der sich mit 80% für eine höhere Anzahl an Röntgenbildern ausspricht, fokussiert im Wesentlichen die Prozedursicherheit.

4a) Definieren eines geeigneten Zeitfensters (Min.) für die Durchführung einer PV:

Das von den Experten definierte Zeitfenster von mindestens 18,0 bis höchstens 41,0 Minuten für die optimal mögliche Durchführung einer PV impliziert, dass eine Dauer von unter 18,00 Minuten als zu kurz angesehen werden könnte. In einer derartigen Situation sei das Outcome der OP des jeweiligen Teilnehmers zu überprüfen. Ein erfolgreiches Resultat, das unterhalb der definierten Mindestzeit erreicht wurde, könnte geradezu noch positiver bewertet werden als ein ebenso erfolgreiches Resultat, das innerhalb des definierten Zeitfensters liegt. Ein negatives Resultat unterhalb der definierten Mindestzeit legt jedoch die Vermutung nahe, dass dies durch unpräzises, vorschnelles Handeln vonseiten des Operators hervorgerufen wurde. Bei Vergleichen zu literarischen Angaben optimaler Zeitfenster sei angemerkt, dass die Start- und Endpunkte bei unterschiedlichen Studien variieren können und dadurch unterschiedliche – und teils nicht vergleichbare - Angaben zustande kommen. (17) (70)

4.2.1.2 qualitative Teilaufgaben

Die folgenden Anmerkungen beziehen sich auf das Ergebnis-Unterkapitel 3.1.2.

1g) Definition des „idealen Eintrittspunktes“ (IEP) in den Pedikel:

- Da sich der Expertenkonsensus beim Auffinden des IEP nach dem Pedikel richtet, ist das Verfahren unsicher, wenn dieser nicht sicher identifiziert werden kann – z.B. aufgrund von Osteoporose, Fraktur und/oder Weichteilüberlagerung.
- Der Grund für die Orientierung leicht dezentral der Pedikelachse ist die etwas konvergierende Stichrichtung.
- Röntgen muss in zwei Ebenen möglich sein.

1l) Definition einer Ebene, anhand derer eine Bewertung der Tiefe der Trokarspitze in der Endlage möglich ist:

Der Expertenkonsensus, laut dem die Nadelspitze in der seitlichen Ansicht des Wirbelkörpers im ventralen Drittel des Wirbelkörpers liegen sollte, wird von der Literatur gestützt. (6)²⁸

2b) Begründung für die negativere Einstufung von einer zu hohen Menge an injiziertem Zement im Vergleich zu einer zu geringen Menge:

Der Expertenkonsensus, laut dem eine zu hohe Menge an injiziertem Zement aufgrund der zunehmenden Gefahr des Zementaustrittes und des Komplikationsrisikos negativer eingestuft wird als eine zu hohe Zementmenge, wird von der Literatur gestützt. Eine Studie von Ryu et al. zu den Hauptkomplikationen durch Zementleckage ergab, dass Zementleckage-bedingte Komplikationen durch ein hohes Injektionsvolumen verursacht wurden. Jin et al. betonen das gesteigerte Komplikationsrisiko durch massive Zementinjektionen. Andererseits wird auch festgestellt, dass die Stabilisierung des Wirbelkörpers bei einer PV nur durch ein ausreichendes Zementvolumen sichergestellt werden kann. (71)

2c) Nennung der Strukturen, deren Kritizität sich ändert, je nachdem, ob sie von einer Punktion oder von einer Zementleckage betroffen wurden:

- Zementleckage in den Spinalkanal kann durch die Hitzeentwicklung bei der Zementpolymerisation zu Thermoschäden an nervalen Strukturen führen. Außerdem sind eine Kompression oder Durchblutungsstörungen des Rückenmarks mitsamt (ggf. bleibender) Querschnittslähmung möglich. Blasen-/ Darm-Inkontinenz, Impotenz, Gefühlsstörungen oder chronische Schmerzen können weitere mögliche klinische Folgen darstellen. Eine Punktion des Spinalkanals kann zu Rückenmarksschädigung oder Nervenverletzungen führen. (4)
- Gefäße resp. Venen können sich nach Punktion durch den Trokar wieder schließen.
Zu den möglichen Folgen einer Zementleckage ins Gefäßsystem sowie in die Bandscheiben s. Anmerkungen zur TA 2c) unter 4.2.1.1.

➔ Insofern wird der Expertenkonsensus gestützt, laut dem es bei den Strukturen „Spinalkanal, Gefäße oder Bandscheiben“ einen Kritizitätsunterschied hinsichtlich

²⁸ s. 7.3, „Anhang 2“ für eine ergänzende „Definition der idealen Nadelposition“

der Tatsache gibt, ob sie von Zementleckage oder Trokarpunktion betroffen wurden.

2f)/3c) Es wird für ein Mehr an Röntgenkontrollbildern zugunsten der Prozedursicherheit votiert, welche damit über die erhöhte Strahlenbelastung gestellt wird.

3a) Ergänzende Richtlinien zur Positionierung des C-Bogens aus der Literatur stimmen erweiternd mit dem Expertenkonsensus überein:

- Orientierung am Pedikel: vor allem die medialen und inferioren Wände des Pedikels müssen klar dargestellt werden, um das Risiko der Verletzung von Spinalnervenzwurzeln oder des Rückenmarks zu minimieren.
- AP-Ebene: der C-Bogen sollte in den drei Raumachsen (nach kranial, kaudal und um die Körperlängs- und Querachse) verstellt bzw. verkippt werden, sodass die vordere Wirbelkörperwand direkt über die hintere Wirbelkörperwand projiziert wird und der Wirbelkörper direkt frontal getroffen wird. AP-Aufnahmen sollten angefertigt werden, um das Vorantreiben des Trokars abzubilden, bis die Ebene der Hinterwand des Wirbelkörpers erreicht wird.
- Sagittale Ebene: intermittierend sollten laterale Röntgenbilder angefertigt werden, um eine geeignete kranial-kaudale Angulation des Trokars sowie den Zugang zum frakturierten Wirbelkörper zu garantieren und ein Erreichen der Ebene der Hinterwand des Wirbelkörpers festzustellen. (6)²⁹

4.2.1.3 simulationsbasierte Teilaufgaben

Die folgenden Anmerkungen beziehen sich auf das Ergebnis-Unterkapitel 3.1.3.

1f) Der Expertenkonsensus des idealen, bzw. akzeptablen EP's ist durch zukünftige Datenakquise zu verifizieren. Falls zukünftige Forschungstendenzen zeigen sollten, dass ein derartig gewählter EP häufig zu Fehllagen des Trokars führt, ist der Expertenkonsensus zum IEP dementsprechend zu korrigieren.

Der Expertenkonsensus dieser TA kam durch eine simulationsbasierte Mittelung der von den Experten im Programm „Meshmixer“ gesetzten Trokar-Planungen zustande. Es bleibt zu verifizieren, ob es dabei zu simulationsbedingten Verzerrungen, bzw. Ungenauigkeiten kam.

²⁹ S. 7.3, „Anhang 1“ für eine ergänzende „Definition der idealen Einstellung des C-Bogens“.

Die Tatsache, dass laut Expertenkonsensus der IEP (grüne Markierung) rechts weitaus kleiner dargestellt wird als links, ließ sich durch Literaturrecherche nicht verifizieren. Auch die Tatsache, dass der „akzeptable EP“ des Expertenkonsensus (gelb-orangene Markierung) sich rechts bis in den Beginn des Facettengelenkbereichs erstreckt, bleibt zu verifizieren.

1I) Die Tabellen 8 und 9 (S. 50/51) zeigen Abbildungen der Expertenkonsensus der „idealen Nadellage“. Die Nadelspitze kommt in den sagittalen Projektionen der computersimulierten Einstellungen am Übergang vom medialen zum ventralen Wirbelkörperdrittel zu liegen. Damit entspricht sie weitestgehend dem Expertenkonsensus des qualitativen Anteils der TA 1I), welcher in der seitlichen Ansicht eine Lage der Nadelspitze im ventralen Drittel fordert.³⁰

3a) Da die Abbildungen der Tabelle 10 (S. 54) zur „passenden Einstellung des Bildwandlers in AP und sagittaler Projektion“ keinen Expertenkonsensus darstellen, sondern jeweils nur die Ansicht *eines* Experten, sind sie unter Vorbehalt zu betrachten und weiterhin zu verifizieren. Der aus dem qualitativen Anteil der TA 3a) abgeleitete Expertenkonsensus, der eine streng-seitliche Einstellung des C-Bogens fordert, sodass die Deck- und Bodenplatte des Wirbelkörpers orthograd getroffen werden, wird in den Abbildungen weitestgehend eingehalten; ebenso die geforderte mittige Positionierung des Wirbelkörpers in der AP Projektion.³¹

4.2.2 Interpretation der Ergebnisse des PV-Scores

Die Durchschnittswerte der Items **1B.** „benötigte Zeit für die Durchführung der Gesamt-OP“, **3B.** „minimaler eingehaltener Abstand zu einer der Wirbelkörperwände“, **3C.** „Gesamtmenge an appliziertem Zement“ und **4B.** „Anzahl der Ansatzpunkte der Nadel am Knochen beim PLV“ liegen innerhalb der entsprechenden Expertenkonsensus. Das Item **3A.** hat keinen entsprechenden Expertenkonsensus, bzw. Durchschnittswert. Die Durchschnittswerte der Items **1A., 2A., 2B. und 4A.** liegen außerhalb des Expertenkonsensus. Damit liegt die Hälfte der Durchschnittswerte der Items innerhalb der definierten Expertenkonsensus, die Hälfte liegt außerhalb. Angesichts des Notendurchschnitts der Teilnehmer von 3,36 erscheint diese Gewichtung plausibel und unterstreicht die insgesamt als „befriedigend“ einzustufende Leistung der Teilnehmer der PV-Trainingssimulation.

³⁰ S. 7.3, „Anhang 2“ für eine ergänzende „Definition der idealen Nadelposition“.

³¹ Zu ergänzenden Richtlinien zur Positionierung des C-Bogens sei auf den qualitativen Anteil der TA 3a), sowie 7.3, „Anhang 1“ verwiesen.

1A. Der Expertenkonsensus von „ ≤ 3 Minuten pro Pedikel beim Platzierungsversuch bis zur Endlage des Trokars“ wurde von keinem der Teilnehmer erreicht. Dementsprechend wurde bei diesem Item eine differenziertere Bepunktung angewendet (s. 2.3). In der vorliegenden PV-Trainingssimulation wurde ein mobiler C-Bogen eingesetzt, welcher zum Wechsel zwischen lateralem und AP-Strahlengang per Hand geschwenkt werden musste. Da im Normalfall bei der Durchführung von PV's am Klinikum für Allgemeine, Unfall- und Wiederherstellungschirurgie biplanare Bildgebung praktiziert wird, liegt die Vermutung nahe, dass die Experten beim Beantworten des BGDV-Fragebogens ein schnelleres Vorgehen visualisierten und dadurch der Konsensus von „ ≤ 3 Minuten“ zustande kam. Der erhöhte Zeitaufwand durch das Schwenken des Bildwandlers während der PV-Trainingssimulation stellt eine mögliche Erklärung für eine erhöhte Dauer beim Platzierungsversuch vonseiten der Teilnehmer dar.

1B. Gemäß der Anweisung des Items 1B. erhält ein Teilnehmer mit einer Gesamtpunktzahl von $\geq 10,8$ Punkten bei einer Gesamtdauer von < 18 Minuten nachträglich 2 Punkte im Sinne von „Strafpunkten“ auf sein Gesamtergebnis. (s. 3.1.4, Abbildung des PV-Scores)

TM 1 und TM 6 weisen mit 17,26, bzw. 9,68 Minuten beide eine niedrige Gesamtdauer auf. Da beide jedoch eine Gesamtpunktzahl von $< 10,8$ Punkten erzielen, fallen keine Strafpunkte für ein „zu schnelles Vorgehen“ an, da sie trotz der zügigen Vorgehensweise gute Ergebnisse (beide Note 2) erzielten.

Um die Dauer der Gesamtprozedur zu beschleunigen, könnte biplanare Fluoroskopie eingesetzt werden. Auch die Anwendung eines unilateralen anstelle des in der vorliegenden PV-Trainingssimulation praktizierten bilateralen Zuganges kann in geeigneten Fällen die Prozedurdauer verkürzen. (6) Allerdings muss dann mit einer dezentralen Zement-Verteilung gerechnet werden.

2A./2B. Die Expertenkonsensus von „ ≤ 11 Röntgenbilder pro Seite beim Platzierungsversuch bis zur Endlage des Trokars“, bzw. „ ≤ 50 Röntgenbilder in der Gesamtprozedur“ wurden von keinem der Teilnehmer eingehalten. Dementsprechend wurde bei diesen Items eine differenziertere Bepunktung angewendet. Ein Faktor, der zu einer erhöhten Anzahl an Röntgenbildern geführt haben kann, liegt in der Tatsache, dass es sich bei der Studie um eine simulierte Operationsumgebung handelte. Da strahlenfreie Bildgebung möglich war, liegt die Vermutung nahe, dass die Teilnehmer öfter das Fußpedal für die Auslösung eines Röntgenbildes betätigten, als es in einer realen OP der Fall gewesen wäre.

3A. Die Tatsache, dass drei der elf Teilnehmer (27,27%) (TM 4, 8 und 11) durch eine Punktion des Rückenmarks ein Killerkriterium des PV-Scores erfüllten und damit direkt in einem Ergebnis „Note 6 – nicht bestanden“ resultierten, ist auffällig. Wie die Bezeichnung „Killerkriterium“ bereits verlauten lässt, handelt es sich dabei um ein Szenario, das unter keinen Umständen geschehen sollte. Ein Blick auf die Tabelle 5 unter Punkt 2.4.4 zeigt, dass es sich bei den TM 4, 8 und 11 um diejenigen Teilnehmer handelt, die die wenigsten „Jahre seit Approbation“ aufzuweisen haben (1,4; 2,0 bzw. 0,0 Jahre). Auch die „Anzahlen bereits durchgeführter PV's“ sind mit einer, zwei, bzw. null sehr gering. Die geringe Berufserfahrung dieser drei Teilnehmer könnte demnach ausschlaggebender Faktor für ein Punktieren des Rückenmarks gewesen sein.

3B. Der vom Expertenkonsensus vorgegebene „Minimalabstand von 3,00 mm zu einer der Wände des Wirbelkörpers“ wurde lediglich von zwei Teilnehmern unterschritten (TM5: L 0,00 mm/ R 1,58 mm und TM 9: R 0,5 mm). Dieses Item stellt demnach für die Teilnehmer der vorliegenden Trainingssimulation eine vergleichsweise niedrige Problematik dar. Bei der Durchsicht der Daten der Teilnehmer ergeben sich teils hohe Abstände (z.B. TM 4: L 11,07 mm).

3C. Keiner der Teilnehmer lag bei der injizierten Zementmenge unterhalb des Expertenkonsensus von „ $\geq 2,1 \leq 6,6\text{ml}$ “, drei der Teilnehmer lagen darüber. Angesichts der unter „4.2.1 Interpretation der Ergebnisse des BGDV-Fragebogens“ - TA 2a) erläuterten Diskussionen um die „ideale Zementmenge“, den in der Literatur stark schwankenden Angaben, und der Tatsache, dass in der vorliegenden PV-Trainingssimulation eine etwaige Zementleckage nicht ausgewertet werden konnte, bleibt die Bewertung dieses Items unter Vorbehalt zu betrachten. Als ausschlaggebende Faktoren für das Vermeiden von Extravasaten sind neben einer nicht zu hohen Zementmenge eine geeignete Kanülenplatzierung, die Verwendung von Zement mit höherer Viskosität, sorgfältige Vorbereitung des Patienten, sowie fortlaufende röntgenologische Kontrolle im AP und lateralen Strahlengang zu nennen. (59)

4A. Dem Expertenkonsensus von „ $\geq 2,01 \leq 4,00\text{mm}$ Abweichung des Trokars vom IEP in den Pedikel“ entspricht lediglich der TM 2 beim rechten Pedikel (1,62mm Abweichung vom IEP). Angesichts der unter „4.2.1 Interpretation der Ergebnisse des BGDV-Fragebogens“ – TA 1f) erläuterten Ungenauigkeiten bei der Bestimmung des IEP, ist die Bewertung dieses Items unter Vorbehalt zu betrachten.

4B. Nur in vier Fällen (der insgesamt betrachteten 22 Fälle dieses Items: resultierend aus Betrachtung des linken und rechten Pedikels jedes Teilnehmers) liegt die Anzahl der „Ansatzpunkte der Nadel am Knochen beim Platzierungsversuch“ über den vom Expertenkonsensus festgelegten fünf Mal. Dieses Item stellt demnach – zumindest für die Teilnehmer der vorliegenden Trainingssimulation – eine vergleichsweise niedrige Problematik dar. Die Auswertung dieses Items fand durch ein Auszählen der Knochenkontakte der Trokarspitze im Programm „3D Slicer sequence files“ statt. Dieses Programm zeigte den zurückgelegten Weg und die Bewegungen des Trokars in einer simulierten Umgebung anhand des Trackingsystems der PV-Trainingssimulation. Dabei sei anzumerken, dass Ungenauigkeiten in der Auswertung durch sog. „Tracking jitter“ (=minimale, „flatternde/ zitternde“ Bewegungen der Trokarspitze) nicht auszuschließen sind.

Es lassen sich keine erkennbaren relevanten Zusammenhänge bei der Betrachtung der Teilnehmer, die die Instrumentierung links begonnen haben (TM 3, 4, 8 und 11) und anderen Parametern feststellen. Die Vermutung, dass ein Teilnehmer an dem Pedikel, mit dem er beginnt, schlechtere Werte in den Items erzielt, als beim zweiten Pedikel, da er bei diesem durch die „Übung“ am ersten Pedikel bereits „trainierter ist“, kann bei Durchschau der Daten nicht bestätigt werden.

Etwaige Differenzen zwischen den Leistungen der Teilnehmer am linken bzw. rechten Pedikel können ihren Ursprung in anatomisch individuellen Seitenunterschieden des Wirbelkörpers haben.

Da jedoch alle Teilnehmer denselben Wirbelkörper behandelt haben, spricht die Tatsache, dass einige der Teilnehmer in bestimmten Items am linken Pedikel bessere Resultate erzielen, andere am rechten, für eine Zufallserscheinung. Auch etwaige Zusammenhänge zwischen den Händigkeiten der Teilnehmer (rechts oder links) in Bezug auf bestimmte Parameter könnten sich erst an einem größeren Teilnehmerkollektiv repräsentativ untersuchen lassen.

Ein Vergleich der Durchschnittswerte (M) der jeweiligen Metriken, bzw. Items (siehe rechte Spalte der Tabellen unter 3.2), lässt erkennen, bei welchen Items die Teilnehmer höhere Punktzahlen ansammelten als bei anderen, d.h. bei welchen Items die Leistungen stärker vom Expertenkonsensus abwichen. Dennoch lässt sich „das schlechteste“ oder „das beste“ Item nicht direkt feststellen, da ein Vergleich durch die unterschiedlichen Höchstpunktzahlen der Items (bzw. Killerkriterien bei 3A.) nur eingeschränkt möglich ist. Dennoch lassen sich Tendenzen erkennen, welche im vorliegenden Kapitel unter dem jeweiligen Item thematisiert wurden.

4.2.3 Interpretation der Ergebnisse der Validierungsmethoden

Die Ergebnisse der nichtparametrischen Spearman-Rangkorrelationen lassen folgende Rückschlüsse zu:

- Die Erfahrung eines Teilnehmers in Form von „Jahren seit Approbation“ (ERF_1), bzw. „Anzahl bereits durchgeführter PV's“ (ERF_2) zeigt eine höhere Korrelation mit den Ergebnissen des PV-Scores als die „Jahre als selbstständiger, verantwortlicher Operateur von PV's“ (ERF_3). Die ERF_3 stellt demnach den am wenigsten geeigneten Parameter dar, um einen Zusammenhang zwischen der Berufserfahrung und dem Ergebnis eines Teilnehmers im PV-Score festzustellen.
- Da sich jedoch bei keiner der drei Ausprägungsformen der Berufserfahrung statistische Signifikanz ergibt und auch die Effektstärken (Spearman-Korrelationskoeffizienten ρ) lediglich schwache bis mittlere Zusammenhänge zeigen, sollte die Bedeutsamkeit dieser Ergebnisse nicht zu hoch eingestuft werden.
- Die Nullhypothese „Die praktische Anwendung des Scores auf die Simulationsdaten der Teilnehmer der PV-Trainingssimulation zeigt keinen signifikanten Zusammenhang zwischen dem Score-Ergebnis und der Berufserfahrung eines Teilnehmers.“ wird beibehalten.

Hingegen wird die Nullhypothese „Die praktische Anwendung des Scores auf die Simulationsdaten der Teilnehmer der PV-Trainingssimulation zeigt keinen signifikanten Zusammenhang zwischen dem Score-Ergebnis eines Teilnehmers und dem Ergebnis eines vergleichbaren, bereits etablierten Bewertungstools.“ aufgrund des Erreichens des Signifikanzniveaus im Mann-Whitney-U- und Chi-Quadrat-Test abgelehnt.

Die Tatsache, dass die drei Teilnehmer, die im PV-Score (mit der Note 6) als „nicht bestanden“ gewertet werden (TM 4, 8 und 11), auch im TSA/OSATS-Score nicht bestehen, spricht dafür, dass kein Teilnehmer durch den PV-Score falsch negativ bewertet wurde (z.B. ein Teilnehmer mit guter Leistung besteht nicht). Die Tatsache jedoch, dass zwei Teilnehmer laut TSA/OSATS-Score nicht bestehen (TM 2 und 5), laut PV-Score jedoch das Ergebnis „bestanden“ erfahren (mit den Noten 2, bzw. 4), spricht für eine falsch positive Bewertung dieser beiden Teilnehmer durch den PV-Score.³² Im Falle einer klinischen Anwendung des PV-Scores wäre diese Variante die unerwünschtere, da sie z.B. dazu führen könnte, dass ein noch nicht kompetenter Teilnehmer die Behandlungslizenz erhält.

³² Dies gilt unter der Annahme, dass die Ergebnisse des TSA/OSATS-Scores reliabel sind.

Es besteht die Möglichkeit, dass diese Fehleinstufung des PV-Scores durch Metriken zustande kam, die der PV-Score nicht erfasst. Da im Verlauf der Konzeption des PV-Scores aus genannten Gründen Metriken – wie beispielsweise der Bereich „Zementinjektion“ - der Kürzung anheimfielen (s. 2.3), lässt sich nicht ausschließen, dass nicht derart leistungskritische darunter waren, dass der PV-Score in seiner vorliegenden Version nicht bei jedem Teilnehmer in der Lage ist, eine angemessene Einstufung vorzunehmen.

Die Durchführung der Validierungsstrategien diene weniger der Postulierung endgültiger Darstellungen, als vielmehr dem Aufzeigen gewisser Tendenzen. Es geht im Wesentlichen darum, die Punktzahlen resp. Outcomes des PV-Scores als Anhaltspunkte zu betrachten, um mithilfe von Datenmaterial gewisse Tendenzen, die die Interpretation des Verfahrens vorschlägt (wie beispielsweise die Szenarien „bestanden vs. nicht bestanden“ eines Teilnehmers), zu unterstützen, bzw. zu widerlegen. (49)

Die Ergebnisse der Validierungsmethoden lassen die Feststellung zu, dass der PV-Score eine objektive Analyse der chirurgischen Kompetenz der Teilnehmer ermöglicht. Durch diese Operationalisierung der Leistung wird wiederum ein Vergleich mit anderen Leistungen ermöglicht. Die Ergebnisse der Validierungsmethoden sind im Sinne vorläufiger Validierung der vorliegenden pilotierenden Forschungsgegebenheiten zu verstehen.

Aussagen über die statistische Signifikanz und damit die Aussagekraft einer Studie sind entscheidend vom Stichprobenumfang abhängig. Etwaige bestehende signifikante Unterschiede resp. Zusammenhänge laufen Gefahr, bei zu kleinen Stichprobengrößen nicht erkannt zu werden. Aufgrund der geringen Fallzahl ($n=11$) der vorliegenden Arbeit konnte bei Anwendung des Spearman-Rangkorrelationstests das Signifikanzniveau nicht erreicht werden.

Da auch der Expertenkonsensus auf einer geringen Fallzahl ($n=5$) beruht und der PV-Score auf diesem basiert, sollte zukünftige Validierungsarbeit eine Vergrößerung des Stichprobenumfangs (sowohl der Experten, als auch der Teilnehmer), sowie klinische Bezüge miteinbeziehen. (72)

Durch zukünftige Forschung auf dem Gebiet der „Sample Size Analysis“ („Bestimmung des Stichprobenumfangs“) ließen sich Optimierungen im Bereich der „Statistical Power“ (Trennschärfe/Teststärke eines statistischen Tests) des PV-Scores und somit an seiner statistischen Aussagekraft erzielen.

Weitere Forschungsarbeit hinsichtlich zukünftiger Validierung könnte zum einen auf inhaltlicher Ebene ansetzen: die Items des PV-Scores könnten Experten vorgelegt werden. Deren Einordnung, ob es sich bei diesen Items um die ausschlaggebenden und leistungskritischen Aspekte einer PV handelt oder nicht, würde die Expertenvalidierung des PV-Scores darstellen.

Entsprechend dieser Expertenvalidierung ließe sich dann eine etwaige Modifikation der Anzahl und Formulierung der Items vornehmen.

Ein Vergleich des PV-Score-Ergebnisses eines Teilnehmers mit seiner Bewertung durch ein bereits etabliertes Messinstrument zu einem späteren Zeitpunkt könnte außerdem eine Überprüfung der prognostischen Kriteriumsvalidität darstellen. Bei dieser Variante der Validitätsbestimmung wird das Außenkriterium zu einem späteren Zeitpunkt als das zu validierende Messinstrument selbst bestimmt. Dadurch wird eruiert, inwieweit die Ergebnisse eines Messverfahrens mit später tatsächlich eingetretenen Ereignissen übereinstimmen. Die Validierung von Zulassungstests für bestimmte Studiengänge durch Vergleich mit dem späteren Berufserfolg stellt ein Beispiel der prognostischen Kriteriumsvalidität dar. (32)

Zum anderen ließe sich zukünftige Forschungsarbeit auch auf statistischer Ebene angehen. Durch epidemiologische Ansätze ließe sich prüfen, ob gewisse Items des PV-Scores prädiktiver für die Bewertung chirurgischer Kompetenz sind als andere. Dementsprechend ließe sich auf dieser Ebene eine etwaige Modifikation der Anzahl und Formulierung der Items mittels eines „statistischen Siebes“ vornehmen.

4.3 Limitationen und mögliche Modifikationen des PV-Scores der vorliegenden Arbeit

Die unter 2.5 beschriebene strukturierende, qualitative Inhaltsanalysemethode ermöglichte durch Systematisierung der qualitativen Expertenangaben eine Identifizierung des darin enthaltenen Expertenkonsensus. (14) (73) Dennoch sind Interpretationsungenauigkeiten durch charakteristische, blinde Flecken, die sowohl qualitative als auch quantitative Datenerhebungs- und Auswertungsverfahren bergen, nicht auszuschließen. (34 S. 21)

Die Teilnahmebereitschaft an der Bearbeitung des BGDV-Fragebogens ist als mittelmäßig einzustufen. Es wurden mehr chirurgische Experten kontaktiert, als sich schließlich zur Teilnahme bereit erklärten. Dies lässt sich auf die starke Einspannung im klinischen Alltag dieser Berufsgruppe und das dementsprechend geringe Maß an Zeit für außerberufliche Aktivitäten zurückführen.

Für eine zukünftige erhöhte Akzeptanz der Teilnahme an einer derartigen Untersuchung bzw. eine erhöhte Rücklaufquote könnte das Verteilen von persönlich formulierten Begleitschreiben oder das Anbringen von Informationsflyern über den Hintergrund der Studie, sowie eine Angabe über die voraussichtliche Dauer zur Beantwortung des Fragebogens in den Kliniken förderlich sein. Auch die Beilage eines Informationsgutscheines mit exklusiver Zusammenfassung der Ergebnisse oder eines Workshopgutscheines könnten wirkungsvolle Maßnahmen sein.

Die optische Gestaltung des Fragebogens ließe sich durch Einbau positiv wirkender Bestandteile, wie beispielsweise eine Interesse weckende Gestaltung der Titelseite, optimieren. (74)

Die Experten, die sich zu einer Teilnahme bereit erklärten, zeigten sich sehr kooperativ. Jedoch war die starke berufliche Einspannung sowohl bei der Suche nach einem geeigneten Termin zur Beantwortung des BGDV-Fragebogens, als auch bei der Durchführung des Fragebogens selbst deutlich wahrnehmbar. Dies äußerte sich teilweise in äußerst komprimiert formulierten schriftlichen Ausführungen der frei formulierbaren TA, sowie in tendenziell lakonischer Beantwortung auf verbale Nachfrage vonseiten der Mitarbeiter der Studie.

Diese Tatsache kann sich teilweise limitierend auf den Informationsgehalt dieser qualitativen Daten ausgewirkt haben, da durch die knappe Formulierung stellenweise Spielraum für Interpretation blieb. Dieser trotz Anwendung der strukturierenden Inhaltsanalysemethode bestehende Interpretationsspielraum kann zu Ungenauigkeiten in der Auswertung dieser unter 3.1.2 dargelegten weichen Daten geführt haben. Da die ausgewerteten weichen Daten des BGDV-Fragebogens aber keinen direkten Einbau in den PV-Score erfuhren, sondern neben erweitertem Erkenntnisgewinn der Untermauerung von Daten mit hartem Informationsgehalt dienten, hatten etwaige interpretatorische Ungenauigkeiten keine Auswirkungen auf den konzipierten PV-Score.

Im Hinblick auf die interpretationsfreie Vergleichbarkeit der Aussagen sind geschlossene Fragestellungen vorzuziehen.

Die bereits erwähnte geringe Anzahl an Experten, auf deren Konsensus der PV-Score beruht, könnte sich limitierend auf die Repräsentativität des PV-Scores auswirken. Um dem entgegenzuwirken, bzw. das etwaige Ausmaß dieser Limitation abschätzen zu können, könnte einerseits die Anzahl der Experten vergrößert werden, andererseits eine Prüfung mittels induktiver Statistik durchgeführt werden. Da diese darauf abzielt, von einer Zufallsstichprobe oder repräsentativen Auswahl auf die Grundgesamtheit zu schließen, ließe sich feststellen, ob es legitim ist, die Meinung der fünf Experten als allgemeingültigen Konsensus für die Durchführung von PV's anzusehen. (75)

Für die Konkretisierung des medizinischen Kompetenzbegriffes liegen seit den 1970er Jahren zahlreiche Definitionen vor. Carraccio et al. befasste sich in seiner Abhandlung „Shifting Paradigms: From Flexner to Competencies“ mit der Frage, was es bedeute, im medizinischen Sinne „kompetent“ zu sein und liefert sinngemäß folgende resümierende Definition: Kompetenz ist als ein komplexer Satz an Handlungsweisen zu verstehen, welcher auf Wissen, Geschick, Einstellung und persönlicher Fähigkeit beruht. (24) Angesichts der Vielschichtigkeit und Komplexität des Kompetenzbegriffes

steht außer Zweifel, dass einzelne Tests kaum eine umfassende Operationalisierung des Generalfaktors der Kompetenz darstellen können. Der PV-Score erfasst ausschließlich technische Facetten (TS) der chirurgischen Kompetenz. Die Tatsache, dass durch das Anwenden des PV-Scores ein hochdimensionaler Sachverhalt auf eine einzige Zahl reduziert wird, liegt in der Natur von Scores. Andererseits wird ebendiese Informationsreduktion ja gerade beabsichtigt und im Rahmen der Score-Konzipierung dafür gesorgt, dass die relevante Information überdauert. (28)

„Soft Skills“ des Chirurgen, welche ebenfalls einen wichtigen Teil seiner Gesamtkompetenz ausmachen, fanden in der vorliegenden Studie keine Berücksichtigung. Nicht-technische Fähigkeiten (NTS) wie Kommunikation, Stressbewältigung, Situationsbewusstsein, Aufgabenbewältigung, sowie Führungsverhalten sind nicht in der Bewertung durch den PV-Score beinhaltet. (18) Objektive Messinstrumente, welche auch die NTS in die Bewertung chirurgischer Leistung mit einbeziehen, existieren.³³

Zudem kann der PV-Score nicht das gesamte Therapieergebnis abbilden, da er nur das klinische Outcome, nicht aber das subjektive Patientenoutcome (inklusive postoperativer Patientensituation, Schmerzreduktion, Lebensqualitätsveränderung, usw.) berücksichtigt. Andere bereits etablierte Wirbelsäulen-Scores³⁴ basieren auf dem subjektiven (Schmerz-) Empfinden des Patienten.

Eine mögliche zukünftige Erweiterung des PV-Scores um die nicht erfassten Kompetenzfacetten durch Erkenntnisse benachbarter Forschung³⁵, bzw. durch eine Integration bereits etablierter Scoringssysteme könnte zu einer umfassenderen Bewertung des Kompetenzbegriffes durch den PV-Score führen.

Die Tatsache, dass in der vorliegenden Studie ein 3D-gedruckter Wirbelkörper behandelt wurde, könnte die Supposition nahelegen, dass die Praxistauglichkeit des PV-Scores zu bestreiten sei, da anatomische Unterschiede resp. Besonderheiten einer realen Wirbelsäule nicht erfasst wurden. Dem sei entgegengebracht, dass die Studie das Ziel verfolgte, sich mit den Abläufen einer PV auseinanderzusetzen und die Prozeduren zu trainieren.

Vergleichbar ist dies mit dem in der Luftfahrt etablierten Flugsimulatortraining, welches ebenfalls darauf abzielt, Abläufe zu trainieren, damit diese dem Piloten unter realen Bedingungen keine

³³ z.B. Non-Technical Skills for Surgeons, Oxford Non-Technical Skills Scoring System II, Observational Teamwork Assessment for Surgery (OTAS), Anesthetists' Non-Technical Skills (17 S. 1361)

³⁴ z.B. VAS-Score: Visual Analog Scale, Oswestry-Low-Back-Pain-Disability Questionnaire, Hannover Wirbelsäulen-Score (84 S. 112)

³⁵ s. z.B. (37)

Probleme mehr bereiten - auch wenn die reale Situation sich anders präsentiert als im SBT. Um größere anatomische Variabilität und damit mehr Realitätsnähe in die PV-Trainingssimulation und nachfolgend die Ergebnisse des PV-Scores zu integrieren, könnte man die Prozedur in zukünftiger Forschungsarbeit an unterschiedlichen 3D-gedruckten Wirbelkörpern verschiedener Patientenfälle und Schwierigkeitsgrade trainieren lassen. Damit ließe sich auch die Verallgemeinerbarkeit des PV-Scores, sowie des zugrunde liegenden Expertenkonsensus verifizieren.

Limitierend auf die Aussagekraft der Ergebnisse des PV-Scores dieser Studie könnte sich die Tatsache auswirken, dass jeder Teilnehmer nach nur *einer* Durchführung an der PV-Trainingssimulation bewertet wurde. Es kann somit nicht ausgeschlossen werden, dass ein Teilnehmer bei diesem *einen* Versuch eine schlechtere chirurgische Performance lieferte, als er eigentlich im Stande war. Dies könnte der Tatsache geschuldet sein, dass ein Teilnehmer sich erst mit den Gegebenheiten der PV-Trainingssimulation vertraut machen musste und demnach möglicherweise im ersten Durchlauf am Simulator noch nicht das gleiche chirurgische Kompetenzlevel liefern konnte, wie er es unter Bedingungen in vivo könnte. Um diese möglichen Schwankungen in der Performance der Teilnehmer auszugleichen, könnte es von Vorteil sein, jeden Teilnehmer eine bestimmte Anzahl *n* an Durchläufen an der PV-Trainingssimulation mit diversen Wirbelsäulen-Modellen ausführen zu lassen und den PV-Score dann auf die Mittelwerte der Ergebnisse der *n*-Durchläufe zu beziehen. Die *n*-Durchläufe könnten an verschiedenen Tagen durchgeführt werden, sodass „tagesformabhängige Schwankungen“ der Chirurgen ausgeglichen würden. Dies könnte eine umfassendere und aussagekräftigere Bewertung der Gesamtkompetenz eines Chirurgen ermöglichen.³⁶

Idealtypischerweise kommt es in einem Standard-Setting-Verfahren zur Konsultation eines Expertenpanels, welches in einem iterativen - aus Einzelurteilen und Gruppendiskussionen bestehenden - Prozess, zur Festlegung von Cut-Scores kommt. Je nach angewandter Methode wird das Expertenpanel genau über empirische Schwierigkeiten des Prozederes und die Folgen informiert, die seine Cut-Score-Setzung für die dadurch entstehenden Kompetenzstufen und die praktische Anwendung des Messinstrumentes haben können. (47)

³⁶ Dieses Vorgehen praktiziert z.B. das „Royal College of Obstetricians & Gynaecologists“. Hier absolvieren Teilnehmer Trainingsprogramme und erfahren Bewertungen durch Messverfahren wie den „summativen OSATS“, wobei gilt: „Trainees require a minimum of three procedures to be deemed competent (...)“ (81)

Limitierend auf die Validität der Cut-Score-Setzung des PV-Scores wirkt sich die Tatsache aus, dass anstelle eines derartigen Expertenpanels eine Festlegung der Cut-Scores auf Basis einer konstruierten Höchstpunktzahl erfolgte (vgl. 2.3). Durch die Tatsache, dass die Gewichtung der Items des PV-Scores stellenweise an die klinischen Outcomes der Teilnehmer der PV-Trainingssimulation angepasst wurde, sind in einigen Items theoretisch sehr hohe Höchstpunktzahlen möglich (z.B. in Item 3A.: bei Mehrfachpunktion von Risikostrukturen), während in anderen durch eine sehr differenzierte Bewertung in Form von Dezimalzahlen trotz großer Abweichung vom Expertenkonsensus „nur“ eine geringe Strafpunktzahl anfällt (z.B. in Item 2A.: trotz eines doppelt so hohen Outcomes wie das vom Expertenkonsensus vorgegebene „nur“ eine Strafpunktzahl von 0,2 Punkten). Diese Festlegung der strukturellen Eigenschaften eines Scores liegt meist im Entscheidungsbereich des Autors, welcher neben statistischen Argumenten meist hauptsächlich der Praktikabilität Bedeutung beimisst. Ein Teil der Entscheidung ist dabei oftmals auch Willkür. (28) Dieser Faktor kann sich limitierend auf die Formulierung des PV-Scores ausgewirkt haben.

Bei der Konzeption der beiden wissenschaftlichen Messinstrumente des vorliegenden Pilotprojekts wurde nach bestem Wissen und Gewissen daran gearbeitet, den geforderten Gütekriterien Rechnung zu tragen. In den entsprechenden Textstellen wurden die explizit überprüften Gütekriterien erläutert. An anderen Stellen, an denen eine explizite Überprüfung eines Gütekriteriums noch aussteht, können mögliche Verfahren zur Güteüberprüfung ansetzen. Nicht alle Gütekriterien wurden im Rahmen des explorativen Vorgehens des vorliegenden Pilotprojekts in kompletter Ausführung überprüft, da die vorliegende Version des PV-Scores im Sinne eines Blueprints zu verstehen ist, an dessen Bestandteilen modifizierend weitergearbeitet werden kann. Als solches wurde die ausführliche Überprüfung gewisser Gütekriterien nicht in den Vordergrund der vorliegenden Arbeit gerückt. An dieser Stelle werden noch einige abschließende Anmerkungen zu ausgewählten Aspekten der Gütekriterien getroffen.

Die Gewährleistung der Auswertungsobjektivität (s. 1.6.1) bei Fragen mit offenem Antwortformat ist häufig deutlich schwieriger sicherzustellen als die geschlossener Antwortformate.

Eine quantitative Bestimmung der Auswertungsobjektivität lässt sich erreichen, indem ein Fragebogen mindestens zwei verschiedenen Auswertern zur Analyse vorgelegt wird. Diese sollen unabhängig voneinander eine Codierung bzw. Interpretation der einzelnen Aufgaben - insbesondere der frei formulierbaren Antworten - vornehmen. Die mittlere Korrelation zwischen den Auswertern

lässt sich dann als Maß der Auswertungsobjektivität interpretieren.³⁷ (32) Da der Einsatz offener Antwortformate im vorliegenden BGDV-Fragebogen nicht dem Generieren quantitativen Datengehaltes diene, sondern erweiternder Hintergrundinformation der harten Daten, wurde im Unterkapitel „2.5 Datenspeicherung und Statistik“ auf eine quantitative Überprüfung der Auswertungsobjektivität verzichtet. (31)

Eine zukünftige sinnvolle Überprüfung der Reliabilität (s. 1.6.2) des PV-Scores nach etwaigen Modifikationen in der Anzahl bzw. Formulierung seiner Items würde etwa die Durchführung einer Konsistenzanalyse darstellen. Aus den verschiedenen Formeln zur Bestimmung der internen Konsistenz würde sich für den vorliegenden Fall der Alpha-Koeffizient nach Cronbach (1951)³⁸ eignen und Auskunft über die Homogenität des Verfahrens geben. Die Korrelation aller Items des Scores miteinander würde sozusagen als „Durchschnittsreliabilität“ des Verfahrens Auskunft über die Konsistenz, also die Homogenität des Scores liefern. (32)

Die vorliegende Arbeit erfüllt demnach en passant auch den Zweck, den Blick für die Güte von wissenschaftlichen Messinstrumenten zu schärfen und Möglichkeiten zu deren empirischer Bestimmung zu vermitteln.

Als zusammenfassende Erkenntnis lässt sich statuieren, dass eine flexible Gestaltung der Forschungsinhalte eines Pilotprojekts zwar ein anpassungsfähiges Vorgehen ermöglicht, das vorzeitige Festlegen eindeutiger Klassifikationsanweisungen zur Dateneingabe und –transformation jedoch ein simultanes oder zukünftiges Überprüfen bestimmter Gütekriterien erleichtert.

Somit beziehen sich etwaige Defizite im Bereich der Gütekriterien auf die Version des PV-Scores der vorliegenden Arbeit, sind aber zugleich als Anknüpfungspunkte zukünftiger Forschung zu verstehen. Letztlich wird eine rationale Entscheidung für oder gegen die routinemäßige Verwendung eines Scores immer die Berücksichtigung der Gütekriterien einschließen.

³⁷ Diese Korrelation sollte erheblich über $r = 0.5$ liegen, um noch von einer akzeptablen Übereinstimmung sprechen zu können. (31)

³⁸ Seine Berechnung wird im Statistikprogramm SPSS als Standardmethode zur Reliabilitätsbestimmung angeboten.

4.4 Schlussfolgerungen und Zukunftsausblicke

Wie bereits in der Einleitung erwähnt, ergaben sich durch die flexible Gestaltung der Rahmenbedingungen der Forschungsfrage Möglichkeiten, auf die im Laufe der Forschung auftretenden neuen Gegebenheiten einzugehen. Im Verlauf der explorativen Herangehensweise des Pilotprojekts wurden TA des BGDV-Fragebogens, die ursprünglich für den direkten Einbau in den PV-Score geplant waren, im Laufe der fortschreitenden Forschungsarbeit zum Teil anderweitigen Zwecken - wie erweitertem Informationsgewinn für zukünftige Forschung - zugeschrieben. Derartige Adaptionen stellen - hinsichtlich der Konzipierung der vorliegenden Studie als Pilotprojekt - die Änderungen dar, die es einem zukünftigen groß angelegten Projekt zur selbigen Thematik ermöglichen, an dem momentanen Stand der vorliegenden Arbeit anzuknüpfen.

Zukünftige normative Forschungsarbeit könnte an der Stelle der Cut-Score-Setzung ansetzen (vgl. 2.3). Ein Expertenpanel könnte neben einer Neueinstufung der Punkteverteilung auch eruieren, ob die Cut-Scores der binären Klassifikation in „bestanden vs. nicht bestanden“ einer Neutaxierung bedürfen. In Kombination mit klinischen Tests ließe sich ermitteln, ob ein „strengerer“ Cut-Score (beispielsweise „nicht bestanden“ ab Note 3) zweckmäßigere Konsequenzen nach sich ziehen würde. Ein Ansatz, den pass-fail-Cut empirisch zu überprüfen, läge im Durchführen einer Standard Setting Studie, bei der spezielle Tests für die Ermittlung eines Cut-Scores für ein Messinstrument durchgeführt werden. (76) Gerade im Hinblick auf etwaige falsch positive (z.B. „bestanden“ für einen nicht kompetenten Anfänger) und falsch negative (z.B. „nicht bestanden“ für einen kompetenten Experten) Bewertungen könnten Untersuchungen angeführt werden (s. 4.2.3). (12)

Der in der vorliegenden Arbeit durch eine PV behandelte LWK2 wies mit seiner Deckplatten-Impressionsfraktur eine epidemiologisch gesehen besonders häufige Lokalisation auf. (s. 1.2 und 2.1) Er entstammte zudem einer typischen Risikopatientin für Wirbelkörperfrakturen: weibliches Geschlecht, höheres Alter, Adipositas, Osteoporose. Die Prävalenz für Osteoporose-assoziierte Wirbelkörperfrakturen liegt bei rund 250.000/Jahr in Deutschland. (5) Da der durch den BGDV-Fragebogen erhobene Expertenkonsensus und dementsprechend der darauf basierende PV-Score sich auf diesen Fall beziehen, ist davon auszugehen, dass die in dieser Arbeit dargelegten Richtlinien sich auf eine große Anzahl der klinischen PV-Fälle übertragen lassen.

Für die Bewertung der Durchführung einer PV an Wirbelkörpern anderer Lokalisation müssten Bestandteile des PV-Scores dementsprechend angepasst werden (z.B. etwaige Änderungen der Risikostrukturen und der Zementmenge bei Wirbelkörpern der Thorakalregion).

Die ständige Verbesserung auf dem Gebiet der PV könnte einer großen Anzahl an Patienten zu einer verbesserten Lebensqualität verhelfen. Studien zeigten erhebliche Schmerzlinderung und Verbesserung der körperlichen Beweglichkeit nach der Behandlung durch PV in 90% der Fälle bei von Osteoporose betroffenen Patienten, in 70% der Fälle bei tumorösen Geschehen und in 80% der Fälle bei Patienten mit schmerzhaften Hämangiomen. (6)

Der PV-Score in seiner gegenwärtigen Version stellt eine Art „gekürztes Ad-Hoc-Messinstrument“ dar. Einige seiner ursprünglich generierten Items, die ihre Validierung durch den Expertenkonsensus des BGDV-Fragebogens bereits erfahren hatten, fanden dennoch keinen Eingang in den PV-Score (s. 2.3). Da die zur Bewertung dieser Items nötigen Daten der PV-Trainingssimulation bis zur Fertigstellung der vorliegenden Arbeit nicht auswertbar waren, fielen die entsprechenden Items der Kürzung anheim. Sie sind jedoch Teil des ursprünglich konzipierten Messinstrumentes „PV-Score“ und demnach unerlässlich für eine umfassende Bewertung des Kompetenzbegriffes.

So konnte beispielsweise das Gebiet der „durch Zementleckage betroffenen Risikostrukturen“ (vgl. TA 2c)) kein Item in der Version des PV-Scores der vorliegenden Arbeit finden, da dieser Parameter der Analyse der PV-Trainingssimulation nicht zugänglich war. Zukünftige Forschungsarbeit sollte eine Integration der gekürzten Items in den PV-Score vornehmen und zukünftige Teilnehmer somit anhand einer umfassender formulierten Version des PV-Scores bewerten. Am Beispiel der Integration des Items „durch Zementleckage betroffene Risikostrukturen“ könnten die Killerkriterien um den Parameter „Zementleckage in das Rückenmark“ erweitert werden, da die polymerisatorische Hitzeentwicklung des Zements dort irreversible Schäden hervorrufen würde. Zementleckage, bzw. -fehleinspritzung um die Spinalnervenwurzeln bzw. in das Foramen intervertebrale oder in den Spinalkanal könnte in Anlehnung an den Expertenkonsensus der TA 2c) (vgl. 4.2.1.1) mit der schlechtesten Note vor dem Killerkriterium resp. einer hohen Punktzahl bewertet werden.

Der PV-Score bietet die Möglichkeit der Erweiterung, bzw. Anpassung für verwandte Operationen. Diese Erweiterungsmöglichkeit war bereits in einer frühen Phase der Forschungsarbeit in Erwägung gezogen worden, siehe die einleitenden Worte zum BGDV-Fragebogen (7.1): „Ziel ist es, dass der Score für Vertebroplastien und (wenn möglich) allgemein für Pedikulierungen eine objektive Bewertungsmöglichkeit darstellt.“ Eine anpassende Umformulierung für die Bewertung chirurgischer Kompetenz bei der Durchführung einer Kyphoplastie, Pedikelschraubeninsertion oder anderer wirbelsäulenchirurgischer Eingriffe ist beispielsweise denkbar.

Das Grundkonstrukt des PV-Scores könnte im Sinne eines „für alle Wirbelsäuleneingriffe anwendbaren Basisteils“ übernommen werden. Übereinstimmende Elemente der Operationen, wie z.B. die Risikostrukturen bei der Punktion des Pedikels, ließen sich somit unverändert aus dem PV-Score übernehmen. Zusätzlich müssten diesen Bestandteilen dann diejenigen Anteile beigelegt werden, die für die jeweilige OP charakteristisch sind. Somit ließe sich der PV-Score beispielsweise durch Integration zusätzlicher Bewertungskriterien zur Schraubeninsertion zum „Pedikulierungs-Score“ abwandeln.

Diese möglichen Abwandlungen des PV-Scores könnten ebenfalls zukünftige Validierungsmöglichkeiten darstellen, mittels derer sich überprüfen ließe, ob das in dieser Arbeit beschriebene Grundvorgehen - auf eine andere Situation resp. OP bezogen und dementsprechend angepasst – zu repräsentativen und praktikablen Ergebnissen führen kann.

Eine andere Erweiterungsmöglichkeit des PV-Scores liegt in der Formulierung verschiedener Varianten des Scores für verschiedene klinische Anforderungen. Durch unterschiedliche Gewichtung der vier Kategorien „1. Zeit, 2. Effizienz, 3. Sicherheit, 4. Geschicklichkeit“³⁹ könnte der Score individuell an situations- bzw. anwenderabhängige Anforderungen angepasst werden. Der in der vorliegenden Arbeit konstruierte PV-Score ist als „All-inclusive-Score“-Variante zu verstehen, in der alle vier Kategorien (und damit die in ihnen enthaltenen Items) die gleiche Gewichtung erfahren und kann im Sinne eines Gesamtüberblickes über die chirurgische Kompetenz eines Behandlers angewendet werden, wenn keine speziellen Anforderungen des Anwenders vorliegen. Beispielsweise ließe sich die Variante „Zeit-/Effizienz-Score“ konstruieren, indem die Kategorien „Zeit“ und „Effizienz“ (und damit die in ihnen enthaltenen Items) eine höhere Gewichtung erfahren als die übrigen Kategorien. Dies könnte in einer Klinik von Bedeutung sein, in der Behandlermangel und Patientenüberschuss herrscht, die vorhandenen Behandler in den Bereichen „Sicherheit/Genauigkeit bei der Durchführung einer PV“ bereits zur vollsten Zufriedenheit ausgebildet sind und somit im Hauptfokus das „Optimieren der Effizienz/Beschleunigung der Durchführung einer PV“ im Vordergrund steht. Auch bei einer weniger spezifischen Ausgangssituation könnten die verschiedenen Score-Varianten dazu dienen, individueller auf die jeweiligen Stärken oder Schwachpunkte in der Leistung eines Chirurgen eingehen zu können und so bei der Weiterbildung gezieltere Ansätze verfolgen zu können.

³⁹ Die Kategorie „5.Outcome“ wird in dieser Ausführung nicht berücksichtigt, da sie bei Erfüllen eines Killerkriteriums direkt zum Nichtbestehen der PV-Trainingssimulation führt.

Im Rahmen dieser Arbeit wurde der PV-Score nur auf die Ergebnisse der Teilnehmer der PV-Trainingssimulation angewendet. Es ist in Erwägung zu ziehen, dass die Teilnehmer bei der Durchführung der PV nicht ihr größtes Potenzial ausgeschöpft haben, da es sich „nur“ um eine Trainingssimulation handelte. Zukünftige Forschungsarbeit könnte zum Ziel haben, die Teilnehmer der Studie bei der Durchführung einer in vivo PV am realen Patienten zu bewerten.

Dies könnte eine erneute Schnittstelle für die Forschungskombination aus Medizin und Informatik darstellen, da durch Vergleich der Ergebnisse des PV-Scores einer realen OP mit den Verhältnissen in einer Trainingssimulation Rückschlüsse auf eventuelle Anregungen zur Gestaltung der Trainingssimulation gezogen werden könnten. Forschungsfortschritte auf dem Gebiet der Gestaltung realitätsnaher Trainingssimulationen in Kombination mit der Gestaltung geeigneter Bewertungstools chirurgischer Kompetenz können dazu beitragen, den Einfluss des SBT im Vergleich zu konventionellen Trainingsmethoden bedeutend voranzutreiben. (s. 1.3)

Durch routinemäßigen Einsatz von SBT ließe sich der praktische Aspekt zu einem früheren Zeitpunkt in die chirurgische Ausbildung integrieren, als dies unter konventionellen Trainingsmethoden möglich ist. So könnte das Verständnis des Chirurgen in der Ausbildung für die Abläufe einer PV verbessert und seine praktischen Fähigkeiten frühzeitig geschult werden. Zum Teil ließen sich dadurch konventionelle Trainingsmethoden, wie Kadaverstudien, ersetzen. (77) (13) (12)

Die Forschungskombination aus Medizin und Informatik könnte auch an der unter 2.3 angeführten „Ausgestaltung des Trainingsprogrammes“, welches die Konsequenz der Noten 5 und 6 des PV-Scores darstellt, ansetzen. Das Trainingsprogramm ließe sich durch eine Verknüpfung aus klinischen (z.B. konventionelles Assistieren bei PV eines Experten) mit simulationsbasierten Aspekten (z.B. Absolvieren eigens konzipierter Übungsszenarien) realisieren. Übung der präoperativen Operationsplanung, der Bildwandlereinstellung, geeigneter Kanülenplatzierung, Beschränkung der Zementmenge und Druckreduktion könnten Möglichkeiten der Risikoabschwächung darstellen. (59) Untersuchungen, ob das Ergebnis des PV-Scores nach Absolvieren des Trainingsprogrammes eine Besserung erfährt, bieten weitere Forschungsansatzpunkte.

Bei positivem Ausfallen der ausstehenden Forschungsinhalte (z.B. Etablierung durch weitere Validierungsprüfungen) könnte der PV-Score als high-stake Assessment angewendet werden.

High-stake Assessments ziehen wichtige Konsequenzen für den Teilnehmer eines Tests (im vorliegenden Fall: Teilnehmer der PV-Trainingssimulation) nach sich. (23) (78)

So könnte der PV-Score ein Medium institutioneller Art darstellen, mittels dessen einem noch in der Ausbildung befindlichen Chirurgen bei „positivem Ergebnis“ die Lizenz zum eigenständigen Durchführen einer PV am realen Patienten erteilt werden könnte. Für bereits qualifizierte Chirurgen könnte ein „positives Ergebnis“ des PV-Scores eine individuelle Verbesserung im Rahmen der vorliegenden Arbeitssituation bedeuten (z.B. psychologische Steigerung des Wohlbefindens durch Bestätigung der eigenen Leistung, Wertschätzung durch Kollegen, etc.). (79)

Indem als „positives Ergebnis“ beispielsweise die Noten 1 und 2 definiert würden, ließen sich so Anreize auf das Erreichen eines möglichst guten Resultates generieren. „Negative Ergebnisse“ (z.B. Note 5 und 6) des PV-Scores könnten dazu führen, dass ein noch in der Ausbildung befindlicher Chirurg wiederholte Trainingsdurchläufe vor der Möglichkeit der selbstständigen Behandlung durchführen müsste. „Negative Ergebnisse“ eines bereits qualifizierten Chirurgen könnten aber zum vorübergehenden Entzug seiner selbstständigen Behandlungslizenz führen, bis ein positives Score-Ergebnis nachgewiesen werden kann.

Im Rahmen der vorliegenden Arbeit wurde der PV-Score, sowie die Daten der PV-Trainingssimulation ausschließlich zur Analyse der operativen Leistung des behandelnden Chirurgen verwendet (s. Begriff der „Single-User-Study“ im Unterkapitel 2.1). Da das simulierte Setup der PV-Trainingssimulation jedoch den kompletten intraoperativen Behandlungsablauf für ein gesamtes OP-Team beinhaltet, ließe sich durch Integration entsprechender Metriken und Items die Bewertung der Kompetenz auf das gesamte beteiligte OP-Team ausweiten. Bei einem demnach als „Multiple-User-Study“ zu bezeichnenden Ansatz stünde jedoch voraussichtlich eine Strapazierung der Aufwand-Nutzen-Relation bevor.

Diese Ansätze sind im Sinne der Qualitätssicherung und Patientenschutz zu verstehen und greifen damit einen aktuellen Schwerpunkt auf, da sich die Qualitätssicherung im Krankenhaus derzeit in einer wichtigen Phase der Weiterentwicklung befindet. (80)

Es wäre vorstellbar, das Ergebnis des PV-Scores eines Teilnehmers nach Absolvieren der PV-Trainingssimulation im Sinne eines Zertifikates darzustellen. In Anlehnung an das „OSATS Assessment of Performance“- Zertifikat, welches das Royal College of Obstetricians and Gynaecologists den Absolventen seiner Trainingseinheiten ausstellt, sei im Anhang die Version eines möglichen Zertifikates ⁴⁰ an Absolventen der in dieser Arbeit thematisierten PV-Trainingssimulation unter Zuhilfenahme des PV-Scores vorgestellt. (81)

⁴⁰ s. 7.3, „Anhang 3“

Ein möglicher zukünftiger Ansatz der praktischen Nutzbarmachung des PV-Scores könnte in der Erstellung einer Website liegen, wie es der Heidelberg VR Score vormacht. (21 S. 5)

Dieser bietet die Möglichkeit, laparoskopische Performance an VR-Trainingssimulatoren zu bewerten und soll auf einer - zum Zeitpunkt des Verfassens dieser Arbeit noch in der Entwicklung befindlichen – Website ⁴¹ der Öffentlichkeit zugänglich gemacht wird.

Den PV-Score ebenfalls auf diese Art und Weise zur Verfügung zu stellen, könnte für andere Forschungsgruppen oder Anwender von PV-Trainingssimulationen von Nutzen sein.

5 Zusammenfassung

Simulationsbasiertes Training (SBT) anhand 3D-gedruckter Knochenmodelle erfährt im Rahmen alternativer Trainingsmethoden - ergänzend zu konventionellen Trainingsmethoden - zunehmend Anwendung in der wirbelsäulenchirurgischen Ausbildung. Vorteile des SBT liegen in der Risikofreiheit der Methode und dementsprechend einer erhöhten Patientensicherheit, Reduktion der langen und beschwerlichen Lernkurve minimal-invasiver Eingriffe, der Möglichkeit gezielter anwenderabhängiger Trainingsgestaltung, sowie postoperativer Analysemöglichkeit der Simulatordaten. Dieser Paradigmenwechsel von zeit-, struktur-, sowie prozessbasiertem Lernen hin zu kompetenz-basiertem Lernen ruft ein gesteigertes Interesse an der Einführung praktikabler wissenschaftlicher Messinstrumente hervor, die chirurgische Kompetenz durch objektive Evaluation messbar machen. Das Ziel der vorliegenden Studie war die Operationalisierung chirurgischer Kompetenz am Beispiel einer Mixed-Reality-basierten Trainingssimulation für perkutane Vertebroplastie (MR-basierte PV-Trainingssimulation).

Im Rahmen eines Pilotprojektes wurde diesem Bestreben durch die Konzeption eines Scoring-Systems, welches als Messinstrument ein sachliches und präzises Bewerten der chirurgischen Kompetenz eines Teilnehmers bei der Durchführung einer PV ermöglichen sollte, nachgekommen. Dabei sollten mögliche Forschungsansätze, Erkenntnisgewinne, sowie ggf. Widrigkeiten oder Komplikationen eruiert werden, welche einem zukünftigen größer angelegten Forschungsprojekt von Nutzen sein können. Als Forschungsmethode wurde demnach neben fortwährend empirischer Analyse ein konzeptionelles Vorgehen praktiziert. Der im Laufe der Studie angewendete Methodenpluralismus steht unter dem Aspekt des Forschungsansatzes der Mixed-methods-Studien (MMS), welcher durch das Generieren eines multiperspektivischen Blickwinkels ein vielschichtiges Verständnis über das zu erfassende Konstrukt der chirurgischen Kompetenz ermöglichen sollte.

⁴¹ <http://heidelbergvrscore.de/> (Zugriff: 04.03.2020)

Die Studie fand in Zusammenarbeit mit der interdisziplinären Forschungsgruppe Navigation and Augmented Reality Visualization System (NARVIS) unter der Leitung von Prof. N. Navab und Prof. Dr. E. Euler statt und umfasst einen experimentellen, prospektiven, sowie einen statistischen, retrospektiven Teil.

Der experimentelle Teil der Studie fand im Zeitraum 02/2018-10/2019 an der Klinik für Allgemeine, Unfall-, und Wiederherstellungschirurgie der Ludwig-Maximilians-Universität München statt. Er enthält die Durchführung der PV-Trainingssimulation durch elf Teilnehmer, sowie die dazugehörige Datenanalyse durch Auswertung von Audio- und Videodokumentationen, sowie Simulatordaten der PV-Trainingssimulation.

Im Folgenden kam es durch Methodiken der Kognitiven Aufgaben-Analyse (CTA), Konsultation einschlägiger Fachliteratur, in situ Observation einer PV, sowie in Anlehnung an den als Prototyp zu verstehenden Bewertungsbogen „Vertebroplasty: **Technical Skills Assessment Scale for Surgeons**“ (TSA-Bogen) zur Konzeption des Fragebogens über die „**Bewertung der Güte der Durchführung einer Vertebroplastie**“ (BGDV-Fragebogen).

Dessen Bearbeitung durch fünf Experten der Wirbelsäulenchirurgie ergab einen Expertenkonsensus. Basierend auf diesem Expertenkonsensus erfolgte die Konzeption des PV-Scores. Durch die Methode des Standard-Settings wurde der Algorithmus des PV-Scores formuliert. Dieser analysiert das zu erfassende Konstrukt der chirurgischen Kompetenz im Sinne einer Multi-Item-Skala. Anhand von neun Items, welche sich den Kategorien Zeit, Effizienz, Sicherheit, manuelle Geschicklichkeit, sowie Outcome zuordnen lassen, evaluiert er die Kompetenz eines Teilnehmers. Die resultierende Gesamtbewertung der Kompetenz erfolgt anhand einer Ordinalskala (Noten 1-6), welche mit der Ausformulierung seines Kompetenzniveaus in „sehr hoch/ hoch/ befriedigend/ ausreichend/ mangelhaft, bzw. ungenügend“ verbunden ist. Im Sinne eines diskriminativen Index wird anhand der Note und dem entsprechenden Kompetenzniveau eine Differenzierung der Teilnehmer in „bestanden vs. nicht bestanden“ vorgenommen.

Der retrospektive Teil der Studie erfolgte im Zeitraum vom 10/2019-02/2020 und enthält die praktische Anwendung des PV-Scores auf die Daten der Teilnehmer der PV-Trainingssimulation.

Die Bewertung der Daten durch den PV-Score ergab einen Notendurchschnitt von 3,36, wobei sich die Verteilung der Noten folgendermaßen präsentierte: 54,55% Note 2 (6 TM), 9,09% Note 3 (1 TM), 9,09% Note 4 (1 TM) und 27,27% Note 6 (3TM).

Anhand der ausgewerteten Ergebnisse des PV-Scores fanden Korrelationsanalysen zur Überprüfung der Kriteriums- und Konstruktvalidität statt.

Für die Überprüfung der Kriteriumsvalidität wurde mittels Spearman-Rangkorrelationen die Berufserfahrung (ERF) der Teilnehmer als Außenkriterium herangezogen, um zu beurteilen, ob diese mit dem Ergebnis des PV-Scores der jeweiligen Teilnehmer korrelierte. Dabei wurde die Berufserfahrung in drei Ausprägungsformen angegeben. In keinem der drei Fälle ergab sich statistische Signifikanz.

- ERF_1: $\rho = -0,43$; p-Wert = 0,19

Die Note des PV-Scores korreliert in mittlerem Maße mit der ERF_1 der Teilnehmer.

- ERF_2: $\rho = -0,34$; p-Wert = 0,30

Die Note des PV-Scores korreliert in mittlerem Maße mit der ERF_2 der Teilnehmer.

- ERF_3: $\rho = 0,10$; p-Wert = 0,76

Die Note des PV-Scores korreliert in geringem Maße mit der ERF_3 der Teilnehmer.

Zur Überprüfung der Konstruktvalidität wurden die Ergebnisse des PV-Scores (Evaluierungstool 1) der Teilnehmer einem Vergleich mit den Ergebnissen eines bereits etablierten, vergleichbaren Evaluierungstools (Evaluierungstool 2) unterzogen. Ein Mann-Whitney-U-Test für ordinal verteilte, unabhängige Stichproben prüfte die zentralen Tendenzen der Ergebnisse der beiden Evaluierungstools mittels einer polytomen Formulierung der Ergebnisse des PV-Scores (Noten 1-6) und der bivariaten Ergebnisformulierung des Evaluierungstools 2 (1=bestanden, 2=nicht bestanden). Die beiden Gruppen zeigten einen tendenziell signifikanten Unterschied (p-Wert=0,021).

Der Chi-Quadrat-Test ergab einen statistisch signifikanten Zusammenhang (p-Wert =0,026) zwischen den Ergebnissen der beiden Evaluierungstools, die für diesen Schritt beide in bivariater Formulierung vorlagen.

Somit sind erste Ergebnisse, bzw. Erfahrungen der praktischen Anwendung des PV-Scores in dieser Arbeit enthalten. Diese sind als die ersten Schritte auf einem lang angelegten Evaluierungsweg zu verstehen. Anknüpfung an den erläuterten Limitationen der vorliegenden Arbeit im Sinne von Modifikationen der Items bzw. weiteren Validierungsmaßnahmen, sowie ein klinischer Testeinsatz des konzipierten PV-Scores stellen Möglichkeiten für Anknüpfungspunkte zukünftiger Forschung dar. Langfristig könnte so ein Beitrag zur Steigerung der Akzeptanz und damit zur fortschreitenden routinemäßigen Anwendung des SBT geleistet werden, welches wiederum einen Meilenstein im Sektor Patientenschutz und Qualitätssicherung setzen kann.

6 Literaturverzeichnis

1. **NARVIS-Labor.** Computer aided medical procedures. [Online] 2018. [Zitat vom: 01. 10. 19.] <http://www.klinikum.uni-muenchen.de/Klinik-fuer-Allgemeine-Unfall-und-Wiederherstellungschirurgie/de/forschung/narvis/index.html>.
2. **Eck, U.** Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality - NARVIS Lab. [Online] 2017. [Zitat vom: 11. 11. 2019.] <http://campar.in.tum.de/Main/NarvisLabNew>.
3. **Mahnken, A.H. / Thomas, Ch. / Wilhelm, K.** Interventionelle Radiologie - Interventionelle Schmerztherapie. Stuttgart : Thieme, 2019.
4. **Wetzel, S.G. / Wilhelm, K.E.** Perkutane Vertebroplastie und Kyphoplastie. Stuttgart, New York : Othopädie und Unfallchirurgie/Thieme, 2008, Bde. 3, No. 1, S. 57-75.
5. **AMBOSS.** AMBOSS - Fachwissen für Mediziner. [Online] 2020. [Zitat vom: 20. 02. 2020.] <https://www.amboss.com/de/wissen/Wirbelk%C3%B6rperfraktur>.
6. **Jay, B. / Ahn, S.H.** Vertebroplasty. Providence : Thieme, 2013, Bde. 30, No. 3, S. 297–306.
7. **Baerlocher, M.O. / et al.** Quality Improvement Guidelines for Percutaneous Vertebroplasty. Barrie : Journal of Vascular and Interventional Radiology, 2014, Bde. 25, No. 2, S. 165–170.
8. **Carus, Th.** Operationsatlas laparoskopische Chirurgie - Ausbildung in der laparoskopischen Chirurgie. Berlin, Heidelberg : Springer, 2014, S. 19-23.
9. **Müller, Ch.W. / Krettek, Ch.** Vertebroplastie und Kyphoplastie bei osteoporotischen Wirbelkörperfrakturen: Indikationen, Technik, Ergebnisse und Komplikationen. Stuttgart, New York : OP-Journal/Thieme, 2009, Bde. 25, No. 01, S. 28-32.
10. **Manke, Ch.** Patientenaufklärung Perkutane Vertebroplastie. [Online] 2013. [Zitat vom: 23. 01. 2020.] <https://webcache.googleusercontent.com/search?q=cache:Sd5D6sHYT3cJ:https://www.klinikum-fulda.de/wp-content/uploads/2017/03/RAD-PIN-Patientenaufklaerung-Perkutane-Gastrostomie-02-0.pdf+&cd=1&hl=it&ct=clnk&gl=de>.
11. **Wucherer, P. / et al.** Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis. München : IEEE transactions on medical imaging, 2015, Bde. 34, No. 8, S. 1730-1737.
12. **Stefan, Ph. / et al.** A radiation-free mixed-reality training environment and assessment. München : International journal of computer assisted radiology and surgery, 2018, Bde. 13, No. 9, S. 1335–1344.
13. **Stefan, Ph. / et al.** Three-dimensional–Printed Computed Tomography–Based Bone Models for Spine Surgery Simulation. München : Simulation in healthcare: journal of the Society for Simulation in Healthcare, 2020, Bde. 15, No. 1.

14. **Pfandler, M. / et al.** Say, What Is on Your Mind? Surgeons' Evaluations of Realism and Usability of a Virtual Reality Vertebroplasty Simulator. München : Surgical Innovation, 2019, Bde. 26, No. 2.
15. **Mühler, K.** Animationen und Explorationstechniken zur Unterstützung der chirurgischen Operationsplanung. Magdeburg : Shaker, 2010.
16. **Kotsis, S.V. / Chung, K.C.** Application of See One, Do One, Teach One Concept in Surgical Training. Michigan : Plastic and reconstructive surgery, 2013, Bde. 131, No. 5, S. 1194–1201.
17. **Pfandler, M. / et al.** Stepwise development of a simulation environment for operating room teams: the example of vertebroplasty. München : Advances in Simulation , 2018, Bde. 3, No. 18.
18. **Pfandler, M. / et al.** Virtual reality-based simulators for spine surgery: a systematic review. München : The spine journal, 2017, Bde. 17, No. 9.
19. **Coelho, G. / Defino, H.L.A.** The Role of Mixed Reality Simulation for Surgical Training in Spine, Phase 1 Validation. Saõ Paulo : Journal of Neurosurgery: Spine, 2018, Bde. 43, No. 22, S. 1609-1616.
20. **Fuerst, D. / Hollensteiner, M. / Schrempf, A.** Assessment Parameters for a Novel Simulator in Minimally Invasive Spine Surgery. Linz : IEEE Engineering in Medicine and Biology Society, 2015, S. 5110-5113.
21. **Schmidt, M.W. / et al.** The Heidelberg VR Score: development and validation of a composite score for laparoscopic virtual reality training. Heidelberg : Springer, 2019, Bd. 33, S. 2093-2103.
22. **Chowriappa, A.J. / et al.** Development and validation of a composite scoring system for robot-assisted surgical training - the Robotic Skills Assessment Score. New York : Journal of Surgical Research, 2013, Bde. 185, No. 2 , S. 561–569.
23. **Goldenberg, M.G. / et al.** Systematic review to establish absolute standards for technical performance in surgery. Chichester : The British journal of surgery, 2017, Bde. 104, No. 1, S. 13-21.
24. **Carraccio, C. / et al.** Shifting Paradigms: From Flexner to Competencies. Baltimore : Academic Medicine, 2002, Bde. 77, No. 5, S. 361-367.
25. **Müller-Wittig, W.** Medizintechnik - Virtuelle Realität in der Medizin. Berlin, Heidelberg : Springer, 2016, S. 845-854.
26. **Hongzhi, H. / et al.** Application of mixed reality technology in surgery. Wuhan : International Journal of Clinical and Experimental Medicine, 2019, Bde. 12, No. 4, S. 3107-3113.
27. **Schlag, P.M. / Eulenstein, S. / Lange, Th.** Computerassistierte Chirurgie. München : Urban & Fischer/Elsevier, 2011, Bde. 26, No. 2.
28. **Holle, R.** Methoden zur Konstruktion und Evaluierung klinischer Scores. Heidelberg : Institut für Medizinische Biometrie und Informatik, 1995, Bd. 27.
29. **Stangl, W.** Werner Stangls Arbeitsblätter - Gütekriterien empirischer Forschung.

[Online] 2020. [Zitat vom: 08. 10. 2019.]

<https://arbeitsblaetter.stangl-taller.at/FORSCHUNGSMETHODEN/Guetekriterien.shtml>.

30. **Wübbenhorst, K.** Gabler Wirtschaftslexikon. [Online] 2018. [Zitat vom: 28. 11. 2019.]

<https://wirtschaftslexikon.gabler.de/definition/guetekriterien-35152>.

31. **Rey, G.D.** Methoden der Entwicklungspsychologie: Datenerhebung und Datenauswertung. Chemnitz : Books on Demand, 2012.

32. **Rammstedt, B.** Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung. Mannheim : Zentrum für Umfragen, Methoden und Analysen -ZUMA-, 2004, Bd. 12.

33. **Becker, F.** Wirtschaftspsychologische Gesellschaft (WPGS) - Gütekriterien wissenschaftlicher Forschung. [Online] 2020. [Zitat vom: 28. 02. 2020.]

<https://wpgs.de/fachtexte/ergebnisinterpretation/guetekriterien-wissenschaftliche-forschung/>.

34. **Niederberger, M. / Peter, L.** Mixed-Methods-Studien in den Gesundheitswissenschaften.

Ein kritischer Überblick. Schwäbisch Gmünd : Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen/Elsevier, 2018, Bd. 133, S. 9-23.

35. **Schreier, M. / Odağ, Ö.** Handbuch Qualitative Forschung in der Psychologie - Mixed Methods. Wiesbaden : VS Verlag für Sozialwissenschaften/Springer, 2010, S. 263-277.

36. **Stefan, Ph. / et al.** Teamtraining und Assessment im Mixed-reality-basierten simulierten OP. München : Der Unfallchirurg, 2018, Bde. 121, No. 6.

37. **Pfandler, M. / et al.** Technical and Non-Technical Skills in Surgery: A Simulated Operating Room Environment Study. München : The spine journal, 2019, Bde. 44, No. 23.

38. **Institut und Poliklinik für Arbeits-, Sozial- und Umweltmedizin.** Klinikum der Universität München - ATMEOS. [Online] 2016. [Zitat vom: 19. 02. 2020.]

http://www.klinikum.uni-muenchen.de/Institut-und-Poliklinik-fuer-Arbeits-Sozial-und-Umweltmedizin/de/forschung/arbeitsgruppen/AMPA/forschung/laufende/DFG_Simulation.html.

39. **Reinhold, M. / et al.** Wirbelsäule kompakt, Der DWG-Pocketguide. Stuttgart : Schattauer, 2017, S. 13.

40. **Sieck, W.** What is Cognitive Task Analysis? [Online] 2019. [Zitat vom: 17. 11. 2019.] <https://www.globalcognition.org/cognitive-task-analysis/>.

41. **Sevdalis, N.** Observational Teamwork Assessment for Surgery (OTAS). [Online] 2011. [Zitat vom: 28. 12. 2019.] <https://www.imperial.ac.uk/surgery-cancer/pstrc/Otasmanualenglish>.

42. **Evolutionizer.** Solyp - Strategielexikon. [Online] 2020. [Zitat vom: 05. 01. 2020.] <https://www.solyp.com/de/know-how/strategielexikon>.

43. **Höffken, O. / Tegenthoff, M.** Neurowissenschaftliche Begutachtung - Traumatisch bedingte Rückenmarkschäden. Stuttgart : Thieme, 2018, S. 525-527.
44. **Ulfig, N.** Kurzlehrbuch Neuroanatomie - Die Spinalnerven und ihre Äste. Stuttgart : Thieme, 2008, S. 33-35.
45. **Klocker, J. / Gratl, A. / Fraedrich, G.** Iatrogene Gefäßverletzungen infolge chirurgischer Eingriffe. Gablitz : Verlag für Medizin und Wissenschaft, 2016, Bde. 13, No. 1, S. 5-8.
46. **Schanze, R.** Was ist ein Algorithmus? – Einfach erklärt. [Online] 2017. [Zitat vom: 05. 02. 20.] <https://www.giga.de/ratgeber/specials/was-ist-ein-algorithmus-einfach-erklart/>.
47. **Pant, H.A. / Tiffin-Richards, S. / Köller, O.** Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes - Standard-Setting für Kompetenztests im Large-Scale-Assessment. Projekt Standardsetting. Weinheim, Basel : Zeitschrift für Pädagogik/Beltz, 2010, S. 175-188.
48. **Becker, F.** Wirtschaftspsychologische Gesellschaft (WPGS) - Bewusste Verfahren: Schneeballverfahren. [Online] 2020. [Zitat vom: 27. 01. 20.] <https://wpgs.de/fachtexte/stichproben/17-bewusste-verfahren-schneeballverfahren/>.
49. **Downing, S.M.** Reliability: on the reproducibility of assessment data. Illinois : Medical Education, 2004, Bd. 38, S. 1006-1012.
50. **Lois, D. - Fakultät für Humanwissenschaften Sozialwissenschaftliche Methodenlehre.** Deskriptive Statistik. [Online] 2015. [Zitat vom: 07. 01. 2020.] <https://www.unibw.de/hum-bildungswissenschaft/professuren/swm/methodenskripte/deskriptive-statistik.pdf>.
51. **Pfeiffer, F.** Scribbr. [Online] 2018. [Zitat vom: 28. 11. 2019.] <https://www.scribbr.de/methodik/auswertung-bachelorarbeit/>.
52. **Kuckartz, U.** Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Berlin : Beltz Juventa, 2018.
53. **Hemmerich, W.A.** Mathe Guru. [Online] 2011-2020. [Zitat vom: 16. 03. 2020.] <https://matheguru.com/stochastik/effektstarke.html>.
54. **Koch, Th. / Bergmann, C. / Ptok, M.** Auditive Verarbeitung und Wahrnehmungsstörungen: Erfahrungen mit Fragebogeninstrumenten. Stuttgart, New York : Laryngo-Rhino-Otologie/Thieme, 2019, Bde. 98, No. 6.
55. **Cohen, J.** Statistical Power Analysis for the Behavioral Sciences - Differences between correlation coefficients. New York : Academic Press/Elsevier, 1977, S. 109-143.

56. **Schwarz, J.** Universität Zürich - Rangkorrelation nach Spearman. [Online] 17. 05 2019. [Zitat vom: 04. 02. 2020.]
https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/rangkorrelation.html.
57. **Smigierski, J.** Smigierski Statistische Beratung. [Online] 2020. [Zitat vom: 16. 03. 2020.]
<https://www.beratung-statistik.de/statistik-beratung-infos/spss-tips/spss-chi-quadrat-test-berechnen/>.
58. **Gertzbein, S.D. / Robbins, S.E.** Accuracy of pedicular screw placement in vivo. Toronto : Lippincott-Raven Publishers, 1990, Bd. 15, S. 11–14.
59. **Oberthaler, W.** Ist die Vertebroplastie wirklich ein risikoreiches Verfahren? Gablitz : Journal für Mineralstoffwechsel & Muskuloskelettale Erkrankungen/Krause & Pachernegg, 2004, Bde. 11, No. 1, S. 20-25.
60. **Heini, P.F. / Wälchli, B. / Berlemann, U.** Percutaneous transpedicular vertebroplasty with PMMA: operative technique and early results - A prospective study for the treatment of osteoporotic compression fractures. Bern : European spine journal, 2000, Bde. 9, No. 5, S. 445–450.
61. **Amar, A.P / et al.** Percutaneous Transpedicular Polymethylmethacrylate Vertebroplasty for the Treatment of Spinal Compression Fractures. Los Angeles : Neurosurgery, 2001, Bde. 49, No. 5, S. 1105-1115.
62. **Al-Ali, F. / Barrow, T. / Luke, K.** Vertebroplasty: what is important and what is not. Kalamazoo : American Journal of Neuroradiology, 2009, Bde. 30, No. 10, S. 1835–1839.
63. **Mathis, J.M. / Deramond, H. / Belkoff, S.M.** Femoral Neck Augmentation - Percutaneous Vertebroplasty: Procedure Technique. New York : Springer, 2006, S. 112-133.
64. **Yuk-Chuen, S. / Man-Hong, Ch. / Chun-Man, M.** Intraspinal Leakage of Cement During Vertebroplasty for an Elderly Woman with Osteoporotic Burst Fracture: A Case Report and Short Review of Prevention and Management. 19. Hong Kong : Journal of Orthopaedics, Trauma and Rehabilitation/Elsevier, 2015, Bd. 19, S. 100-106.
65. **Euler, E. / et al.** Thermotoxicity of palacos cement. Clinical observation. s.l. : Der Unfallchirurg, 1989, Bde. 92, No. 12, S. 606-610.
66. **Schulz, C. / et al.** Erfahrungen mit intravasalen Zementaustritten nach Ballonkyphoplastie. Berlin, Heidelberg : Der Orthopäde, 2012, Bd. 41, S. 881-888.
67. **Weinfurth, A.** via 4 spine - Wirbelkörperaugmentation. [Online] 2020. [Zitat vom: 18. 03. 2020.]
<http://www.via4spine.de/wirbelk%C3%B6rperaugmentation.html>.

68. **Fribourg, D. / et al.** Incidence of subsequent vertebral fracture after kyphoplasty. Los Angeles : The spine journal, 2004, Bde. 29, No. 20, S. 2270-2276.
69. **Schmidt, R. / et al.** Cement leakage during vertebroplasty: an underestimated problem? Ulm : European spine journal, 2005, Bde. 14, No. 5, S. 466-473.
70. **Xu, R.-J. / et al.** A method of percutaneous vertebroplasty under the guidance of two C-arm fluoroscopes. Suzhou : Pakistan Journal of Medical Sciences, 2014, Bde. 30, No. 2, S. 335-338.
71. **Kwon, H.M. / et al.** Appropriate Cement Volume in Vertebroplasty: A Multivariate Analysis with Short-Term Follow-Up. s.l. : Korean Journal of Neurotrauma, 2016, Bde. 12, No.2 , S. 128-134.
72. **Metzler, P. / Krause, B.** Methodischer Standard bei Studien zur Therapieevaluation. [Online] 1997. [Zitat vom: 12. 12. 2019.]
<https://www.dgps.de/fachgruppen/methoden/mpr-online/issue3/art7/node6.html>.
73. **Halbmayer, E. - Institut für Kultur- und Sozialanthropologie, Universität Wien.** Einführung in die empirischen Methoden der Kultur- und Sozialanthropologie. [Online] 2010. [Zitat vom: 17. 12. 2019.]
<https://www.univie.ac.at/ksa/elearning/cp/ksamethoden/ksamethoden-94.html>.
74. **Werner, M.** Einflussfaktoren des Wissenstransfers in wissensintensiven Dienstleistungsunternehmen - Konzeptionelle Gestaltung und Durchführung der empirischen Untersuchung. Wiesbaden : Deutscher Universitätsverlag, 2004, S. 159-182.
75. **Glück, O.** Welt der BWL. [Online] 2011-2020. [Zitat vom: 17. 03. 2020.]
<https://welt-der-bwl.de/Induktive-Statistik>.
76. **Hardison, C.M. / Vilamovska, A.M.** The Collegiate Learning Assessment: Setting Standards for Performance at a College or University. Berkeley : RAND Corporation, 2009, S. 19–28.
77. **Haaker, R.G. / et al.** Verification of the position of pedicle screws in lumbar spinal fusion. Bochum : European spine journal, 1997, Bd. 6, S. 125-128.
78. **Foundation, Wikipedia.** Wikipedia The Free Encyclopedia - High-stakes testing. [Online] 2020. [Zitat vom: 17. 03. 2020.] https://en.wikipedia.org/wiki/High-stakes_testing.
79. **Schmidt-Atzert, L. / Amelang, M.** Psychologische Diagnostik - Zuordnungs- und Klassifikationsstrategien. Berlin, Heidelberg : Springer, 2012.
80. **Bundesministerium für Gesundheit/Referat L7 "Presse, Internet, Soziale Netzwerke".** Bundesministerium für Gesundheit - Qualitätssicherung im Krankenhausbereich. [Online] 2020. [Zitat vom: 28. 02. 2020.]
<https://www.bundesgesundheitsministerium.de/qualitaet-krankenhausversorgung.html>.

81. **Gynaecologists, The Royal College of Obstetricians & Gynaecologists.** Royal College of Obstetricians and Gynaecologists. [Online] 2020. [Zitat vom: 19. 01. 2020.]
<https://www.rcog.org.uk/en/careers-training/about-specialty-training-in-og/assessment-and-progression-through-training/workplace-based-assessments/osats/>.
82. **Hanzal, E. / et al.** Palpation für das Beckenbodentraining - Was ist OSATS? Wien : De Gruyter, 2015, S. 61-65.
83. **Schmidt, H.J.** 3D make. [Online] 2020. [Zitat vom: 25. 02. 2020.]
3dmake.de/thema/fff-fused-filament-fabrication/.
84. **Schulze, D.** Einsatz und Wertigkeit der CT gestützten Navigation bei der operativen Versorgung instabiler thorakolumbaler Wirbelfrakturen. Leipzig : Universität Leipzig, 2013.
85. **Schmidt-Atzert, L. / Amelang, M.** Psychologische Diagnostik - Zuordnungs- und Klassifikationsstrategien. Berlin, Heidelberg : Springer, 2012, S. 409-428.

7 Anhang

7.1 BGDV-Fragebogen mit Fragebogenanhang

Fragebogen über die Bewertung der Güte der Durchführung einer Vertebroplastie zur Erstellung eines „Pedikel-Zugangs-Scores“

Der Fragebogen bezieht sich auf eine Simulation, bei der Chirurgen verschiedener Erfahrungsstandards eine Vertebroplastie an einem 3D-gedruckten Wirbelsäulenabschnitt (Höhe LWK2) durchführten. Um nachvollziehen zu können, ob die Teilnehmer bei dem Eingriff Risikostrukturen punktierten, bzw. welchen Weg die Nadel zurückgelegt hat, wurde die Lage der verwendeten Nadel nachverfolgt.

Auf Basis der ausgewerteten Fragebögen soll ein „Pedikel-Zugangs-Score“ erstellt werden, anhand dessen eine Bewertung der Performance der Teilnehmer der oben genannten Vertebroplastiestudie ermöglicht werden soll.

Ziel ist es, dass der Score für Vertebroplastien und (wenn möglich) allgemein für Pedikulierungen eine objektive Bewertungsmöglichkeit darstellt.

1. In Kategorie „Trokar-basiert“

a) Bitte nummerieren Sie die folgenden Risikostrukturen absteigend nach Ihrer Kritizität bei Durchstechen/Verletzung durch den Trokar: (von 1=am kritischsten bis 13=am unkritischsten)

___ anderer Wirbelkörper	___ Bandscheiben	___ anteriore WK-Wand
___ Deck-/Bodenplatte	___ Aorta	___ V.cava
___ Segmentgefäße	___ Muskulatur	___ Spinalkanal
___ Spinalnervenzwurzeln	___ Facettengelenk	___ Lunge
___ Rückenmark		

Gruppieren Sie die Reihung danach in die Kategorien...

a = Verletzung der Patientensicherheit, fügt dem Patienten Schaden zu
b = Gefährdung der Patientensicherheit
c = nicht optimal, aber keine Gefährdung der Patientensicherheit
d = unkritisch

... indem sie die Buchstaben (a-d) neben die jeweilige Struktur schreiben.

b) Halten Sie die vier Kategorien (a-d) für sinnvoll?

Ja: _____

Nein: _____. Wenn nein, welche Kategorisierung finden Sie sinnvoller? _____

c) Fallen Ihnen noch weitere relevante Risikostrukturen in Höhe LWK2 ein?

d) Bitte geben Sie in der vorliegenden Tabelle Werte (in mm) an, die Sie bei einer Punktion der jeweiligen Risikostruktur als kritisch bzw. noch unkritisch einstufen würden. Als Anhaltspunkt für die Dimensionen der jeweiligen Risikostrukturen sind die gemittelten Werte in mm angegeben, gemessen im CT-Datensatz (Spine 2017, Höhe LWK2), in Richtung der ungefähren Verlängerung der Pedikelachse.

Struktur	Dimension im behandelten WS-Abschnitt (in mm)	unkritisch bis zu einem Abstand von ... mm zur genannten Struktur	Bedingt kritisch bis zu einem Abstand von... mm zur genannten Struktur (bzw. ohne gravierende Auswirkungen aufs Patienten-Outcome)	Kritisch ab ... mm Abstand zur genannten Struktur
Rückenmark	6,7			
Spinalkanal	10,0			
Aorta	15,9			
V. cava	6,7			
Spinalnervenzurwurzeln	3,7			
Lunge	k.A.			
Segment-Gefäß	4,3			
Bandscheiben	19,0			
Facettengelenk	5,5			
Anderer Wirbelkörper	32,3			
Muskulatur	21,7			

e) Wie wichtig ist es für Sie für die Bewertung der Performance eines Chirurgen wie oft eine Risikostruktur punktiert wurde? —

welche Risikostruktur punktiert wurde? —

über welche Strecke, bzw. Tiefe hinweg eine Risikostruktur punktiert wurde? —

Beantwortung der Fragen:

1 sehr wichtig 2 wichtig 3 egal 4 unwichtig 5 vollkommen unwichtig

f) Bitte zeichnen Sie am Ihnen vorliegenden Wirbelmodell (Programm „Meshmixer“ im PC) in grün den Bereich ein, den Sie als den „idealen Eintritt in den Pedikel“ ansehen und in orange den Bereich, den sie als „akzeptablen Eintritt in den Pedikel“ ansehen.

g) Können Sie den „idealen Eintrittspunkt“ in den Pedikel definieren? (z.B. anatomische Bezugspunkte, Leitstrukturen, Orientierungsebenen/-linien, ...)

h) Bis zu wie viel Millimeter Abweichung von diesem idealen Eintrittspunkt würden Sie als ideal/akzeptabel/unpassend werten?

Abweichung: ideal bis zu ___ mm

nicht ideal, aber akzeptabel bis zu ___ mm

nicht akzeptabel ab ___ mm

i) Bis zu wie viel maliges Ansetzen der Nadel am Eintrittspunkt am Knochen halten Sie beim Platzierungsversuch für angemessen? Bis zu _____ mal.

j) Wie oft ist ein Zurückziehen der Nadel mit nachfolgend erneuter Hautperforation beim Platzierungsversuch noch akzeptabel? Bis zu _____ mal.

k) Bis zu wie viel Zeit und bis zu wie viele benötigte Röntgenbilder erachten Sie beim Platzierungsversuch bis zur Endlage des Trokars noch für angemessen?

Anzahl Röntgenbilder: bis zu _____ Röntgenbilder

Zeit: bis zu _____ Minuten

l) Können Sie eine Ebene definieren/festlegen, anhand derer wir eine Bewertung der Tiefe der Trokarspitze in der Endlage vornehmen können?

Ziel ist eine Bewertung der Trokartiefe im Sinne von: „Die Tiefe der Trokarspitze für die zweiseitige Zementinjektion ist N-mm zu wenig tief/passend/zu tief.“

Wir legen fest, dass es sich bei unserer Betrachtung ausschließlich um einen bipedikulären Zugang handelt, da dieser bei der Pedikulierung den relevanteren im Vergleich zum unipedikulären darstellt.

Bitte zeichnen Sie in der Ihnen vorliegenden 3D-Darstellung (Programm “3D Slicer Trajectories” am PC) die ideale Nadelposition beim bipedikulären Zugang in drei Ebenen ein (zweimal in der lateralen, einmal in der seitlichen Projektion).

m) Gibt es einen minimalen Abstand der Nadelspitze zur Deck-/Bodenplatte, der eingehalten werden sollte? Wenn ja, welchen?

Nein: _____

Ja: _____, zur Bodenplatte einen Abstand von _____ mm und zur Deckplatte einen Abstand von _____ mm.

n) Gibt es einen minimalen Abstand der Nadelspitze vom anterioren WK-Rand, der eingehalten werden sollte? Wenn ja, welchen?

Nein: _____

Ja: _____, einen Abstand von _____ mm

2. In Kategorie „Zement-basiert“

a) Welche Zementmenge erachten Sie für den Ihnen vorliegenden Wirbel im CT-Datensatz als zu wenig/passend/zuviel?

zu wenig: bis zu _____ ml

passend: _____ bis _____ ml

zu viel: ab _____ ml

b) Ist die Injektion von zu wenig oder von zu viel Zement als negativer zu werten?

(Bitte ankreuzen und Antwort begründen)

Zu viel Zement negativer: _____

Zu wenig Zement negativer: _____

Begründung: _____

c) Bitte nummerieren Sie die folgenden Risikostrukturen, die bei einer Zementleckage betroffen sein können, in absteigender Reihenfolge nach ihrer Kritizität: (von 1=am kritischsten bis 6=am unkritischsten)

____ Subduralraum	____ prävertebrale Venen
____ intraspinale Venen	____ Bandscheiben
____ prävertebraler Raum	____ Segmentgefäße

Gruppieren Sie die Reihung danach in die Kategorien ...

a = Verletzung der Patientensicherheit, fügt dem Patienten Schaden zu
b = Gefährdung der Patientensicherheit
c = nicht optimal, aber keine Gefährdung der Patientensicherheit
d = unkritisch

... indem Sie den jeweiligen Buchstaben (a-d) hinter die Struktur schreiben.

Ändert sich etwas an der Kritizität einer der genannten Strukturen, je nachdem, ob sie von der Nadel punktiert oder von Zementaustritt betroffen wurde?

Nein: _____

Ja: _____

Wenn ja: an welchen Strukturen und warum?

d) Fallen Ihnen außer den genannten noch weitere bei einer Zementleckage relevante Risikostrukturen in Höhe LWK2 ein?

Nein: _____

Ja: _____, folgende:

e) Welche Zementmenge (in ml) sollte idealerweise injiziert werden, bis ein Röntgenkontrollbild angefertigt wird? Antwort: _____ ml

f) Ist es als negativer zu werten, wenn das Röntgenkontrollbild nach zu großer oder nach zu kleiner Zementmenge angefertigt wurde? Bitte ankreuzen und begründen.

Negativer nach zu großer Zementmenge: _____

Negativer nach zu kleiner Zementmenge: _____

Begründung: _____

3. In Kategorie „Bildgebungs-basiert“

a) Können Sie eine „passende Einstellung“ des C-Bogens definieren?

(z.B. durch Festlegung zweier Ebenen in dem Wirbelkörper, anhand derer dann die Abweichung der Röntgenquelle von ihrer „Idealposition“ gemessen werden kann)

Bitte stellen Sie an der Ihnen vorliegenden Röntgensimulation (Programm „3D Slicer Röntgen Einstellung“ am PC) den Bildwandler für beide Seiten (Trokarpazierungen linker und rechter Pedikel) passend ein (jeweils in a.p. und lateral).

b) Welche Anzahl an angefertigten Röntgenbildern erachten Sie in der Gesamtprozedur als angemessen? Bis zu _____ Röntgenbilder erachte ich als angemessen.

c) Ist ein Zuviel oder ein Zuwenig an angefertigten Bildern negativer zu werten?

Zuviel Fluoroskopiebilder negativer: _____

Zuwenig Fluoroskopiebilder negativer: _____

Begründung: _____

4. In Kategorie „Anderes“

a) Bitte definieren Sie ein Zeitfenster, innerhalb dessen ein optimales Verhältnis von Zeit zu Qualität des Operationsergebnisses möglich sein sollte.

In einem Zeitfenster von mindestens _____ bis zu maximal _____ Minuten sollte ein optimales Operationsergebnis durchführbar sein.

„Fragebogenanhang: Personenangaben zum Vergleich und zur Differenzierung der Ergebnisse“

Im Rahmen der Studie sind folgende Angaben zu Ihrer Person für einen Vergleich und eine Differenzierung der Ergebnisse sehr wichtig. Dabei dienen die Angaben lediglich statistischen Zwecken. Ihre Anonymität wird selbstverständlich gewährleistet.

Teilnehmer-Nummer: _____

1. Berufserfahrung seit Approbation:	_____ Jahre
2. Fachrichtung:	<input type="checkbox"/> Unfallchirurgie <input type="checkbox"/> Orthopädie <input type="checkbox"/> Neurochirurgie <input type="checkbox"/> Radiologie <input type="checkbox"/> Andere: _____
3. Wie viele wirbelsäulenchirurgische Eingriffe haben Sie bereits durchgeführt?	_____ Stück
4. Wie viele Vertebroplastien haben Sie bereits durchgeführt?	_____ Stück
5. Wie viele Jahre führen Sie bereits selbständig wirbelsäulenchirurgische Eingriffe durch (d.h. als verantwortlicher OperateurIn)?	_____ Jahre

Wir bedanken uns sehr herzlich bei Ihnen für Ihre Mitwirkung!

7.2 TSA-Bogen⁴²

Vertebroplasty: TECHNICAL SKILLS ASSESSMENT SCALE FOR SURGEONS Please assess the surgeons performance among following questions				
For the LEFT SIDE				
TS_1_L	With this trocar position, No-Go-areas (spinal canal, vessels, ...) are harmed?	<input type="checkbox"/> no	<input type="checkbox"/> yes	
TS_2_L	If yes: which one(s)? _____			
TS_3_L	ENTRY POINT: The entry point ist chosen right... (craniocaudal, mediolateral)	<input type="checkbox"/> appropriate	<input type="checkbox"/> not appropriate on one axis	<input type="checkbox"/> not appropriate on both axis
TS_4_L	POSITIONING ATTEMPT: To find the entry point the trocar is placed _____ times.			
TS_5_L	I rate this as ...	<input type="checkbox"/> appropriate	<input type="checkbox"/> too often	
TS_6_L	TROCAR POSITION: The trocar position within the pedicle and the vertebra is...	<input type="checkbox"/> ideal	<input type="checkbox"/> not ideal but uncritical	<input type="checkbox"/> critical
TS_7_L	PEDICLE PERFORATION: (Gertzbein and Robbins, 1990) The trocar is ...	<input type="checkbox"/> within in pedicle (grade 0)	<input type="checkbox"/> 0-2 mm outside the pedicle (grade 1)	<input type="checkbox"/> 2-4 mm outside the pedicle (grade 2)
TS_8_L	If the pedicle was perforated, in which direction was it perforated?	Answer: _____		
TS_9_L	VERTEBRA PERFORATION: The vertebra was perforated:	<input type="checkbox"/> no	<input type="checkbox"/> yes	

⁴² Die Abbildung dieser Version des TSA-Bogens erfolgt mit freundlicher Genehmigung durch das Institut für Arbeits-, Sozial- und Umweltmedizin, sowie durch das NARVIS-Labor. Außerdem ist der TSA-Bogen in einer leicht abgewandelten Version auffindbar unter: <http://links.lww.com/BRS/B445> (Appendix 1). (37)

TS_10_L	TROCAR DEPTH: For a two-sided cement injection, the depth of the tip of the trocar is...	<input type="checkbox"/> not deep enough	<input type="checkbox"/> appropriate	<input type="checkbox"/> too deep
---------	---	--	--------------------------------------	-----------------------------------

For the RIGHT SIDE

TS_1_R	With this trocar position, No-Go-areas (spinal canal, vessels, ...) are harmed?	<input type="checkbox"/> no	<input type="checkbox"/> yes
--------	---	-----------------------------	------------------------------

TS_2_R	If yes: which one(s)? _____		
--------	--------------------------------	--	--

TS_3_R	ENTRY POINT: The entry point is chosen right... (craniocaudal, mediolateral)	<input type="checkbox"/> appropriate	<input type="checkbox"/> not appropriate on one axis	<input type="checkbox"/> not appropriate on both axes
--------	--	--------------------------------------	--	---

TS_4_R	POSITIONING ATTEMPT: To find the entry point the trocar is placed _____ times.
--------	--

TS_5_R	I rate this as ...	<input type="checkbox"/> appropriate	<input type="checkbox"/> too often
--------	--------------------	--------------------------------------	------------------------------------

TS_6_R	TROCAR POSITION: The trocar position within the pedicle and the vertebra is...	<input type="checkbox"/> ideal	<input type="checkbox"/> not ideal but uncritical	<input type="checkbox"/> critical
--------	--	--------------------------------	---	-----------------------------------

TS_7_R	PEDICLE PERFORATION: (Gertzbein and Robbins, 1990) The trocar is ...	<input type="checkbox"/> within in pedicle (grade 0)	<input type="checkbox"/> 0-2 mm outside the pedicle (grade 1)	<input type="checkbox"/> 2-4 mm outside the pedicle (grade 2)	<input type="checkbox"/> > 4 mm outside the pedicle (grade 3)
--------	--	--	---	---	---

TS_8_R	If the pedicle was perforated, in which direction was it perforated?	Answer: _____
--------	--	------------------

TS_9_R	VERTEBRA PERFORATION: The vertebra was perforated:	<input type="checkbox"/> no	<input type="checkbox"/> yes
--------	--	-----------------------------	------------------------------

TS_10_R	TROCAR DEPTH: For a two-sided cement injection, the depth of the tip of the trocar is...	<input type="checkbox"/> not deep enough	<input type="checkbox"/> appropriate	<input type="checkbox"/> too deep
---------	---	--	--------------------------------------	-----------------------------------

For BOTH SIDES

TS_11	Cement is leaking...	<input type="checkbox"/> no	<input type="checkbox"/> yes
-------	----------------------	-----------------------------	------------------------------

TS_12	If cement is leaking, the cement is leaking into: (please name the regions)	Regions: _____
-------	---	----------------

TS_13	The REGION(S) of the leakage is (are) ...	<input type="checkbox"/> uncritical	<input type="checkbox"/> critical
-------	---	-------------------------------------	-----------------------------------

TS_14	If cement is leaking in one region: the AMOUNT of leakage is...	<input type="checkbox"/> uncritical	<input type="checkbox"/> critical
-------	---	-------------------------------------	-----------------------------------

TS_15/16	If cement is leaking in more than one region: the AMOUNT of leakage in region _____ is ...	<input type="checkbox"/> uncritical	<input type="checkbox"/> critical
----------	--	-------------------------------------	-----------------------------------

TS_17/18	the AMOUNT of leakage in region _____ is...	<input type="checkbox"/> uncritical	<input type="checkbox"/> critical
----------	---	-------------------------------------	-----------------------------------

TS_19/20	the AMOUNT of leakage in region _____ is...	<input type="checkbox"/> uncritical	<input type="checkbox"/> critical
----------	---	-------------------------------------	-----------------------------------

TS_21	CEMENT AMOUNT: The injected amount of cement into the vertebra is...	<input type="checkbox"/> not enough	<input type="checkbox"/> appropriate	<input type="checkbox"/> too much
-------	---	-------------------------------------	--------------------------------------	-----------------------------------

TS_22	INJECTION SPEED: The injection speed of the cement is... (please remind the 5 second waiting period)	<input type="checkbox"/> too slow	<input type="checkbox"/> appropriate	<input type="checkbox"/> too fast
-------	---	-----------------------------------	--------------------------------------	-----------------------------------

TS_23	C-ARM SETTINGS: The C-Arm was adjusted appropriately	<input type="checkbox"/> yes	<input type="checkbox"/> partly	<input type="checkbox"/> no
-------	---	------------------------------	---------------------------------	-----------------------------

TS_24	Compared to the subject an expert would have switched from AP to lateral view...	<input type="checkbox"/> less often	<input type="checkbox"/> as often	<input type="checkbox"/> more often
-------	--	-------------------------------------	-----------------------------------	-------------------------------------

TS_25	FLUOROSCOPY 1: The number of fluoroscopy shots taken is...	<input type="checkbox"/> not enough	<input type="checkbox"/> appropriate	<input type="checkbox"/> too often
TS_26	FLUOROSCOPY 2: The fluoroscopy shots are taken at the right time and angle	<input type="checkbox"/> never or nearly never	<input type="checkbox"/> partly	<input type="checkbox"/> nearly always or always

DURATION: The duration for... was...

TS_27	the whole procedure	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow
TS_28	the team time out	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow
TS_29	locating the point of insertion on the skin (approximately)	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow
TS_30	the incision and insertion of the trocar up to the pedicle	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow
TS_31	the insertion of the trocar through the pedicle and to the endpoint in the vertebra	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow
TS_32	the cement injection	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow
TS_33	the trocar removal and suturing	<input type="checkbox"/> too fast	<input type="checkbox"/> appropriate	<input type="checkbox"/> too slow

7.3 textbezogener Anhang, weitere theoretische Grundlagen

Anhang 1: ergänzende „**Definition der idealen Einstellung des C-Bogens**“ zur TA 3a) in 4.2.1.2 qualitative Teilaufgaben

„The C arm should be skewed cranially or caudally to view the vertebral body ‚straight on‘ that is, anterior wall projected directly over the posterior wall. Prior to needle placement, the C arm is positioned in a 20- to 30-degree ipsilateral oblique view. The pedicle, especially the medial and the inferior walls, must be clearly visualized. While advancing the needle, it is imperative that the needle does not violate the medial and inferior walls of the pedicle to decrease the risks of nerve root or spinal cord injury. As such, the needle should be advanced toward the upper and outer half of the face of the pedicle.

Intermittent lateral views should also be obtained to ensure appropriate cranial to caudal angulation, access into the fractured vertebral body, and to determine when the needle reaches the posterior (dorsal) wall of the vertebral body. Until the needle enters the posterior vertebral body, AP (ipsilateral oblique) projection should be used to advance the needle. With experience, the posterior wall can often be detected with slight increased resistance in the needle, as well as a subtle change in the ‚tapping sound‘ encountered while advancing the needle with a mallet. From this point, the needle should be advanced in the lateral projection with continuous visualization.

(vgl. Jay B., Ahn S. H., 2013, S. 300) (6)

Anhang 2: ergänzende „**Definition der idealen Nadelposition**“ zur TA 1l) in 4.2.1.2 qualitative Teilaufgaben

Ultimately, the tip of the needle should be positioned at the junction of the anterior (ventral) and middle third of the vertebral body, and as close to midline as possible. Opposite direction traction on the outer portion of the needle to drive the tip in the desired direction can help ideal tip positioning. Advancing the needle further anteriorly than the junction of anterior and middle third of the vertebral body could result in violating the anterior cortex of the vertebral body. Final needle positioning should be confirmed in both the lateral and frontal projections prior to bone cement injection prior to the injection of cement. Ideally, the tips of the needles should be positioned as close to midline as possible at the junction of the anterior and middle third of the vertebral body.“ (vgl. Jay B., Ahn S. H., 2013, S. 300) (6)

Zertifikat PV-Score-Assessment chirurgischer Kompetenz

Name des Teilnehmers:	XY
Name des Bewerters/Gutachters:	XY
Procedere:	Durchführung einer PV an einer MR-basierten PV-Trainingssimulation
Klinische Details und Komplexität:	bi- und transpedikuläre Durchführung an einem LWK2 mit Deckplatten- Impressionsfraktur mit geringer Keilwirbelbildung und ohne Beteiligung der WK-Hinterkante <i>(ggf. hier besondere Anforderungen/ Gegebenheiten eintragen)</i>
Datum:	XY
Schwierigkeitsgrad:	Basic/Mittel/Fortgeschritten

Dieses Assessment in Form des PV-Scores stellt ein verbindliches Messinstrument dar und dient folgenden Zwecken:

1. Bewertung chirurgischer Kompetenz bei der Durchführung einer PV
2. Bereitstellung spezifischen, konstruktiven Feedbacks über die chirurgische Performance des Teilnehmers

Der Teilnehmer wird aufgrund seiner Gesamtpformance mit der Beurteilung „bestanden“ oder „nicht bestanden“ bewertet, was gleichbedeutend mit „kompetent“, bzw. „nicht kompetent“ angesehen werden kann.

(ggf.: Ein Teilnehmer benötigt eine Mindestzahl von drei als „bestanden“ gewerteten Durchläufen.)

Diese Beurteilung ist nur spezifisch auf dieses Assessment anzuwenden.

Die folgenden Kernaussagen dienen der generellen Orientierung über das Level der Gesamtpformance. An den Teilnehmer, der als „bestanden, bzw. kompetent“ eingestuft wurde, lauten die Anforderungen wie folgt:

Der Teilnehmer war in der Lage, alle Aspekte des Prozederes sicher und ohne jegliche, bzw. mit lediglich minimaler Hilfestellung durchzuführen. Im Falle eines unerwartet schwierigen Falles kann eine Unterstützung bei der Durchführung der schwierigeren Aspekte des Prozederes stattgefunden haben.

Der Teilnehmer, der als „nicht bestanden, bzw. nicht kompetent“ eingestuft wurde, weist folgende Merkmale auf:

„Der Teilnehmer benötigte signifikante Hilfestellungen während des gesamten Prozederes, bzw. während eines Großteils des Gesamtprozederes.“

„Der Teilnehmer war nicht in der Lage, grundlegende Bestandteile des Prozederes sicher und kompetent auszuführen.“

Der Teilnehmer führte diese PV unter Beobachtung kompetent aus. *
Der Teilnehmer kann für die Durchführung dieser PV unter Beobachtung nicht als kompetent eingestuft werden. *

* Bitte Entsprechendes durchstreichen.

Vorschläge für zu beachtende Bestandteile der PV, die keine Berücksichtigung im Assessment durch den PV-Score gefunden haben, werden in der folgenden Tabelle aufgelistet.

Checken von Instrumenten/Umgebung
Kommunikation mit Patienten und/oder Angehörigen
Peri-operative Planung, z.B. Positionieren
Gebrauch der Assistenz
Kommunikation mit der Belegschaft
Auswahl von Instrumenten und Ausrüstung
Zukunftsplanung
Wirtschaftlichkeit/Ökonomie der Bewegungen
Umgang mit Problemen und/oder Schwierigkeiten
Umgang mit Gewebe
Dokumentation
Vorsichtsmaßnahmen

Feedback:

Bitte geben Sie dem Teilnehmer in dieser Tabelle noch schriftliches Feedback zu seiner abgelieferten Performance.

Was ist gut gelaufen?

Was hätte besser laufen können?

Lernplan/ -strategie:

Unterschrift des Teilnehmers:

Unterschrift des Bewerters/Gutachters:

7.4 Abkürzungsverzeichnis

Abkürzung	Erklärung
χ^2	Chi-Quadrat
ρ (Rho)	Spearman's Rangkorrelationskoeffizient
3D	dreidimensional
ant.	anterior
AP	anterior-posterior
AR	Augmented Reality
ATMEOS	Assessment and Training of Medical Experts based on Objective Standards
BGDV-Fragebogen	(Fragebogen über die) Bewertung der Güte der Durchführung einer Vertebroplastie zur Erstellung eines „Pedikel-Zugangs-Scores“
CAS	Computer Assisted Surgery
CAOS	Computer Assisted Orthopedic Surgery
CT	Computerthomographie
CTA	Cognitive Task Analysis (Kognitive Aufgaben- Analyse)
DRR	Digitally Reconstructed Radiographs
EP/IEP	Eintrittspunkt/idealer Eintrittspunkt
ERF	(Berufs-) Erfahrung
FFF	Fused Filament Fabrication
HF	Human Factor
KM	Knochenmark
LW	Lendenwirbel
LWK	Lendenwirbelkörper
L	links
LMU	Ludwig-Maximilians-Universität

M	Mittelwert (arithmetisches Mittel)
MMS	Mixed-methods-Studien
MR	Mixed-Reality
MC	Multiple Choice
n	Anzahl der Elemente in der Stichprobe
NARVIS	Navigation and Augmented Reality Visualization System
NTS	non-technical skills
OSATS	Objective Structured Assessment of Technical Skill
OTAS	Observational Teamwork Assessment for Surgery
OP	Operation
PV	Perkutane Vertebroplastie
PLA	Poly lactide
PMMA	Polymethylmetacrylat
R	rechts
RS	Risikostruktur
RSA-Score	Robotic-Skills-Assessment-Score
Rö	Röntgen
RM	Rückenmark
sag.	sagittal
SBT	simulation-based training (Simulatortraining)
SD	Standardabweichung
SPSS	Statistical Package for the Social Sciences
TA	Teilaufgabe
TS	Technical Skills
Th	Thorakalwirbel

TSA-Bogen	Vertebroplasty: Technical Skills Assessment Scale for Surgeons
TTO	Team-Time-Out
TU	Technische Universität
VR	Virtual Reality
WK(W)	Wirbelkörper(wand)
WS	Wirbelsäule

7.5 Abbildungsverzeichnis

7.5.1 Tabellen

Tabelle 1: Vergleich von struktur- und prozessbasierten (\triangleq konventionellen) mit kompetenzbasierten (\triangleq alternativen/simulationsbasierten) Trainingsprogrammen (Quelle: in Anlehnung an Carraccio et al., 2002, (22 S. 362)).....	6
Tabelle 2: Gegenüberstellung der TA des TSA-Bogens mit den entsprechend präzisierenden TA des BGDV-Fragebogens, eigene Darstellung.....	23
Tabelle 3: Inhalt der Kategorien des PV-Scores sowie entsprechende TA des BGDV-Fragebogens und mögliche Höchstpunktzahlen der Einzel-Items, der Kategorien (durch Addition der HPZ der Einzel-Items) und des PV-Scores (durch Addition der HPZ der Kategorien)	32
Tabelle 4: Algorithmus des PV-Scores in Form eines Notenschlüssels, mit zugehörigem Resultat, Einstufung des chirurgischen Kompetenzniveaus, sowie daraus resultierender Konsequenz für den Teilnehmer. KOMP=Kompetenz, befr.=befriedigend, ausr.=ausreichend, ungen.=ungenügend.....	35
Tabelle 5: Berufserfahrung der elf Teilnehmer in drei Ausprägungsformen (ERF_1/2/3).....	39
Tabelle 6: ERG (Ergebnisse) der elf TM des TSA/OSATS-Scores in biv. (bivariater) Formulierung, bzw. des PV-Scores in biv. und polytomer Formulierung (erscheint kursiv, in Klammern), eigene Darstellung	39
Tabelle 7: Kritizitätsabstufungen (in mm) im Umfeld einer potenziell punktierten Risikostruktur...	45
Tabelle 8: ideale Nadelposition im linken Pedikel in AP und sagittaler Projektion	51
Tabelle 9: ideale Nadelposition im rechten Pedikel in AP und sagittaler Projektion	52
Tabelle 10: passende Einstellung des Bildwandlers, in AP und sagittaler Projektion	53
Tabelle 11: benötigte Zeit beim Platzierungsversuch.....	57
Tabelle 12: Dauer der Gesamt-OP	57
Tabelle 13: Anzahl angefertigter Röntgenbilder beim Platzierungsversuch.....	58

Tabelle 14: Anzahl angefertigter Röntgenbilder während Gesamt-OP	58
Tabelle 15: Punktion von Risikostrukturen	59
Tabelle 16: minimaler eingehaltener Abstand der Trokarspitze zu einer der Wirbelkörperwände ..	59
Tabelle 17: Gesamtmengen an appliziertem Zement	60
Tabelle 18: Abweichungen der Trokarspitze vom idealen Eintrittspunkt in den Pedikel	60
Tabelle 19: Anzahl der Ansatzpunkte des Trokars am Knochen	60
Tabelle 20: Auswertung der Punktzahlen der Teilnehmer anhand des PV-Scores	61
Tabelle 21: SPSS-Output der nichtparametrischen Spearman-Rangkorrelationen, Spearman-Korrelationskoeffizient ρ und p-Werte für die Noten des PV-Scores mit drei Ausprägungsformen der Erfahrung der Teilnehmer (ERF_1, ERF_2, ERF_3)	62
Tabelle 22: SPSS-Output des Mann-Whitney-U-Tests, Verteilung der Ränge und Teststatistik, signifikanter Unterschied in den Noten des PV-Scores je Bewertungsgruppe, * Gruppenvariable: Bewertungsgruppen TM mit „Note 1,00“ bzw. TM mit „Note 2,00“ (bezogen auf die Bewertung durch den TSA/OSATS-Score).....	65
Tabelle 23: SPSS-Output des Chi-Quadrat-Tests, Kreuztabelle zu Note_TSA/OSATS mit Note_PV und Teststatistik.....	67
Tabelle 24: vergleichende Gegenüberstellung der Angaben des Expertenkonsensus des PV-Scores zur „Abweichung des Trokars vom IEP“ mit der G&R-Klassifikation zur „Abweichung von Pedikelschraubenlagen“	71

7.5.2 Abbildungen

Abbildung 1: Versuchsdesign-Grafik, blau=prospektive Anteile, grün=retrospektive Anteile, eigene Darstellung.....	14
Abbildung 2: OP-Team (anonymisiert): behandelnder Chirurg, sterile und unsterile Schwester, Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU	16
Abbildung 3: Schematisierte Abb. des Trackingsystems (=W), C-Bogen (=C), Patienten-modell (=P), virtuelle Kamera (=S), Instrument/Trokar (=T), Quelle: (11 S. 1337).....	16
Abbildung 4: künstliche Orientierungsmarkierungen (blau und gelb): (a) im CT-Datensatz der Patientin, (b) im 3D- Knochenmodell, (c) im CT des 3D-Knochenmodells, (d) 3D-Knochenmodell in Gußform mit rotem Gelwachs, Quelle: (12 S. 1338)	17
Abbildung 5: mit Kunsthaut überzogener OP-Situs des Mannequins, Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU	18

Abbildung 6: OP-Umgebung mit Instrumententisch, C-Bogen und Mannequin, Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU	18
Abbildung 7: AP (anterior-posterior) und sag. (sagittal) Rö- (Röntgen-) Aufnahme des LWK2 mit Trokar, Quelle: (35 S. 276)	19
Abbildung 8: Instrumente (Tisch der sterilen OP-Schwester), Quelle: Institut für Arbeits-, Sozial- und Umweltmedizin der LMU	19
Abbildung 9: Überblick über die Gliederungsebenen der Teilaufgaben des BGDV-Fragebogens: Art der Bearbeitung, Art des Dateninformationsgehaltes, Art der Fragestellung, Verwendungszweck, inhaltliche Komponenten, eigene Darstellung	24
Abbildung 10: Schritte der Konzeption des PV-Scores, eigene Darstellung	28
Abbildung 11: generelles Ablaufschema einer qualitativen Inhaltsanalyse, vgl. Kuckartz, 2016, S.45, Abb.4, (52).....	41
Abbildung 12: Ranking der Kritizitätsangaben und Kritizitätskategorien bei Punktion einer Risikostruktur, (aufsteigend, 1,0=am kritischsten, 13,0=am unkritischsten), die Kritizitätskategorien a-d wurden empirisch festgelegt.....	43
Abbildung 13: Ranking der Kritizitätsangaben und -abstufungen, (aufsteigend, 1=am kritischsten, 4,8=am unkritischsten), die Kritizitätskategorien a-d wurden empirisch festgelegt.....	47
Abbildung 14: ideale und akzeptable Eintrittspunkte des Trokars in die Pedikel, beidseitig	50
Abbildung 15: idealer und akzeptabler Eintrittspunkt des Trokars in den linken Pedikel	50
Abbildung 16: idealer und akzeptabler Eintrittspunkt des Trokars in den rechten Pedikel	50
Abbildung 17: Perkutane-Vertebroplastie-Score (PV-Score)	56
Abbildung 18: prozentuale Häufigkeiten der Noten des PV-Scores, eigene Darstellung	61
Abbildung 19: Histogramm der „Ergebnisse im PV-Score" (Noten 1-6) je Bewertungsgruppe „Note 1,00, bzw. 2,00" des TSA/OSATS-Scores: absolute Häufigkeiten und Mittelwerte der erzielten Noten des PV-Scores	64
Abbildung 20: Histogramm der absoluten Häufigkeiten der Teilnehmer einer „TSA/OSATS-Note" (bivariat, 1 oder 2) je „PV-Note" (bivariat, 1 oder 2).....	66

8 Danksagung

Ich möchte mich bei Herrn Prof. Dr. Wolfgang Böcker für die Möglichkeit bedanken, die vorliegende Arbeit an der Klinik für Allgemeine, Unfall-, und Wiederherstellungschirurgie der Ludwig-Maximilians-Universität München durchführen zu können.

Mein besonderer Dank gilt Herrn Prof. Dr. Ekkehard Euler, für die Überlassung des Themas, seine engagierte fachliche Begleitung und fortwährende Präsenz bei der Bearbeitung des Themas, sowie für seine unkonventionelle Art, Doktorvater zu sein und für seine allgemein sehr bereichernde Persönlichkeit.

Vielen Dank an Frau Dr. med. Anna Maria von der Heide für ihr Mitwirken an der vorliegenden Arbeit und ihre Hilfsbereitschaft, insbesondere bei der Erschließung des Themengebietes und Forschungsinhalte, sowie bei der Gestaltung des experimentellen Teils der Arbeit.

Ich danke allen beteiligten Informatikern, Psychologen, Pädagogen, Chirurgen und weiteren Mitarbeitern des NARVIS Labors, sowie allen Probanden, die durch ihr Mitwirken und ihre Teilnahme diese Studie erst ermöglicht haben.

Ich danke den Herren Dipl.-Inf. Philipp Stefan und B.Sc. Björn-Aljoscha Kullmann sehr für ihr Mitwirken in zahlreichen Belangen und für ihre ständige informatische Unterstützung, welche fundamental und unentbehrlich für das Entstehen der vorliegenden Arbeit war.

Ich danke Herrn Dr. Matthias Weigl vom Institut für Arbeits-, Sozial- und Umweltmedizin für seine Unterstützung im statistischen Teil der vorliegenden Studie und für seine wunderbare Gabe, scheinbar komplizierte Dinge einfach und verständlich auszudrücken.

Ich danke Herrn Sebastian Schumacher für seine graphischen Anregungen und für seine beständige Gelassenheit.

Nicht zuletzt danke ich meiner unvergleichlichen Familie, die mir - auch außerhalb der vorliegenden Arbeit – die Leichtigkeit und Lebensfreude schenkt, die ich brauche.