

## 3 Material und Methoden

### 3.1 Studiengruppe

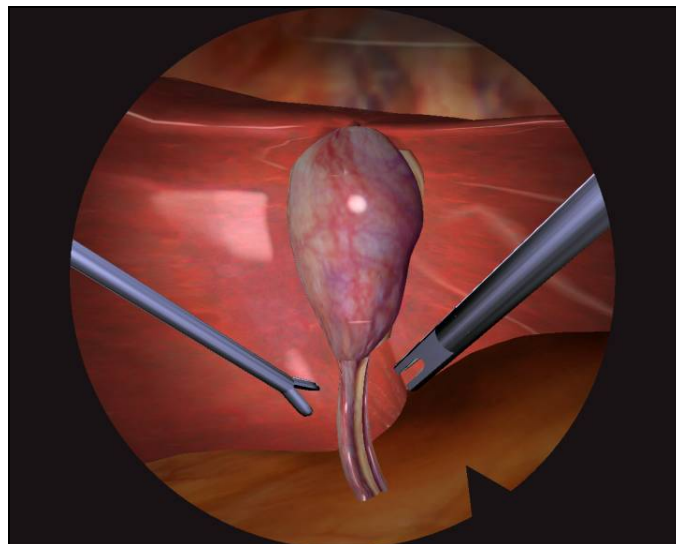
Die Studiengruppe bestand aus 17 Chirurgen unterschiedlicher Ausbildungsgrade einer chirurgischen Universitätsklinik, sowie einer Kontrollgruppe aus 13 Studenten ohne chirurgische bzw. laparoskopische Erfahrung.

Zu Beginn der Versuchsreihe beantwortete jeder Proband zunächst einen Fragebogen „0“, der sich nicht ändernde Eigenschaften über die Person erfasste, wie Geschlecht, bevorzugte Hand, Berufserfahrung usw.

Zum Zweck der Anonymisierung erhielt jede Testperson eine vierstellige Probandennummer.

### 3.2 Simulator

Der Virtual Endoscopic Surgery Trainer (VEST) wurde entwickelt, um möglichst realistisch minimalinvasive Operationen zu simulieren und es Chirurgen zu ermöglichen, das Manövrieren der Instrumente im dreidimensionalen Raum, den man nur über eine zweidimensionale Übertragung auf einem Bildschirm einsehen kann, zu trainieren (siehe Abb. 1).

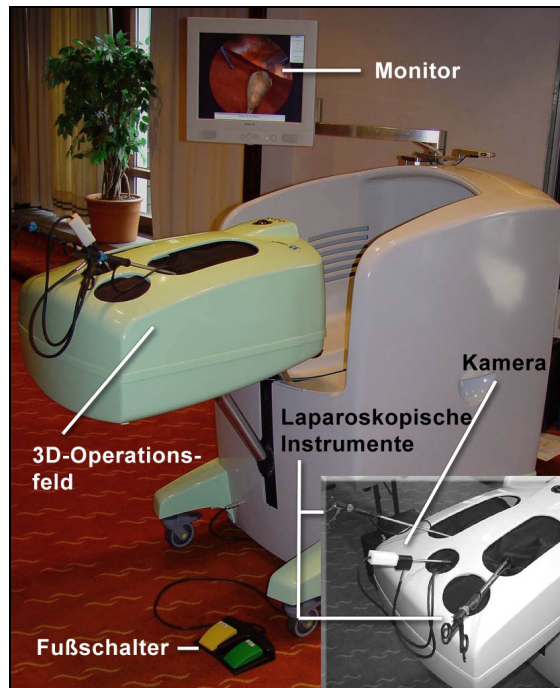


**Abb. 1:** OP-Simulation einer Cholezystektomie

Für dieses Studiendesign wurden Übungen verwendet, die bereits validiert und für die Erfassung von Trainingsfortschritten optimierten worden sind [5].

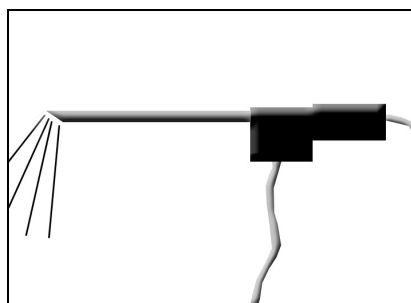
### 3.2.1 Aufbau und Funktionsprinzip

Die Hardware des Trainers befand sich in einem mobilem Gehäuse, auf dem ein Flachbildschirm und an dem ein höhenverstellbarer Korpus angebracht war, der das Operationsfeld simulierte (siehe Abb. 2).



**Abb. 2:** Virtueller OP-Simulator

Im Korpus befanden sich zwei laparoskopische Instrumente, jeweils für die nicht-dominante und für die dominante Hand sowie mittig eine Kamera mit simulierter 30° Winkeloptik (siehe Abb. 3).



**Abb. 3:** Laparoskopische Kamera mit 30° Optik

Alle Instrumente waren analog zu einem realistischen Operationsfeld beweglich. Vom System wurde ein dreidimensionaler Versuchsaufbau simuliert und dem Anwender auf dem Bildschirm als zweidimensionales Bild dargestellt.

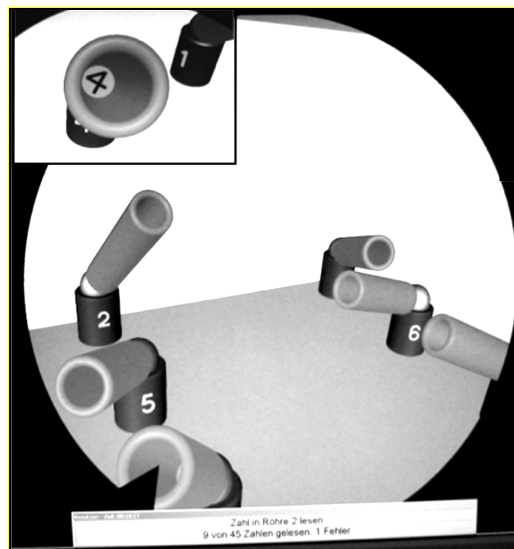
Ferner bot der Simulator ein realistisches motorisches Feedback. Wenn in der Simulation ein Gegenstand mit einem Instrument berührt wurde, so spürte der Anwender den Widerstand, es erfolgte also eine haptische Rückkopplung.

Weiterhin war ein Fußpedal vorhanden, welches mit dem Gehäuse in Verbindung stand und über welches ein Bestätigungssignal an das Programm gesendet werden konnte (siehe Kameraübung, 3.2.2).

Jeder Proband durchlief am Simulator zu jedem Trainings- und Testzeitpunkt jeweils drei Übungen: 1) die Kameraübung, 2) die Podesteübung und 3) die Kugelübung.

### 3.2.2 Kameraübung

Bei dieser Übung bestand die Aufgabe darin, mit der Kamera in Röhren hineinzuschauen, die im Innern auf ihrem Grund eine Zahl abgebildet hatten. Jede dieser Röhren stand in einem etwas anderen Winkel im simulierten dreidimensionalen Raum, so dass die Kamera in eine definierte Position gebracht werden musste, um die Zahl im Innern der Röhre vollständig erkennen zu können (siehe Abb. 4).



**Abb. 4:** Kameraübung: Mit der 30°-Optik müssen die Zahlen auf dem Grund der Röhren fokussiert werden (Einschub oben links)

Wenn dieses Ziel erreicht wurde, musste über das Fußpedal ein Bestätigungssignal an das Programm gesendet werden. Wurde die Zahl vollständig mit der Kamera fokussiert, ertönte ein Erfolgssignal, wenn nicht, so ertönte ein Geräusch, welches den Misserfolg signalisierte und vom System als Fehler erfasst wurde. Im letzteren Fall musste die Kameraposition so verbessert werden, dass die Zahl im Ganzen auf dem Bildschirm zu

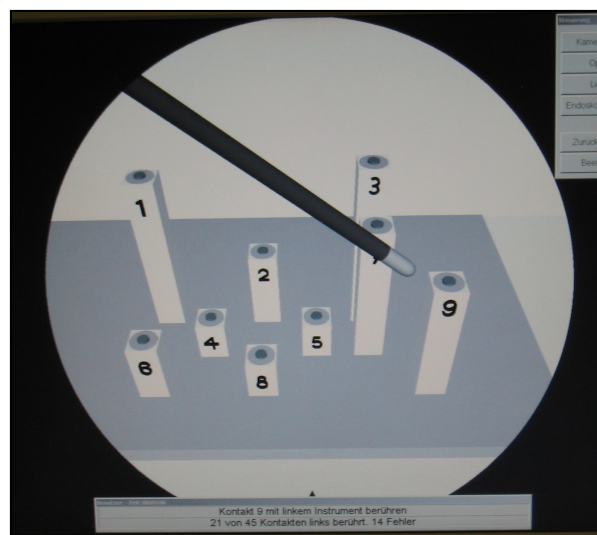
erkennen war. Glaubte der Proband, dieses Ziel erreicht zu haben, gab er wieder das Bestätigungszeichen über das Pedal. Wieder signalisierte ihm das Programm über ein Tonsignal den Erfolg oder Misserfolg.

Insgesamt mussten 45 Zahlen von neun unterschiedlich im Raum stehende Röhren fokussiert werden, um die Aufgabe vollständig zu erfüllen.

Die Messparameter dieser Übung waren 1.) die Zeit (als Parameter der Bearbeitungsgeschwindigkeit), 2.) der zurückgelegte Weg (als Maßzahl der Präzision) und 3.) die Fehlerzahl.

### 3.2.3 Podeste

In dieser Übung simulierte das Programm neun kubische Podeste unterschiedlicher Höhe. Auf jedem dieser Podeste war ein Kontakt angebracht, der mit den Instrumenten berührt werden musste, zunächst 45 mal mit der nicht-dominanten Hand, anschließend 45 mal mit der dominanten (siehe Abb. 5). Um den Kontakt war eine Fehlerzone angebracht. Wurde diese berührt, ertönte ein Geräusch, welches den Misserfolg signalisierte und vom System als Fehler erfasst wurde. Nur wenn der Kontakt genau mit der Instrumentspitze getroffen wurde, ertönte ein Erfolgssignal. Auf diese Weise wurde der Proband zur Präzision veranlasst.



**Abb. 5:** Podeste-Übung: Die vom Programm vorgegebenen 45 Kontakte mussten zunächst mit einem Instrument der nicht-dominanten, dann mit einem Instrument der dominanten Hand berührt werden, ohne dabei die Fehlerzone um den Kontaktpunkt herum zu berühren

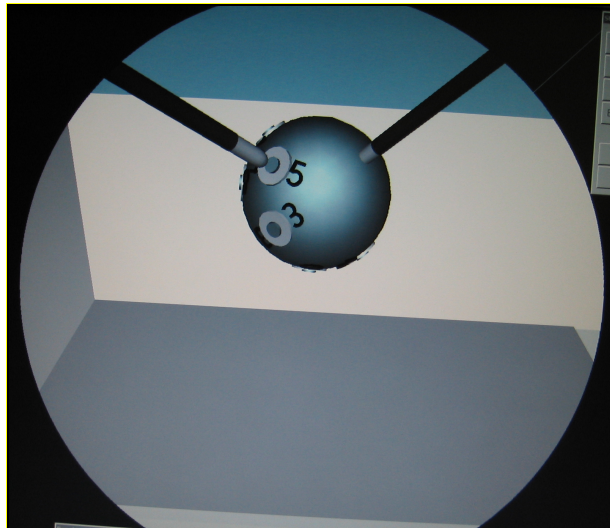
Die einzelnen Podeste waren nummeriert. Das Programm gab die Reihenfolge der zu berührenden Podeste vor. Die Aufgabe war erst dann abgeschlossen, wenn alle 90

Kontakte erfolgreich berührt wurden. (Ein Bestätigungssignal über das Pedal musste in dieser Übung, sowie auch in der Kugelübung nicht gegeben werden.)

Messparameter waren hier: 1.) die Zeit, 2.) der Weg für die nicht-dominante und dominante Hand und 3.) Fehler, ebenfalls für beide Hände getrennt.

### 3.2.4 Kugel

Diese Übung wurde beidhändig durchgeführt. Die dominante Hand hielt das Führungsinstrument, die andere hielt ein Instrument, an dessen Ende eine um 360 Grad drehbare Kugel angebracht war. Auf dieser befanden sich neun nummerierte Kontakte, die auch in dieser Übung von einer Fehlerzone umgeben waren (siehe Abb. 6). Wieder mussten 45 Kontakte in einer vom Programm vorgegebenen Reihenfolge berührt werden.



**Abb. 6:** Kugel-Übung: Die auf einer drehbaren Kugel angebrachten Kontakte mussten in der vom Programm vorgegebenen Reihenfolge mit einem Instrument der zweiten Hand berührt werden, ohne dabei die Fehlerzone um den Kontaktpunkt herum zu berühren

Dem Probanden wurden Erfolg und Misserfolg durch Tonsignale mitgeteilt. Gemessen wurden die Zeit zum Vollenden der Übung, der zurückgelegte Weg für jede Hand, sowie die Fehler.

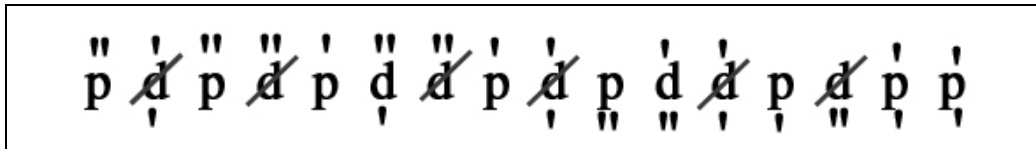
## 3.3 Psychometrische Tests

Die im Folgenden beschriebenen psychometrischen Tests gelten als hoch validiert und stabil. [6,7]

### 3.3.1 d2-Konzentrationsleistungstest

Beim d2-Durchstreichtest (d2-Test) handelt es sich um einen psychometrischen Test, der die Konzentrationsleistung erfasst. Seine Durchführung erfolgte standardisiert nach dem Test-Manual von Brickenkamp und Rolf, 9. Auflage [7].

Ziel dieses Tests ist es, innerhalb einer vorgegebenen Zeit für jede Testzeile alle Buchstaben „d“, die mit zwei Strichen versehen sind, durchzustreichen (siehe Abb. 7).



**Abb. 7: Testprinzip:** Alle Buchstaben „d“, die mit zwei Strichen versehen sind, müssen durchgestrichen werden. Alle „d“s mit mehr oder weniger als zwei Strichen und alle „p“s, unabhängig davon, mit wie vielen Strichen sie versehen sind, dürfen nicht durchgestrichen werden.

Zunächst erhielt der Proband einen Testbogen und einen flüssig schreibenden Kugelschreiber. Auf der Bogenvorderseite wurden die Probandennummer eingetragen sowie der Testzeitpunkt.

Ferner befand sich auf dieser Seite eine Übungszeile, die es der Testperson ermöglichen sollte, sich vor Beginn des eigentlichen Tests mit der Aufgabe vertraut zu machen. Diese wurde bearbeitet. War das Testprinzip verstanden, erhielt der Proband die Anweisungen zur Bearbeitung der Testbogenrückseite. Dann wurde das Blatt umgedreht und die Rückseite bearbeitet. Insgesamt wurden 14 verschiedene Zeilen bearbeitet, die aus jeweils 47 Zeichen bestanden. Die durchzustreichenden Zeichen galten als relevant, alle anderen als irrelevant. Das Verhältnis relevanter Zeichen zu irrelevanten Zeichen betrug 1:1,2

Zwei dem d2-Test zugehörige Schablonen ermöglichten die Auswertung. Schablone 1 diente der Identifikation nicht durchgestrichener relevanter Zeichen, also der so genannten Auslassungsfehler (Typ F1). Mit Hilfe der Schablone 2 konnten die durchgestrichenen nicht relevanten Zeichen einfach erfasst werden. Dabei handelte es sich um falsch durchgestrichene Buchstaben, demnach als Verwechslungsfehler bezeichnet (Typ F2).

Die Summe beider Fehlertypen ergab den Wert F, also die Gesamtfehleranzahl.

Schablone 1 ermöglichte ferner ein Auswerten des Wertes KL, der ein Maß für die tatsächlich erbrachte Konzentrationsleistung darstellte. Er setzte sich zusammen aus

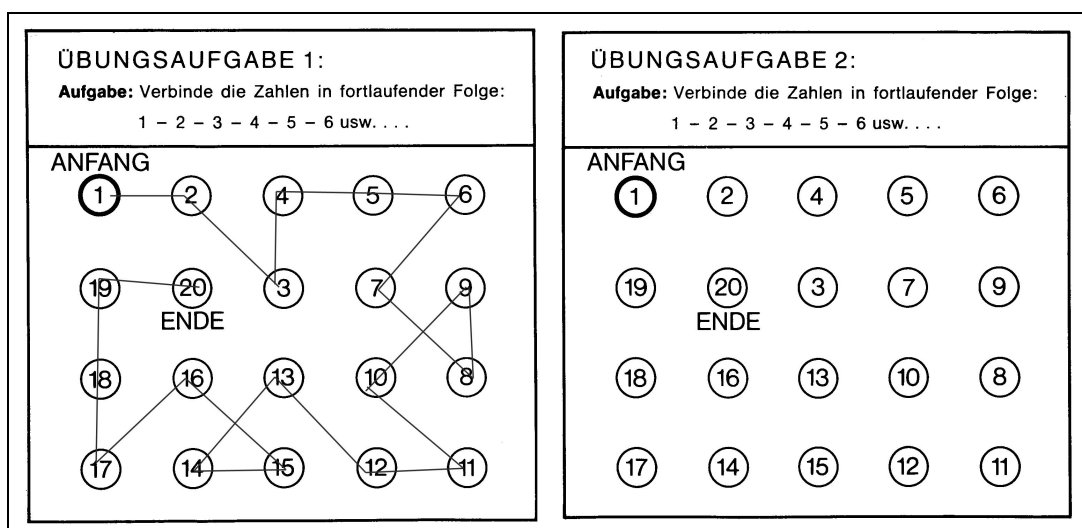
der Anzahl richtig durchgestrichener (relevanter) Zeichen abzüglich der Summe der gemachten Fehler vom Typ F2 (der Verwechslungsfehler).

### 3.3.2 Zahlenverbindungstest (ZVT)

Beim ZVT handelt es sich um einen Test zur Erfassung der Bearbeitungsgeschwindigkeit. Seine Durchführung erfolgte nach standardisierten Anweisungen (Oswald und Roth, der Zahlenverbindungstest, Handanweisung, 2. Auflage [6]).

Dieser Test bestand aus insgesamt vier Zahlenmatrizen (A-D) auf denen die Zahlen 1-90 jeweils unterschiedlich angeordnet waren. Aufgabe des Probanden war es, diese Zahlen in der richtigen Reihenfolge und so schnell wie möglich miteinander zu verbinden. Die als nächstes aufzusuchende Ziffer befand sich stets in unmittelbarer Nähe zur aktuell erreichten Zahl. In allen vier Matrizen war die zu bearbeitende Wegstrecke annähernd gleich groß.

Zunächst erhielt der Proband einen flüssig schreibenden Kugelschreiber, mit dem er auf einem Auswertungs-/Übungsbogen seine Probandennummer eintrug. Anschließend wurden zwei kleine Übungsmatrizen auf diesem Bogen von der Testperson bearbeitet, nachdem ihr die Versuchsaufgabe verständlich erläutert wurde (siehe Abb. 8).



**Abb. 8:** Zahlenverbindungstest – Testbogenvorderseite. Hier wurden die Probandennummer und der Testzeitpunkt eingetragen. Anhand der Übungsmatrizen konnte sich der Proband mit dem Testverfahren vertraut machen

Wurden die Übungsaufgaben korrekt durchgeführt, konnte der eigentliche Test beginnen, wobei alle vier Testbögen (von A bis D) in beliebiger Reihenfolge bearbeitet werden mussten. Die Bearbeitungszeit für jede einzelne Matrize wurde von der Versuchsleiterin mit einer Stoppuhr erfasst.

Zur Auswertung der Bearbeitungszeit wurde schließlich der Mittelwert aus allen vier gemessenen Zeiten gebildet.

### 3.4 Studiendesign

#### 3.4.1 Trainingsphase

Zunächst absolvierten die Probanden eine Trainingsphase an fünf aufeinander folgenden Tagen, um den Umgang mit dem Simulator zu trainieren. Frühere Studien mit OP-Simulatoren haben gezeigt, dass die Lernkurven der Testpersonen nach ein paar Tagen in ein Plateau übergehen, in dem kaum ein weiterer Lernzuwachs mehr stattfindet [5,8-12].

Mit der Trainingsphase sollte ein weiterer Lernzuwachs während der eigentlichen Testphase weitgehend vermieden und ein einheitliches Leistungsniveau aller Probanden erreicht werden (Trainingstag 1 – 5, im Folgenden T1-T5 genannt) (siehe Abb.9).



**Abb. 9: Schema Trainingsphase und Bereitschaftsdienstphase.** Trainingstag 1- 5, Testzeitpunkte Bereitschaftsdienstbeginn 8 Uhr (BD8), nach 8 Stunden Dienst (BD16), am Folgetag um 8 Uhr (FD8) und am Recovery-Day (nach normalem Nachtschlaf) um 8 Uhr (RD8)



### 3.4.2 Bereitschaftsdienstphase

Auf die Trainingsphase folgte innerhalb von zwei bis drei Tagen die Testphase (siehe Abb. 9) mit vier Testzeitpunkten während eines Bereitschaftsdienstes.

Der in dieser Studie untersuchte ärztliche Bereitschaftsdienst ist ein Dienst der Stufe D nach BAT (Bundesangestelltentarifvertrag). Ein regulärer Arbeitstag beginnt um 7.30 Uhr und endet um 16 Uhr. Danach beginnt der Bereitschaftsdienst. Dieser endet um 9 Uhr am Folgetag. Während des Bereitschaftsdienstes muss der Arzt im Krankenhaus bleiben. Er darf sich in dieser Zeit schlafen legen, wenn seine Dienste nicht benötigt werden. Schlaf ist jedoch nicht garantiert.

Zeitlicher Ausgangspunkt der Testphase war der Bereitschaftsdienstbeginn um etwa 8 Uhr (BD8). Zweiter Testzeitpunkt war um 16 Uhr (BD16), also nach acht Stunden Arbeitszeit. Der dritte Testzeitpunkt fand am Folgetag um 8 Uhr statt (FD8), also nach 24 Stunden Bereitschaftsdienst in der Regel mit deutlichem Schlafentzug. Der vierte und damit letzte Test fand einen Tag nach dem Dienst, am so genannten Recovery Day statt, um 8 Uhr morgens in der Regel nach einer Nacht mit Normalschlaf ( $\geq 6$  h) (RD8).

Die Kontrollgruppe (Studenten) durchlief ebenfalls die fünftägige Trainingsphase sowie, analog zum Bereitschaftsdienst, die gleichen vier Testzeitpunkte allerdings nicht unter Dienstbedingungen und ohne Schlafentzug.

Jede Versuchsperson führte zu jedem Testzeitpunkt drei Übungen am Operationssimulator durch. Unmittelbar im Anschluss wurden die beiden psychometrischen Tests durchgeführt, der d2-Durchstreichetest, sowie der Zahlenverbindungstest (ZVT).

Die psychometrischen Tests gelten als nur geringfügig trainierbar und stabil [6,7]. So wurden diese nur am ersten Trainingstag einmal durchgeführt, um die Probanden mit dem Procedere vertraut zu machen, und zu allen vier Testzeitpunkten des Bereitschaftsdienstes. Die Simulatorübungen hat jede Testperson zu allen Trainings- und Testzeitpunkten durchlaufen.

Zusätzlich machte der Proband vor den Übungen und Tests auf einem Fragebogen Angaben über seinen Nachtschlaf und beantwortete subjektive Fragen zu seinem Befinden bzw. seiner Einschätzung über seine aktuelle Konzentrationsfähigkeit. Die Fragebögen wurden nach den Testzeitpunkten BD8, BD16, FD8 und RD8 benannt.

### 3.4.3 Cross-Over-Design

Um zu untersuchen, ob die Zeitpunkte BD8 und BD16 den Übungseffekt in die Testphase hinein fortführen, wurde die Gruppe der Ärzte randomisiert aufgeteilt in die zwei gleich großen Gruppen „Ärzte-Normal“ und „Ärzte-Cross-Over“.

Probanden der „Ärzte-Normal“-Gruppe wurden in der bereits beschriebenen Reihenfolge getestet: BD8, BD16, FD8 und RD8. Die Testteilnehmer der „Cross-Over“-Gruppe wurden zunächst zum Zeitpunkt FD8 getestet, also an einem Morgen nach einem Bereitschaftsdienst, dann zu RD8 und schließlich, bei einem erneuten Bereitschaftsdienst zu den Zeitpunkten BD8 und BD16.

Anschließend wurde statistisch untersucht, ob signifikante Leistungsunterschiede zwischen den beiden Gruppen bestanden.

### 3.5 Fragebögen

Der „Fragebogen 0“ diente dazu, allgemeine und über den Versuchszeitraum sich nicht ändernde Informationen über die Testperson einzuholen.

Die Fragebögen BD8, BD16, FD8 und RD8 waren zu den jeweiligen Testzeitpunkten auszufüllen. Die erfragten Parameter sind Tabelle 1 zu entnehmen:

**Tabelle 1: Übersicht der Fragebögen, erfasste Parameter**

Fragebogen	Erfasste Parameter
0	Alter, Geschlecht, dominante Hand, *Berufsjahre in einem chirurgischen Fach, Ausbildungsstand, *Fachrichtung, *Anzahl eigenständiger laparoskopischer Eingriffe, Erfahrung im Umgang mit Computern, Erfahrung im Umgang mit Videospiele, Erfahrung mit OP-Simulatoren
BD8	* „heute im Dienst seit...“, Angaben zum Nachtschlaf, Angaben zur subjektiven Leistungseinschätzung, zur Stimmung und zum gesundheitlichen Wohlbefinden
BD16	Wie BD8 + *zwei Zusatzfragen zur aktuellen subjektiven Leistungseinschätzung im Vergleich zum Zeitpunkt BD8
FD8	Wie BD8 + *zwei Zusatzfragen zur aktuellen subjektiven Leistungseinschätzung im Vergleich zum Zeitpunkt BD8
RD8	Wie BD8 + *zwei Zusatzfragen zur aktuellen subjektiven Leistungseinschätzung im Vergleich zum Zeitpunkt BD8
* = von den Studenten nicht zu beantworten	

### 3.6 Statistische Methoden

Die gemessenen Einzelparameter aus den Simulatorübungen und den psychometrischen Tests wurden zunächst tabellarisch erfasst. Für die statistische Analyse wurde das Programm SPSS® für Windows, Version 11.5 verwendet.

Die Auswertung der Parameter erfolgte mit dem Mann-Whitney-Test und dem Wilcoxon-Test.

Der Mann-Whitney-Test ist ein Signifikanztest zur Überprüfung einer Hypothese, die besagt, dass zwei unabhängige Stichproben aus derselben Grundgesamtheit stammen. Dieser Test wurde demnach für sämtliche Vergleiche zwischen den unterschiedlichen Probandengruppen bzw. Subgruppen herangezogen (z.B. Vergleich Ärzte gegen Studenten).

Der Wilcoxon-Test ist ein nichtparametrischer Test für verbundene Stichproben. Er fand Anwendung bei Vergleichen innerhalb einer Gruppe, z.B. in der Gruppe Ärzte, Vergleich der Leistungen BD8 zu BD16, FD8 und RD8.

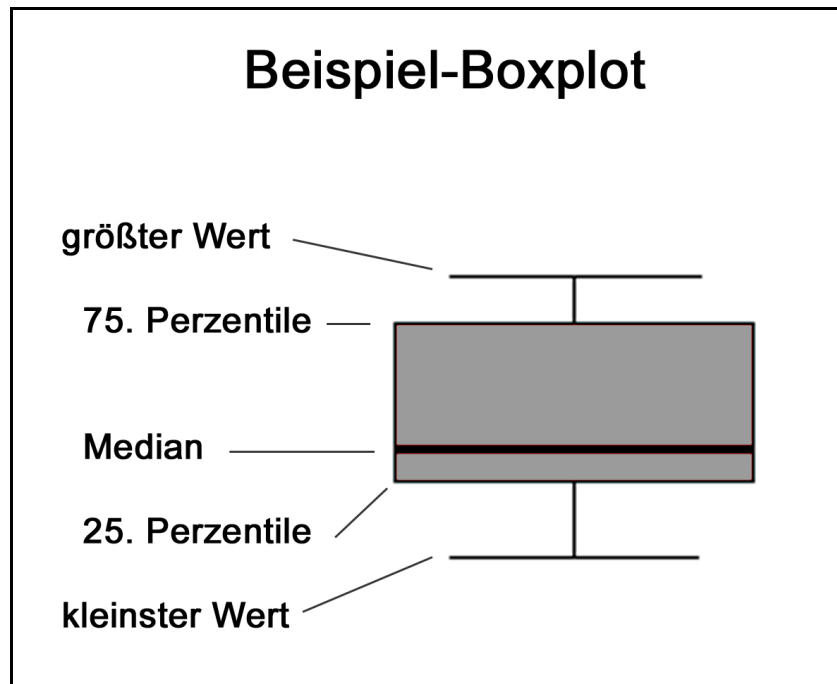
Im Folgenden ist die Nomenklatur des p-Werts erklärt:

- $p > 0,05$ : nicht signifikant, Symbol : ns
- $p \leq 0,05$ : signifikant, Symbol : \*
- $p \leq 0,01$ : sehr signifikant, Symbol: \*\*
- $p \leq 0,001$ : höchst signifikant, Symbol: \*\*\*

Um die Auswertung der psychomotorischen Leistungen am Simulator überschaubarer zu machen, wurden die einzelnen Variablen (Einzelparameter) mittels einer Hauptkomponentenanalyse zu Gesamtscores zusammengefasst. Diese Scores wurden so konstruiert, dass sie die maximale Varianz der Einzelvariablen erklären und untereinander unabhängig sind:

- 1) Gesamtscore 1: fasste alle Parameter des Weges und der Zeit zusammen
- 2) Gesamtscore 2: fasste alle Fehler zusammen

Grafisch wurden die Ergebnisse in Form von Boxplots dargestellt (siehe Abb 10). Der Boxplot präsentiert die statistische Verteilung metrischer Daten. Er stellt den Median dar (50. Perzentile), sowie die 25. und die 75. Perzentile und die Extremwerte.



**Abb. 10:** Der Boxplot stellt die statistische Verteilung der Parameter dar. Ablesbar sind die Extremwerte, der Median sowie die 25. und die 75. Perzentile.

### 3.6.1 Statistische Analysen

Folgende Analysen wurden durchgeführt:

1. Vergleich Studenten gegen Ärzte bzgl. der Trainingsphase und der Bereitschaftsdienstphase

- a) anhand der Einzelparameter
- b) anhand der Gesamtscores

Anschließend erfolgte eine Einteilung der Probanden in die Subgruppen

a) Erfahrung:

- Anfänger (Studenten, keine Erfahrung)
- Lernende (Ärzte,  $\leq 10$  eigenständige Laparoskopien)
- Experten (Ärzte,  $\geq 21$  eigenständige Laparoskopien)

b) Schlaf:

- Normalschlaf (Studenten,  $\geq 6$  h Schlaf zwischen BD8 und FD8)
- $< 3$  h Schlaf (Ärzte,  $< 3$  h Schlaf zwischen BD8 und FD8)
- $\geq 3$  h Schlaf (Ärzte,  $\geq 3$  h bis  $< 6$  h Schlaf zwischen BD8 und FD8)

2. Vergleich Anfänger gegen Lernende gegen Experten bezüglich der Trainingsphase und der Bereitschaftsdienstphase anhand der Gesamtscores
3. Vergleich Normalschlaf gegen  $< 3$  h Schlaf gegen  $\geq 3$  h Schlaf bzgl. der Trainingsphase und der Bereitschaftsdienstphase anhand der Gesamtscores
4. BD-Analyse: Vergleich der Leistungen innerhalb einer Gruppe von BD8 zu BD16, BD8 zu FD8 und BD8 zu RD8 für
  - a) Studenten = Anfänger = Normalschlaf
  - b) Ärzte
  - c) Lernende
  - d) Experten
  - e) Ärzte  $< 3$  h Schlaf
  - f) Ärzte  $\geq 3$  h Schlaf