Institut für Klinische und Experimentelle Trauma-Immunologie - Universität Ulm

Ärztlicher Direktor: Prof. Dr. med. Markus S. Huber-Lang

Urteiler*innenübereinstimmung von Tutor*innen und Ärzt*innen bei Skills Lab-Prüfungen

Dissertation zur Erlangung des Doktorgrades der Medizin der Medizinischen Fakultät der Universität Ulm

Verfassername: Lukas Keßling

Geburtsort: München

Jahr: 2021

Amtierender Dekan: Prof. Dr. T. Wirth

1. Berichterstatter: PD Dr. D. Messerer

2. Berichterstatter: Prof. Dr. A. Böckers

Tag der Promotion: 16.12.2022

Inhaltsverzeichnis

Inhaltsverzeichnis I

Abkürzungsverzeichnis III

1 Einleitung 1

1.1 Studium der Humanmedizin 1

1.2 Skills Lab 1

1.3 Prüfungsformate in der medizinischen Ausbildung 2

1.4 Peer Teaching 3

1.5 Peer Assessment 4

1.6 Fragestellungen und Arbeitsschritte 5

2 Material und Methoden 6

2.1 Skills Lab der Universität Ulm 6

2.2 Prüferstudie 10

2.3 Aufbau, Methodik und Durchführung der Online-Praxis-Check-Simulation 11

2.4 Statistische Auswertung 18

3 Ergebnisse 20

3.1 Analyse der Praxis-Checks mit einem*r Prüfer*in 20

3.2 Bewertungen von Haupt- und Co-Prüfer*innen bei realen Praxis-Checks 26

3.3 Vergleich der Bewertung mehrerer Prüfer*innengruppen durch Online-Praxis-Check-Simulation 31

4 Diskussion 42

4.1 Urteiler*innenübereinstimmung bei Praxis-Checks mit einem*r Prüfer*in 42

4.2 Urteiler*innenübereinstimmung von Haupt- und Co-Prüfer*innen bei realen Praxis-Checks 43

4.3 Urteiler*innenübereinstimmung mehrerer Prüfer*innengruppen durch Online-Praxis-Check-Simulation 45

4.4 Prüfsystem Praxis-Check 53
4.5 Evaluation der Fragestellungen 54
4.6 Weitere Fragestellungen 54
5 Zusammenfassung 55
6 Literaturverzeichnis 56
7 Anhang 63
8 Danksagung 106
9 Lebenslauf 107

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ÄApprO	ärztlichen Approbationsordnung
CP	Co-Prüfer*in
CP-A	Co-Prüfer*in Ärzt*in
CP-T	Co-Prüfer*in Tutor*in
HP	Hauptprüfer*in
IQR	Interquartilsabstand
n	Gesamtanzahl der Untersuchungen
NKLM	Nationaler Kompetenzbasierten Lernzielkatalog Medizin
OSCE	Objective Structured Clinical Examination (Objektive Strukturierte Klinische Prüfung)
PB	Prüfungsbogen
PC	Praxis-Check
PJ	Praktisches Jahr
POL	Problemorientiertes Lernen
SoSe	Sommersemester
SP	Schauspielpatient*in
TC	Theorie-Check
WiSe	Wintersemester

1 Einleitung

1.1 Studium der Humanmedizin

Studierende der Humanmedizin in Deutschland durchlaufen im Rahmen ihres Studiums einen vorklinischen Teil (zwei Jahre), einen klinischen Teil (drei Jahre) und ein praktisches Jahr. In dieser Zeit werden den Studierenden Wissen und praktische Fertigkeiten vermittelt. Die dafür geltende rechtliche Grundlage ist die Approbationsordnung für Ärzt*innen (ÄApprO) [9]. Nach verstärkter Kritik über zu geringe Vermittlung praktischer Fertigkeiten im Studium wurde im Zuge des „Masterplans Medizin 2020“ unter anderem beschlossen, das Humanmedizinstudium praktischer zu gestalten [44]. Von Studierenden „in der Ausbildung soll frühzeitig die konsequente Orientierung am Patienten und seinen Bedürfnissen erlernt und eingeübt werden“ [24]. Die praktischen Fertigkeiten werden vom Nationalen Kompetenzbasierten Lernzielkatalog Medizin (NKLM) klar definiert [40].

1.2 Skills Lab

Alle Universitäten stellen ihren Studierenden ein Skills Lab zur Verfügung [7]. Da die Ausbildung angehender Mediziner*innen in praktischen Fertigkeiten einen immer höheren Stellenwert einnimmt, ist auch das Angebot an Trainingsmöglichkeiten weiter gestiegen [39; 44]. Diese Skills Labs sind im Regelfall speziell ausgestattete Räumlichkeiten, in denen Studierende praktische Fertigkeiten trainieren können. Dafür stehen medizinische Instrumentarien, Phantompuppen und zum Teil Schauspielpatient*innen zur Verfügung [39]. So können Studierende beispielweise möglichst wirklichkeitsnahe Patient*innen-Ärzt*innen-Situationen simulieren, um sich auf zukünftige Begegnungen in der Rolle des*der Ärzt*in vorzubereiten. Praktische ärztliche Fertigkeiten können so in einem sicheren und geschützten Rahmen von den Studierenden erlernt werden [18; 47]. Das Lehrangebot ist oft breit gefächert und reicht je nach Skills Lab von peripher venöser Blutentnahme über abdominelle Untersuchungen bis hin zur endotrachealen Intubation oder Anlage einer Thoraxdrainage [37; 39]. Das Angebot hängt von den zur Verfügung stehenden Materialien und Räumen ab und, viel wichtiger, ob eine in der Fertigkeit ausgebildete Person bereit steht, um das Wissen weiterzugeben [47]. Oft wird diese Aufgabe von studentischen Tutor*innen übernommen, welche unter ärztlicher Supervision stehen [7].

1.3 Prüfungsformate in der medizinischen Ausbildung

In der medizinischen Ausbildung gibt es verschiedene Prüfungsformate. In einem Großteil der Prüfungen wird Wissen in Form einer schriftlichen Klausur mit Multiple-Choice-Fragen abgefragt [3; 12; 34; 37]. Gelegentlich werden Studierende auch mündlich oder in Form von einer „**O**bjektive **S**tructured **C**linical **E**xamination“ (OSCE) geprüft [12; 54]. So standen beispielsweise an der Universität Ulm im klinischen Studienabschnitt im Wintersemester 2019/20 (WiSe 19/20) 48 Multiple-Choice-Fragen-Prüfungen sieben praktischen Prüfungen gegenüber. Die ärztlichen Basisfertigkeiten wurden im NKLM zusammengefasst. Dieser wurde unter Beachtung der gesetzlichen Vorgaben der ÄApprO erstellt [40]. Der NKLM definiert das Absolventenprofil von angehenden Ärzt*innen, welches in ihrem Studium erlangt werden soll [40]. So hilft der NKLM die praxisnahe Ausbildung der Studierenden zu vereinheitlichen und weiter zu fördern [40]. Inhalte für praktische Prüfungen und Skills Labs werden idealerweise unter Beachtung des NKLM erstellt [11; 32].

Um Prüfungsinhalte hinsichtlich des Kompetenzniveaus zu kategorisieren, wurde von Miller 1990 ein Pyramidenmodell entwickelt, in dem die Kompetenz der Studierenden in verschiedenen Stufen dargestellt wird [36] (Abbildung 1). Das faktische Wissen „Weiß es“ bildet die Grundlage der Pyramide. Aufbauend auf dem faktischen Wissen steht das angewandte Wissen „Weiß wie“. Umsetzung von Wissen in praktische Fertigkeiten wird von der Stufe „Zeigt es“ dargestellt, gefolgt von „Macht es“ als repräsentative Stufe für die reale regelmäßige Anwendung im ärztlichen Alltag.

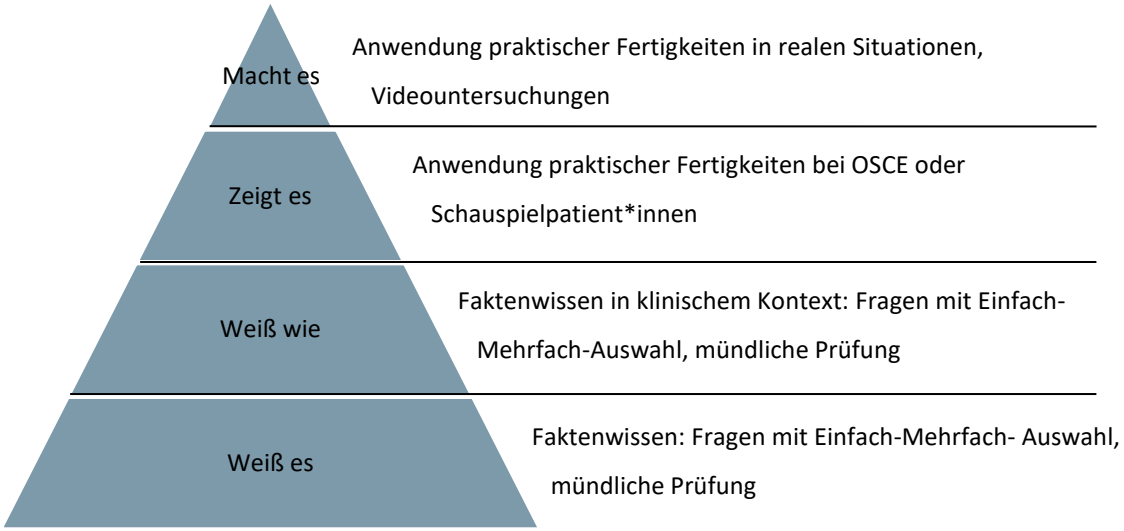


Abbildung 1: Darstellung der Kompetenzstufen mit möglicher Prüfungsform nach Miller 1990 [36]

Schriftliche Prüfungen werden verwendet, um die unteren zwei Stufen der Pyramide abzufragen: faktisches Wissen und theoretische Anwendung [12; 25; 40]. Eine praktische Fertigkeit mit einer schriftlichen Klausur zu beurteilen, ist aber nur eingeschränkt möglich [48]. Um die Stufe „Zeigt es“, die praktische Anwendung, zu prüfen, wurde unter anderem die OSCE entwickelt. Praktische Prüfungen in Form von OSCEs nehmen in der Ausbildung von Medizinstudierenden einen immer größeren Stellenwert ein [33; 45]. OSCEs sind ein wichtiges Beurteilungssystem für die Qualität und Entwicklung praktischer Fertigkeiten von Studierenden, welche von herkömmlichen Prüfungsmethoden, beispielsweise Multiple-Choice-Tests, nicht erfasst werden können [56]. Praktische Prüfungen sollten die Kriterien Validität, Reliabilität und Durchführbarkeit erfüllen [22].

1.3.1 **Prüfungsform OSCE**

Die OSCE ist eine mittlerweile weit verbreitete Prüfungsform für praktische Fertigkeiten [23; 30; 49; 67]. Oft besteht die OSCE aus mehreren unabhängigen Stationen, welche von den Teilnehmenden nacheinander durchlaufen werden [12]. Für jede Station gibt es eine festgelegte Zeitdauer, in der eine bestimmte Aufgabe zu erfüllen ist [63]. Die Varietät der Prüfungsaufgaben ist aufgrund der Anzahl möglicher Fälle sehr groß. So kann dieses Prüfungsformat prinzipiell in jeder Fachrichtung eingesetzt werden. Oft werden an den Stationen Phantompuppen, Bilder, Schauspielpatient*innen oder klinische Befunde zur Interpretation herangezogen. Bewertet werden die Studierenden im Regelfall von Ärzt*innen oder geschulten studentischen Tutor*innen [50; 51]. Die Leistung wird mit Hilfe eines vorab erstellten, klar definierten Kriterienkataloges bewertet, welcher den geprüften Studierenden im Regelfall nicht bekannt ist [4; 28; 51]. Die Vergleichbarkeit einer Prüfung wird durch Strukturierung und Standardisierung verbessert [23]. Auch das Lernverhalten der Studierenden wird durch die praktische Prüfungsform einer OSCE beeinflusst, sodass Inhalte zielgerichteter erlernt werden können [16].

1.4 Peer Teaching

Unter Peer Teaching wird im Kontext des Medizinstudiums die Lehre von Studierenden für Studierende verstanden [20; 53; 59]. Studierende übernehmen die Rolle des Lehrenden und vermitteln als Tutor*in bereits erworbenes Wissen an Kommiliton*innen weiter [58; 65]. Vor allem praktische Fertigkeiten können mit diesem System gut vermittelt werden,

Tutor*innen werden oft in Seminaren oder Skills Labs eingesetzt [5; 54]. Die Vorteile sind sowohl für die Tutor*innen als auch für die Studierenden vielfältig [7]. Studentische Tutor*innen können die aktuelle Lernsituation von Studierenden nachvollziehen und so Lerninhalte aus studentischer Sicht vermitteln [1; 66]: Tendenziell gibt es eine hohe kognitive und soziale Kongruenz [8]. Lerndefizite können so gut aufgedeckt und behoben werden. Gleichzeitig können Tutor*innen ihr Wissen und Können beim aktiven Lehren vertiefen und verfestigen [62]. So wird sowohl deren fachliche als auch soziale Kompetenz gefördert [14; 26]. Wichtig ist dabei, dass die studentischen Tutor*innen gut und ausreichend ausgebildet wurden [66]. Eine gute Ausbildung sichert die richtige Wissensvermittlung und hilft, Fehler der Studierenden zu finden und diese kompetent zu korrigieren. So ist das Skills Lab ein geeigneter Ort, um Peer Teaching einzusetzen [5].

1.5 Peer Assessment

Peer Assessment bedeutet, dass studentische Tutor*innen die Leistungen von Studierenden bewerten [2; 19; 41; 59]. Oft wird Peer Assessment, ähnlich zu Peer Teaching, in Skills Labs oder OSCEs eingesetzt [52]. Studien von Chenot et al. 2007 und Koch 2008 zeigten, dass Studierende praktische Prüfungen reliabel bewerten können [13; 28]. Wichtig ist dabei, dass die Tutor*innen gut ausgebildet wurden, um eine qualitative Bewertung der praktischen Leistung der Studierenden abzugeben [19]. Studentische Tutor*innen trauen sich in der Regel zu, die Leistung von Studierenden zu bewerten [13; 15]. Die Akzeptanz der Studierenden bei Peer Assessment ist hoch [13; 28].

1.6 Fragestellungen und Arbeitsschritte

Im Skills Lab der Universität Ulm wurde in den letzten Jahren eine neuartige Prüfungsform eingeführt. Praktische Fertigkeiten der Studierenden werden mit Hilfe von Praxis-Checks (PCs) überprüft. Als Prüfer*innen werden ausgebildete studentische Tutor*innen eingesetzt. Im Hinblick auf steigende Zahlen praktischer Prüfungen wäre es von Vorteil zu wissen, ob studentische Tutor*innen ähnlich treffsichere Bewertungen wie ärztliche Prüfer*innen in dieser Prüfungsform geben.

Die Kernfrage dieser Dissertation war also: Sind geschulte studentische Prüfende in der Bewertung praktischer Fertigkeiten in Praxis-Checks ähnlich treffsicher wie ärztliche Prüfende? Des Weiteren wurde untersucht, ob und wie gut diese Prüfer*innengruppen miteinander übereinstimmen, ob bestimmte Punkte Prüfer*innen eher Schwierigkeiten bereiten und inwiefern Tutor*innen als eigenständige Prüfer*innen für curriculare PC-Prüfungen eingesetzt werden könnten.

Die Fragestellungen wurden in dieser Dissertation mit Hilfe folgender Arbeitsschritte untersucht:

Arbeitsschritt 1: PCs mit einem*r Prüfer*in

Im ersten Arbeitsschritt wurden zunächst allgemein erfasste Daten von PCs mit einem*r Prüfer*in analysiert und ausgewertet.

Arbeitsschritt 2: PCs mit je zwei Prüfer*innen

Im zweiten Arbeitsschritt wurden Bewertungen von jeweils zwei Prüfer*innen (Haupt- und Co-Prüfer*in) bei realen PCs verglichen.

Arbeitsschritt 3: Online-PC-Simulation mit mehreren Prüfer*innengruppen

Im dritten Arbeitsschritt wurden die Bewertungen mehrerer Prüfer*innengruppen (Studierende, Tutor*innen und Ärzt*innen) durch eine Online-PC-Simulation hinsichtlich ihrer Qualität und Vergleichbarkeit analysiert. Diese Online-PC-Simulation ist eine Videostudie, die im SoSe 2020 entwickelt und durchgeführt wurde. Die Online-PC-Simulation ermöglichte einen direkten Vergleich der Bewertungen mehrerer Prüfer*innen durch Verwendung von einheitlichem Prüfmaterial in Form von vier Filmen eines PC.

2 Material und Methoden

2.1 Skills Lab der Universität Ulm

Das Skills Lab der Universität Ulm stand im Untersuchungszeitraum (WiSe 18/19 – WiSe 19/20, vor der Covid-19 Pandemie) für alle Studierende der Humanmedizin an den von Tutor*innen betreuten Übungsnachmittagen (im Regelfall Montag bis Donnerstag 16 - 20 Uhr während der Vorlesungszeit) zur freien Verfügung. Die Räumlichkeiten des Skills Lab boten mit diversen Trainingsmöglichkeiten allen Studierenden eine gesicherte Umgebung, um praktische Fertigkeiten möglichst stressfrei erlernen zu können. Es war möglich, entweder selbstständig oder unter Anleitung studentischer Tutor*innen die erwünschten Fertigkeiten zu trainieren. Während den Übungsnachmittagen standen immer mindestens zwei Tutor*innen für die Vermittlung, Supervision und Feedback zur Verfügung. Um die Übungen so real wie möglich zu gestalten, wurden echte medizinische Materialien verwendet, beispielsweise Ultraschallgeräte und Dopplerstiftsonden oder auch täuschend echte Phantompuppen. Viele der Fertigkeiten konnten und sollten auch direkt an Kommiliton*innen geübt werden, auch unter dem Aspekt, die Patient*innenperspektive einzunehmen. Andere, beispielsweise eine Hautnaht, wurden an speziellen künstlichen Häuten erlernt. Insgesamt standen den Studierenden über 20 vorbereitete Übungsstationen zur Verfügung.

2.1.1 Studentische Tutor*innen und deren Ausbildung

Unter ärztlicher Leitung standen im Skills Lab der Universität Ulm mehrere, ca. 20, geschulte studentische Hilfskräfte zur Verfügung. Um die Studierenden in ihrem Lernprozess optimal zu unterstützen, wurde von den Tutor*innen erwartet, dass diese sich bereits im klinischen Abschnitt ihres Studiums befanden und Interesse an der Lehre von Kommiliton*innen bestand. Tutor*innen des Skills Lab durchliefen in der Regel eine siebenstufige Ausbildung (Abbildung 2) [31].

Angehende Tutor*innen erwarben in vorklinischen Tutoraten erste Erfahrungen als Peer Teacher und nahmen meist zusätzlich am Lehren-Lernen-Studientrack der Universität Ulm teil. In diesem großen Wahlfach mit 28 Stunden wurden medizindidaktische Grundlagen zu Lernzielen, Unterrichtsplanung, Lehrmethoden, Gruppendynamik und Feedback vermittelt [60]. Parallel erwarben die angehenden Tutor*innen bei Skills Lab-Übungsabenden

praktische Fertigkeiten, welche als Teilnehmende im PC abgelegt wurden. Nach Aufnahme in das Skills Lab-Tutor*innen-Team durchliefen die Tutor*innen am Anfang jedes Semesters ein Training, in dem vor allem neu angebotene Übungsstationen besprochen und vorbereitet wurden. Neue Tutor*innen begannen dann unter Aufsicht erfahrener Tutor*innen erste Lehr- und Prüfungstätigkeiten. Als ausgebildete Tutor*innen wurden eigenständige Lehr- und Prüfungstätigkeiten in extracurricularen Bereichen übernommen.

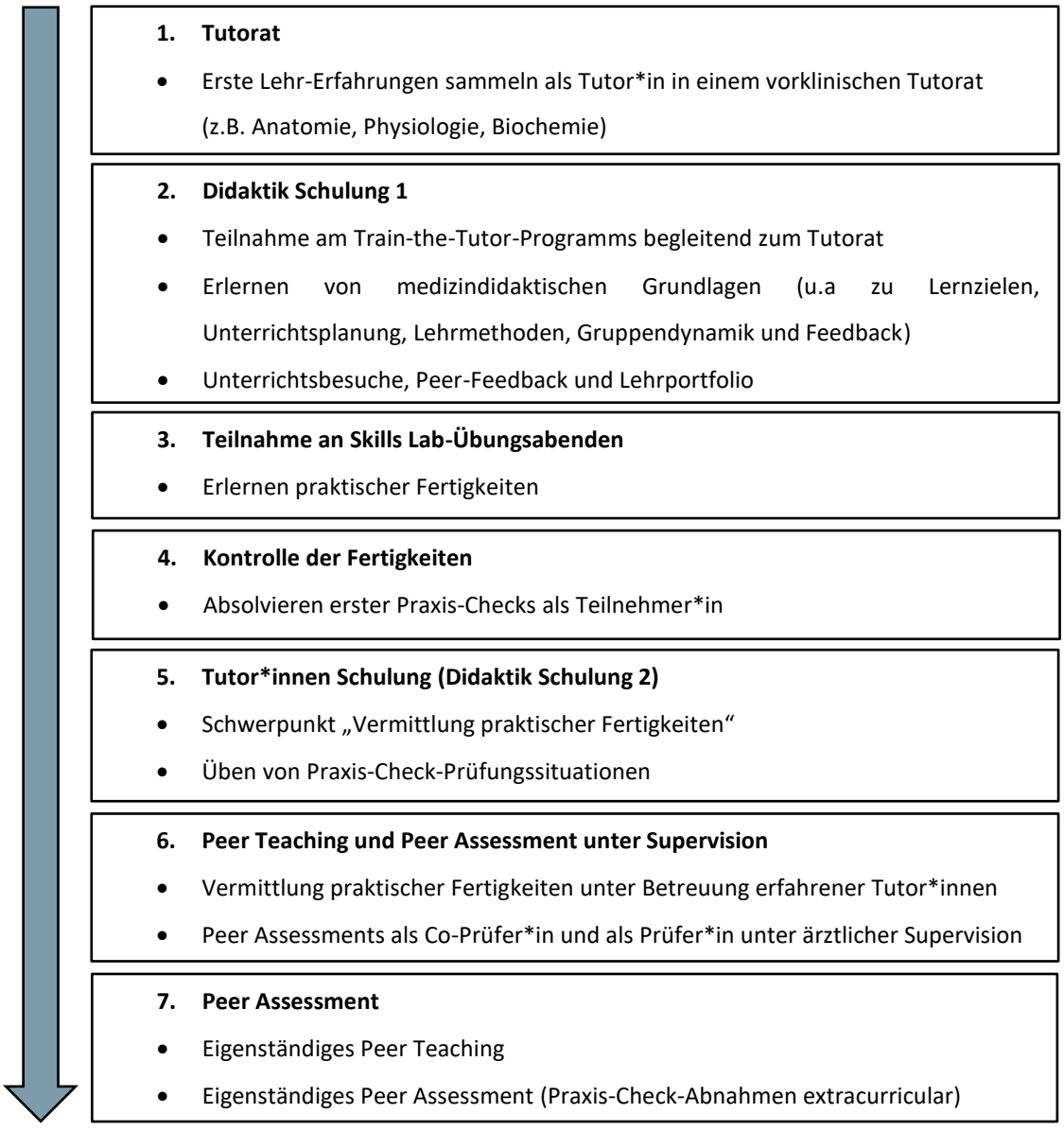


Abbildung 2: Flussdiagramm der Ausbildung der Skills Lab-Tutor*innen [52]

2.1.2 **Theorie-Check, Praxis-Check**

Zur Kontrolle ihres Lernfortschrittes konnten Studierende ihr theoretisches Wissen und ihre praktischen Fertigkeiten in den Theorie-Checks (TCs) und Praxis-Checks (PCs) unter Beweis stellen. Sowohl die TCs als auch die PCs jeder einzelnen Übungsstation wurden unter Beachtung des NKLM entwickelt. In einer Testphase, bestehend aus zwei Semestern (WiSe 16/17 und SoSe 17), wurden die Anforderungen der PCs überprüft und regelmäßig verbessert. Zusätzlich wurde jede PC-Station fachärztlich validiert und freigegeben, um die Qualität, Aktualität und Richtigkeit zu gewährleisten. In den PCs wurden verschiedene Punktekategorien geprüft. Einerseits gab es Skillpunkte, welche sich auf die korrekte Ausführung einer Tätigkeit bezogen, beispielsweise die Auskultation des Abdomens in allen vier Quadranten. Andererseits gab es Kontextpunkte, die wiederum in Hygiene-, Dokumentations- und Kommunikationspunkte unterteilt wurden. Die Kontextpunkte überprüften vor allem das richtige Verhalten während der Durchführung der einzelnen Maßnahmen, beispielsweise hygienische Händedesinfektion vor und nach Patientenkontakt. Um praxisnahe Handlungsabläufe einzuüben enthielten viele PCs beide Punktekategorien, sogenannte Kontext-Skill-PCs, z.B. PC Untersuchung Abdomen, andere waren reine Skill-PCs. Die Bewertungskriterien waren für die Studierenden bereits während des Trainings und vor und nach der Prüfung über die Lernplattform Moodle frei zugänglich und klar definiert. Die Bestehensgrenze lag bei 80 % oder 16 von 20 Punkten.

Mit Hilfe der Tutor*innen konnten sich die Studierenden gezielt auf die im PC abgefragten Fertigkeiten vorbereiten. Entschieden sich Studierende nach ausreichender Vorbereitung für den PC anzumelden, wurde ein fixer Prüfungstermin vereinbart. Der TC fungierte in diesem Fall als Zulassungsbeschränkung für den PC. Im TC wurden Studierende über ihr theoretisches Hintergrundwissen, welches für die praktische Fertigkeit benötigt wurde, abgefragt. Nach bestandenem TC konnten sich die Studierenden für den PC anmelden. Am Prüfungstag wurden von den Tutor*innen die für den PC benötigten Materialien bereitgestellt. Die Prüfung begann, nachdem der*die Tutor*in die erste Anweisung vorgelesen hatte und die Stoppuhr mit der vorgegebenen Zeit, im Regelfall zehn Minuten, gestartet wurde. Nach der Prüfung gab es ein ca. 5 minütiges Feedbackgespräch, in dem unter anderem die erreichte Punktzahl mitgeteilt wurde und der Prüfungsablauf im Detail durchgesprochen wurde.

Geprüft wurden die Studierenden von den studentischen Tutor*innen. Die Prüfungsberechtigung für den jeweiligen PC erhielten die Tutor*innen, wenn sie diesen PC selber bestanden, sie diesen PC selbst mehrmals co-geprüft und mindestens eine Prüfung unter ärztlicher Supervision absolviert hatten. Zur Qualitätssicherung und zu Studienzwecken gab es die Möglichkeit, als Co-Prüfer*in (CP), parallel zum*r Hauptprüfer*in (HP), den PC zu bewerten. Um als CP agieren zu dürfen, war es allerdings nicht notwendig, den PC vorher erfolgreich abgelegt zu haben. Allerdings fiel dem CP eine reine beobachtende Position zu. Entscheidungen zur Note und Punktevergabe traf der HP ohne Absprache mit dem CP. Als CP an einem PC konnten prinzipiell beliebig viele Personen teilnehmen. So war es möglich, dass CP und HP im Anschluss ihre Prüfungseindrücke besprechen und so die notwendige Prüferfahrung gesammelt werden konnte.

2.1.3 **Prüfungsbogen**

Als Bewertungsbogen wurde bei den PCs der, speziell für das Skills Lab entwickelte, Prüfungsbogen (PB) verwendet (s. Anhang 7.2). Dieser wurde sowohl von den Tutor*innen als auch von den Geprüften ausgefüllt und diente der Prüfungsdokumentation.

Ein erster Teil wurde von den Studierenden vor dem Start des PCs ausgefüllt. Erhoben wurde: ein achtstelliger anonymisierter Code, ob die Lernziele verständlich formuliert waren, ob die praktische Fertigkeit von den Tutor*innen gut vermittelt wurde und wie oft die praktische Fertigkeit im Skills Lab von dem*der Geprüften trainiert wurde. Nach dem PC sollte der*die Geprüfte einschätzen, ob der PC entsprechend der angekündigten Kriterien durchgeführt wurde, auf welchem Niveau er*sie sich selber einordnen würde, wann diese Fertigkeit das erste Mal von ihm*ihr verlangt wurde und mit welcher Schulnote er*sie die Übungsstation und den PC insgesamt bewerten würde.

Von dem*der HP wurden Informationen zu Prüfer*innennummer, PC-Nummer, Datum, restlichen Prüfungszeit und zur Anzahl bereits abgelegter PCs des*der Geprüften erhoben. Des Weiteren wurde von dem*der HP eine Einschätzung abgegeben, ob verglichen mit der objektiven Note, die Leistung des*der Geprüften subjektiv abwich, auf welchem Niveau sich der*die Geprüfte befand, ob der*die Geprüfte für den*die HP subjektiv bestanden hätte und ob der Ablauf des PC zufriedenstellend war. Parallel zur Prüfung wurden die vergebenen Punkte 1 - 20 angekreuzt. Da der PB ein universell für alle PCs nutzbarer Bogen

war, wurden die jeweils verlangten Kriterien auf einem separaten Blatt beschrieben und nur die Punkte zur Dokumentation angekreuzt.

2.1.4 Lehrevaluationssystem EvaSys

Für die Auswertung aller PB wurde das Programm EvaSys (EvaSys Version 8.1. EvaSys GmBH, Lüneburg, Deutschland) verwendet [17]. Die mit Hilfe von EvaSys maschinell lesbar konstruierten Bögen wurden an einer speziell dafür eingerichteten Scanstation an der Universität Ulm automatisiert digitalisiert. Die resultierenden Rohdaten wurden in einer Tabelle zahlenkodiert wiedergegeben. Um mögliche Fehler zu finden, wurden alle relevanten PB manuell auf ihre Richtigkeit kontrolliert. Rekonstruierbare Fehler, beispielsweise mehrfach gesetzte Kreuze bei einer Single-Choice-Frage, wurden händisch verbessert.

2.2 Prüferstudie

Die Studie wurde in folgende Teile gegliedert: Im ersten Teil wurden zunächst die gesammelten Daten der PCs mit einem*r Prüfer*in analysiert. Im zweiten Teil wurden PCs von jeweils zwei Prüfer*innen bewertet, welche hinsichtlich ihrer Übereinstimmung der Bewertung analysiert wurden. Im dritten Teil wurden die Übereinstimmung mehrerer Prüfer*innen (Studierende, Tutor*innen und Ärzt*innen) und die Qualität der Bewertungen systematisch an 4 Filmen des PC Untersuchung Abdomen analysiert.

2.2.1 Teil 1: Praxis-Checks mit einem*r Prüfer*in

Um das System der PCs in Zusammenhang mit Peer Teaching und Peer Assessment zunächst zu beschreiben, wurden im Zeitraum WiSe 18/19, SoSe 19 und WiSe 19/20 alle PB gesammelt und analysiert. Als HP zugelassen waren alle aktiven Tutor*innen im Skills Lab, welche im genannten Zeitraum angestellt waren. Geprüft wurden Studierende, die sich freiwillig oder im Rahmen von Wahlpflichtfächern zu PCs angemeldet hatten.

2.2.2 Teil 2: Vergleich der Bewertungen von Haupt- und Co-Prüfer*innen bei realen Praxis-Checks

Um Prüfer*innen vergleichen zu können, wurden PCs von mehreren Prüfer*innen bewertet. Ein PC wurde jeweils gleichzeitig von einem*r HP und einem*r CP bewertet. Hierfür wurden im SoSe 19 und WiSe 19/20 Prüfungen mit mehreren CP durchgeführt. Die Daten der CP wurden rein zu Studienzwecken und zur Qualitätssicherung erhoben. Als CP

wurden Ärzt*innen und Tutor*innen eingesetzt. Die Seite der ärztlichen Prüfer*innen wurde von der medizinischen Leitung des Skills Lab Universität Ulm übernommen.

2.2.3 Teil 3: Vergleich der Bewertungen mehrerer Prüfer*innengruppen durch Online-Praxis-Check-Simulation

Um die Nachteile der vorangegangenen Studienabschnitte zu eliminieren, wurde ein einheitliches Prüfmaterial in Form von vier Filmen eines PC erstellt. Dies ermöglichte einen direkten Vergleich der Bewertungen mehrerer Prüfer*innen an identischen Prüfungsfilmen. Hierfür wurde im SoSe 20 ein Onlinestudienformat entwickelt und durchgeführt.

2.3 Aufbau, Methodik und Durchführung der Online-Praxis-Check-Simulation

Für die Prüfungsfilme wurde der PC Untersuchung Abdomen ausgewählt, da dieser PC einer der häufigsten geprüften PCs im Skills Lab war (s. Anhang 7.5). Außerdem waren Kontext- und Skillpunkte ausgeglichen zu jeweils 10 Punkten vertreten. Für jeden Prüfungsfilm wurde ein detailliertes Drehbuch mit präzisen Anweisungen für die Schauspieler*innen des Untersuchers, des Patienten und der Tutorin erstellt (s. Anhang 7.10). Die vom Untersucher gemachten Fehler wurden alle anhand des PB festgelegt.

Jeder der 20 erreichbaren Punkte wurde, wenn möglich, nur einmal als Fehler eingebaut, um einen Wiederholungseffekt zu vermeiden. In den Prüfungsfilmen erreichte der Untersucher in einem Film die volle Punktzahl von 20 Punkten, fiel einmal mit 11 von 20 Punkten durch und bestand in zwei Fällen mit 17 von 20 Punkten. Insgesamt wurden von 20 Punkten 15 explizit falsch durchgeführt. Diese mussten von den Probanden als fehlerhaft identifiziert werden. Diese Filme werden im Kapitel 0 in gezeigter Reihenfolge beschrieben.

Mit allen Schauspieler*innen wurden deren Rollen detailliert besprochen. Die Rolle des Untersuchers übernahm ein freiwilliger Tutor aus dem Skills Lab. Die gezeigte Prüferin wurde von einer freiwilligen Tutorin übernommen. Der Patient wurde von einem Schauspieler ohne medizinischen Hintergrund gespielt, um dessen Reaktionen unverfälscht dazustellen zu können.

2.3.1 Filmvarianten: Praxis-Check Untersuchung Abdomen

Film 1: Skillfehler-Variante

Als erster Film wurde den Prüfer*innen die Skillfehler-Variante gezeigt. Der Schauspieler führte drei Skillpunkte fehlerhaft aus. Als erstes wurden nur die oberen Quadranten des Abdomens auskultiert (Punkt 4) anstatt alle vier Quadranten des Abdomens abzuhören. Als zweiter Fehler wurde die Reihenfolge vertauscht. Zuerst wurde die Perkussion, dann die Palpation und als letztes die Auskultation durchgeführt (Punkt 9), wobei die Auskultation der erste Schritt gewesen wäre. Als letzter Fehler wurde die falsche Palpation der Milz auf den Rippen demonstriert (Punkt 11). Die Zielpunktzahl lag bei 17 Punkten.

Film 2: Fehlerfrei-Variante

Als zweiter Film wurde die Fehlerfrei-Variante gezeigt. Die Zielpunktzahl lag bei 20 Punkten.

Film 3: Kontextfehler-Variante

Als dritter Film wurden den Prüfer*innen die Kontextfehler-Variante gezeigt. Der Schauspieler führte drei Kontextpunkte fehlerhaft aus. Als erstes wurde die nachfolgende Untersuchung nicht angekündigt (Punkt 2). Der zweite Fehler war die, mit 15 Sekunden, zu kurze hygienische Händedesinfektion (Punkt 15). Als dritter Fehler wurden bei der Dokumentation die Stammdaten auf dem Befundbogen weggelassen (Punkt 17). Die Zielpunktzahl lag bei 17 Punkten.

Film 4: Mischfehler-Variante

Als vierter Film wurde eine Kombination aus mehreren Skill- und Kontextfehlern gezeigt. Der Schauspieler führte insgesamt 9 Punkte fehlerhaft aus. Beginnend mit einer rhetorischen Frage bei Ansprache des Patienten (Punkt 1), wurde als nächster Fehler bei der hygienischen Händedesinfektion vor Patientenkontakt ein Armband getragen (Punkt 3). Des Weiteren wurde die Kontrolle des Loslass-Schmerzes (Punkt 8), die Kratzauskultation der Leber (Punkt 10) und die Prüfung des Nierenklopfschmerzes (Punkt 12) mangelhaft demonstriert. Die hygienische Händedesinfektion nach Patientenkontakt wurde komplett weggelassen (Punkt 15). Die Antwort des Untersuchers auf die Frage der weiterführenden Diagnostik war falsch (Punkt 18) und die Dokumentation der Leber und

der Milz auf dem Befundbogen wurde weggelassen (Punkt 20). Die Zielpunktzahl lag bei 11 Punkten.

2.3.2 **Ethik und Einverständniserklärung**

Die gezeigten Schauspieler*innen gaben vorab ihr schriftliches Einverständnis zur Verwendung des Filmmaterials. Hinsichtlich der Auswertung der erhobenen Daten entschied die Ethikkommission der Universität Ulm am 15.05.2020 und 16.04.2019, dass kein Ethikvotum benötigt wird.



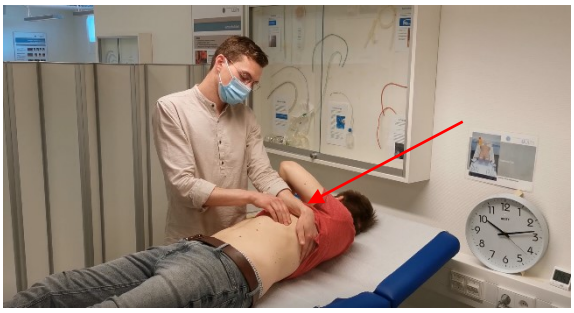
a



b



c



d

Befundbogen Praxis-Check

Datum: 27.5.20 Uhrzeit: 10 30

Untersucher (Name, Funktion): Schmidt, Medizinstudent

Patient (Name, Geburtsdatum): Holz, Maximilian, 23.1.1992

Befund:

schlanker Ernährungszustand, ø Leberhautzeichen, ø OP-Narben
lebhafte Darmgeräusche in allen 4 Quadranten
ø aortale Strömungsgeräusche
tympanitischer Klopfschall über allen 4 Quadranten

Bauchdecke weich, ø Abwehrspannung, ø Druckschmerz, ø Loslass-Schmerz, ø tastbaren Resistenzen
Leber 11cm in MCL, unter Rippenbogen ø tastbar
Milz ø tastbar
Nierenlager beidseits ø klopfschmerzhaft

e

Befundbogen Praxis-Check

Datum: 27.5.2020 Uhrzeit: 9 45

Untersucher (Name, Funktion): Schmidt, Medizinstudent

Patient (Name, Geburtsdatum): Holz

Befund:

schlanker Ernährungszustand, ø Leberhautzeichen, ø OP-Narben
lebhafte Darmgeräusche in allen 4 Quadranten
ø aortalen Strömungsgeräusche,
tympanitischer Klopfschall über allen 4 Quadranten
Bauchdecke weich, ø Abwehrspannung, ø Druckschmerz, ø Loslass-Schmerz, ø tastbare Resistenzen,
Leber 11cm in MCL, unter Rippenbogen ø tastbar
Milz ø tastbar
Nierenlager beidseits ø klopfschmerzhaft

f

Abbildung 3a-f: Beispiel für Kontext- und Skillfehler in den Prüfungsfilmen. a) richtige Händehygiene, b) falsche Händehygiene (Armband am linken Handgelenk, Kontextpunkt); c) richtige Milzpalpation, d) falsche Milzpalpation (Palpation auf den Rippen, Skillpunkt); e) richtige Dokumentation Stammdaten, f) falsche Dokumentation Stammdaten; Fehler mit rotem Pfeil oder Rahmen gekennzeichnet; Einverständnis der gezeigten Personen liegt vor. Fiktive Patientendaten.

2.3.3 **Ablauf der Online-Praxis-Check-Simulation**

Die Online-PC-Simulation auf der Lernplattform Moodle präsentierte sich im Ablauf für alle Prüfer*innen identisch. Insgesamt war mit einem zeitlichen Aufwand von einer Stunde zu rechnen. Für die erfolgreiche Bearbeitung wurde ein Computer oder Laptop, stabiles Internet und gegebenenfalls ein Kopfhörer benötigt.

Von den Prüfer*innen wurden drei Fragebögen und vier Filme bearbeitet. Zu Beginn wurden im ersten Fragebogen persönliche Daten wie beispielsweise Alter, Geschlecht und Ausbildungsstand erhoben. Im nächsten Schritt wurde in einem Kurzfilm der Umgang mit dem PB erklärt. Nach der Erklärung erhielten die Teilnehmenden die Möglichkeit, sich selbst mit dem PB vertraut zu machen.

Für die erfolgreiche Bearbeitung der Prüfungssituation anhand der vier Prüfungsfilme war es erforderlich, einen geteilten Bildschirm zu erstellen (s. Anhang 7.5). Die Erstellung des geteilten Bildschirms wurde in einem Kurzfilm demonstriert. Alle Prüfer*innen mussten probeweise einen geteilten Bildschirm selbst erstellen und die erfolgreiche Erstellung bestätigen. Dieser Schritt wurde benötigt, da einige Internetbrowser unterschiedliche Vorgehensweisen erforderten.

Nach dem Absolvieren der ersten technischen Schritte wurden die vier Prüfungsfilme in der in Kapitel 3.3.1. beschriebenen Reihenfolge gezeigt. Um parallel zum Film den PB bewerten zu können, mussten die Teilnehmenden zum Start jedes Prüfungsfilmes ihren Bildschirm teilen. Auf einer Hälfte war der Film mit der Prüfungssituation sichtbar, während auf der anderen Hälfte der digitalisierte PB bearbeitet werden konnte. Die Prüfer*innen übernahmen in diesem Fall die Rolle des*der HP. Für jede Prüfung wurde so von den Prüfer*innen ein kompletter digitaler PB ausgefüllt. Die Reihenfolge der Filme war bei allen Prüfer*innen identisch.

Beendet wurde die Studie mit einem letzten Fragebogen. In diesem wurden die Teilnehmenden um Rückmeldung gebeten, unter anderem zur Selbsteinschätzung.

2.3.4 **Kollektiv der Teilnehmenden**

Die Prüfer*innen der Studie setzten sich aus Studierenden, Tutor*innen und Ärzt*innen zusammen und wurden in gleichnamige Gruppen unterteilt.

Studierende für die Vergleichsgruppe wurden aus den Humanmedizinstudierenden der Universität Ulm rekrutiert. Die teilnehmenden Studierenden sollten sich im SoSe 20 im klinischen Abschnitt ihres Studiums befinden und das PJ nicht begonnen haben. Studierende, die vor weniger als einem Jahr als aktive Skills Lab-Tutor*innen tätig waren, wurden zu der Tutor*innengruppe gezählt. Als Dank für eine erfolgreiche Teilnahme an der Studie wurde den Studierenden ein Amazon-Gutschein im Wert von 10 € angeboten.

Die Gruppe der Tutor*innen setzte sich aus allen im SoSe 20 aktiven Skills Lab-Tutor*innen zusammen, die nicht bei der Erstellung der Studie beteiligt waren. Es wurde nicht unterschieden, ob der*die Tutor*in das erste Mal im Skills Lab aktiv war oder diese Tätigkeit bereits über mehrere Semester ausgeübt wurde. Da die Tutor*innen im Rahmen ihres Arbeitsvertrages aufgefordert wurden, an der Studie teilzunehmen, ersetzte die Studie das Training der Tutor*innen zu Beginn des SoSe 20.

Für die ärztliche Referenzgruppe wurden Ärzt*innen jeder Fachrichtung, jedes Ausbildungstandes und jeder Erfahrung rekrutiert. Es gab keine standortspezifische Bindung. Voraussetzung für die ärztlichen Teilnehmenden war lediglich die Approbation in der Humanmedizin. Als Dank für eine erfolgreiche Teilnahme an der Studie wurde den Ärzt*innen ein Amazon-Gutschein im Wert von 20 € angeboten.

Studierende und Ärzt*innen wurden über standardisierte Emails auf die Studie aufmerksam gemacht und um Teilnahme gebeten (s. Anhang 7.7). Potenzielle Teilnehmende wurden über die Forschungsfrage, Dauer und Honorierung der Studie informiert. Im Falle einer positiven Rückmeldung wurde den Teilnehmenden eine zweite standardisierte Email gesendet. In dieser wurde den Teilnehmenden unter anderem der Link zur Studie geschickt. Ärzt*innen, die keinen Zugang zur Lernplattform Moodle der Universität Ulm hatten, wurden zusätzlich mit einem personalisierten temporären Moodlezugang mit Benutzername und Passwort ausgestattet.

2.3.5 **Fragebögen**

Die von allen Teilnehmenden zu bearbeitenden Fragebögen setzten sich aus einem Eingangsfragebogen, den digitalisierten PB und dem Abschlussfragebogen zusammen. Als Fragetypen wurden Einfach- und Mehrfachauswahlfragen verwendet. Bei Fragen zur persönlichen Einstellung der Teilnehmenden wurde eine 6-stufige Likert-Skala angewendet.

Im Eingangsfragebogen wurden personenbezogene Daten erhoben, wie das Einverständnis zur Studienteilnahme, ein pseudonymisierter Code, Alter und Geschlecht (s. Anhang 7.7). Weitere Fragen bezogen sich auf die Erfahrung mit dem Umgang des PC Untersuchung Abdomen, den Erfahrungen als Prüfer*in im Allgemeinen, inwieweit die Prüfer*innen sich selber als objektive*n Prüfer*in einschätzen würden und ob sie sich sicher fühlen würden, studentische Leistungen zu beurteilen. Zusätzlich sollten die Teilnehmenden ihre Fähigkeit einschätzen, eine strukturierte Abdomen Untersuchung durchzuführen und eine Abdomen Untersuchung qualitativ zu bewerten.

Studierende und Tutor*innen wurden im Gegensatz zu den Ärzt*innen nach der genauen Fachsemesterzahl gefragt, ob diese bereits als Tutor*in tätig gewesen waren, in welchem Fach dieses Tutorat ausgeübt wurde und nach ihrer Berufserfahrung im medizinischen Bereich. Nur die ärztliche Gruppe wurde des Weiteren nach Fachrichtung und ihrer Berufserfahrung gefragt.

Für die Online-Simulation wurden der PB auf der Lernplattform Moodle digital rekonstruiert (s. Anhang 7.3). Die Fragen und Antwortmöglichkeiten wurden vom PB eins zu eins übernommen. Aufgrund der von Moodle vorgegebenen Layoutmöglichkeiten mussten einige Fragen in ihrem Erscheinungsbild modifiziert werden. Für die Onlinestudie nicht notwendige Fragen, wie beispielsweise das Datum, wurden zu Gunsten einer besseren Übersicht aus dem Online-PB entfernt.

Im Abschlussfragebogen wurden mehrere Fragen zur Prüfungssituation und dem PB gestellt (s. Anhang 7.8). Die Teilnehmenden wurden um ihre Einschätzung gebeten, ob ihnen die Bewertung der Prüfungen leichtgefallen war, ob sie den PC als geeignet empfanden, um praktische Leistungen abzufragen und ob der PB hilfreich bei der Punktevergabe war. Des Weiteren wurden die Teilnehmenden befragt, wie gut sie die

Urteiler*innenübereinstimmung in der ärztlichen Gruppe und in der Tutor*innen Gruppe einschätzten und ob diese untereinander ebenfalls übereinstimmten oder abwichen.

2.3.6 Film und Schnitt

Bilder und Ton für die gezeigten Prüfungsfilme wurden mit dem Mobiltelefon Huawei P30 in 4K-Auflösung aufgenommen. Um ein stabiles und bewegungsarmes Bild zu erhalten, wurde das Handy auf einem Fotostativ befestigt und während den Aufnahmen nicht bewegt. Der jeweils beste gefilmte Durchlauf des PC wurde im Anschluss mit dem frei erhältlichen Programm DaVinci Resolve 17 bearbeitet (Blackmagic Design Pty., Port Melbourne, Australien [6]).

Mit diesem Programm wurden alle Filme nachbearbeitet, Titel eingeblendet, Filmteile geschnitten und das Hintergrundrauschen in der Tonspur minimiert. Ortswechsel wurden mit einem Schnitt verdeutlicht. Der Anfang jedes Filmes wurde mit dem identischen Intro versehen, um den Prüfer*innen bei jedem Film eine gleich lange Vorbereitungsphase zu ermöglichen.

2.4 Statistische Auswertung

Die Bewertungen der Haupt- und Co-Prüfer*innen im zweiten Studienteil wurden als übereinstimmend definiert, wenn die Differenz der vergebenen Punkte ≤ 2 Punkte betrug. In Rücksprache mit der Leitung des Skills Lab wurden für den dritten Studienteil 18 von 20 (90 %) korrekt gewerteter Punkte als eine gute Leistung der Prüfer*innen definiert. Ein Punkt wird als korrekt gewertet, wenn dieser als „zutreffend richtig“ oder „zutreffend fehlerhaft“ ausgeführt erkannt wurde. Beide Möglichkeiten werden als korrekt bezeichnet und bewertet.

Unrealistische Prüfungswerte wurden meist durch eine instabile Internetverbindung oder durch andere technische Probleme ausgelöst. Diese PB wurden in der Auswertung nicht berücksichtigt. Die Daten wurden mit Median ± Interquartilsabstand (IQR) dargestellt. Die Analyse der Daten wurde mit GraphPad Prism durchgeführt (GraphPad Software Version 8, San Diego, USA [46]).

2.4.1 **Urteiler*innenübereinstimmung zwischen den Prüfer*innen**

Um die Übereinstimmung der Prüfer*innen zu analysieren, wurden die medianen Durchschnittswerte (± IQR) für die Allgemeinpunktzahl, für die jeweiligen PCs und für die einzelnen Punkte verglichen. Es wurde die Prüfer*innen-Übereinstimmung in Prozent angegeben. Im dritten Studienteil wurde die Übereinstimmung mehrerer Prüfer*innen mit Hilfe von SPSS die Intraklassen-Korrelation errechnet (IBM SPSS Statistics Version 27, Armonk, USA [55]). Die Übereinstimmung zwischen den Prüfer*innen wurde auf einer Scala von 0 – 1 angegeben und wurde nach Koo et al. 2016 interpretiert [29]. Werte < 0,5 wurden als eine schlechte, 0,5 – 0,75 als eine moderate, 0,75 – 0,9 als eine gute und > 0,9 als eine exzellente Übereinstimmung interpretiert [29].

Die statistische Testung erfolgte unter Annahme einer nichtparametrischen Verteilung der ungepaarten Daten. Der Vergleich zweier Gruppen wurde daher mit dem Mann-Whitney-Test bei ungepaarten Daten durchgeführt. Der Vergleich mehrerer Gruppen bei ungepaarten Daten wurde mit dem Kruskal-Wallis-Test und anschließendem unkorrigiertem Dunn-Test durchgeführt. Der unkorrigierte Dunn-Test wurde gewählt, da maximal ein Vergleich von drei Gruppen stattfand. Zum Vergleich der Daten mit einem hypothetischen Median (z.B. 0) wurde der One sample Wilcoxon-Test (auch: Wilcoxon Signed Rank Test) verwendet. Ein signifikanter Unterschied wurde ab $p < 0,05$ angenommen. Die statistische Analyse erfolgt mit GraphPad Prism [46].

3 Ergebnisse

Das Prüfungssystem der PCs im Skills Lab der Universität Ulm ist nach unserem Wissen einmalig in Deutschland. Viele Universitäten setzen die OSCE als ähnliches Lernkontrollsystem in verschiedenen medizinischen Fächern bereits ein. Von einer allgemeinen standardisierten und freiwilligen Lernzielkontrolle in Skills Labs wie den PCs lagen uns aber keinerlei Informationen vor. Aus diesem Grund und zur Qualitätssicherung wurden die PC-bezogenen Daten systematisch ab WiSe 18/19 untersucht.

3.1 Analyse der Praxis-Checks mit einem*r Prüfer*in

3.1.1 Datensatz

Im Zeitraum WiSe 18/19, SoSe 19 und WiSe 19/20 wurden insgesamt 919 PB erhoben (Abbildung 4). Fehlerhaft ausgefüllte PB wurden ebenso wie PB ärztlicher HP exkludiert. PB von PCs mit einer Gesamtanzahl unter zehn und PB von Tutor*innen die weniger als zehn PC-Prüfungen abgenommen hatten wurden aus dem Datensatz entfernt. Insgesamt wurden 775 auswertbare PB durch 28 Tutor*innen erstellt.

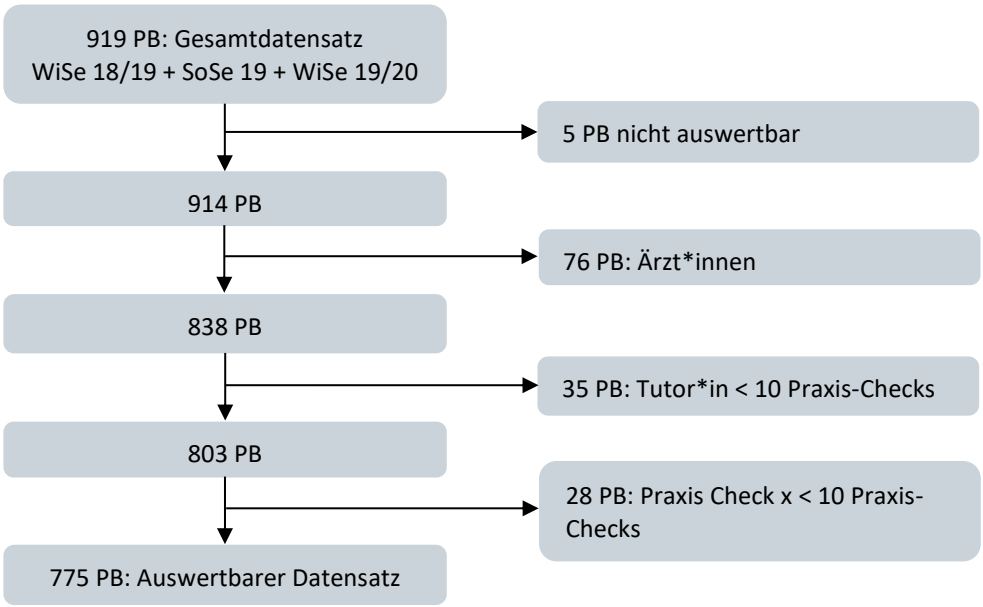


Abbildung 4: Ausschlusskriterien und Anzahl eingeschlossener Prüfungsbögen (PB). Insgesamt wurden 775 von 919 (entspricht 84,3%) der PB ausgewertet. Wintersemester (WiSe), Sommersemester (SoSe)

3.1.2 **Überblick über die Praxis-Checks**

Die am häufigsten abgelegten PCs waren Händehygiene (n = 278; 35,9 %), Vitalparameter (n = 86; 11,1 %), Knoten (n = 55; 7,1 %), Nähen (n = 52; 6,7 %) und Untersuchung Abdomen (n = 51; 6,6 %) (Abbildung 5).

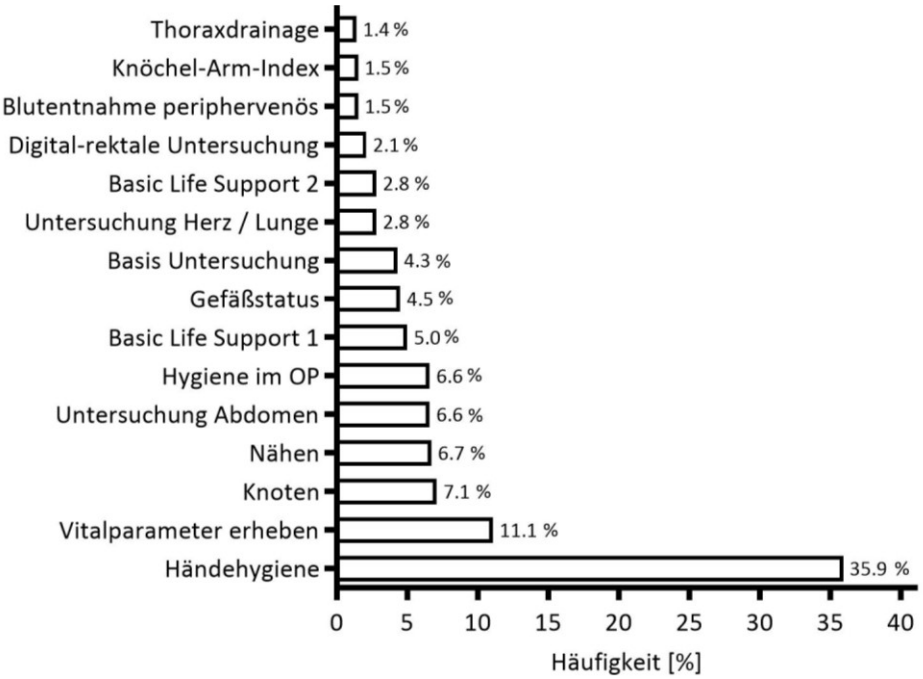


Abbildung 5: Häufigkeit der Praxis-Checks in % im Skills Lab Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; n = 775; Hygiene n = 278, Vitalparameter n = 86, Basis Untersuchung n = 33, Basic Life Support 1 n = 39, Blutentnahme peripher venös n = 12, Untersuchung Herz / Lunge n = 22, Gefäßstatus n = 35, Knöchel-Arm-Index n = 12, Basic Life Support 2 n = 22, Hygiene im OP n = 51, Untersuchung Abdomen n = 51, Digital-rektale Untersuchung n = 16, Nähen n = 52, Knoten n = 55, Thoraxdrainage n = 11

Der vergebene mediane Punktewert der PCs lag bei 20 (18;20) Punkten (Abbildung 6).

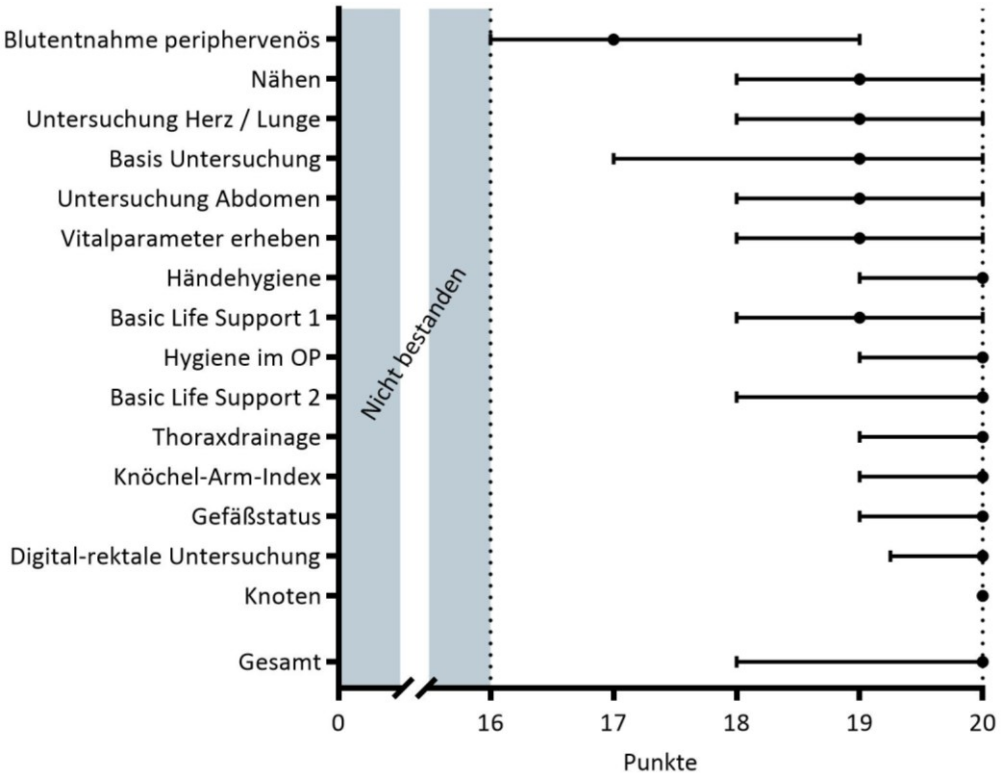


Abbildung 6: Geringer Unterschied im Punktedurchschnitt der Praxis-Checks (PC). Vergebene Durchschnittspunktzahl je PC im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; sortiert nach Mittelwert (MW): Knoten hoher MW, Blutabnahme periphervenös niedriger MW; Gestrichelte Linie: Gesamtmedian (20 Punkte); Gesamt n = 750, Hygiene n = 253, Vitalparameter n = 86, Basis Untersuchung n = 33, Basic Life Support 1 n = 39, Blutentnahme periphervenös n = 12, Untersuchung Herz / Lunge n = 22, Gefäßstatus n = 35, Knöchel-Arm-Index n = 12, Basic Life Support 2 n = 22, Hygiene im OP n = 51, Untersuchung Abdomen n = 51, Digital-rektale Untersuchung n = 16, Nähen n = 52, Knoten n = 55, Thoraxdrainage n = 11

3.1.3 **Vergleich der Bewertungen bei Praxis-Checks mit einem*r Prüfer*in**

Um die Tutor*innen miteinander vergleichen zu können, wurde die mediane Punktezahl aller Tutor*innen ermittelt (Abbildung 7). Die allgemeine Punktzahl lag bei 20 (18;20) Punkten. Ein Großteil der Tutor*innen wies eine Differenz von 5 % auf, während die maximale Differenz zwischen den Tutor*innen bei 10 % (entsprechend 2 Punkten) lag.

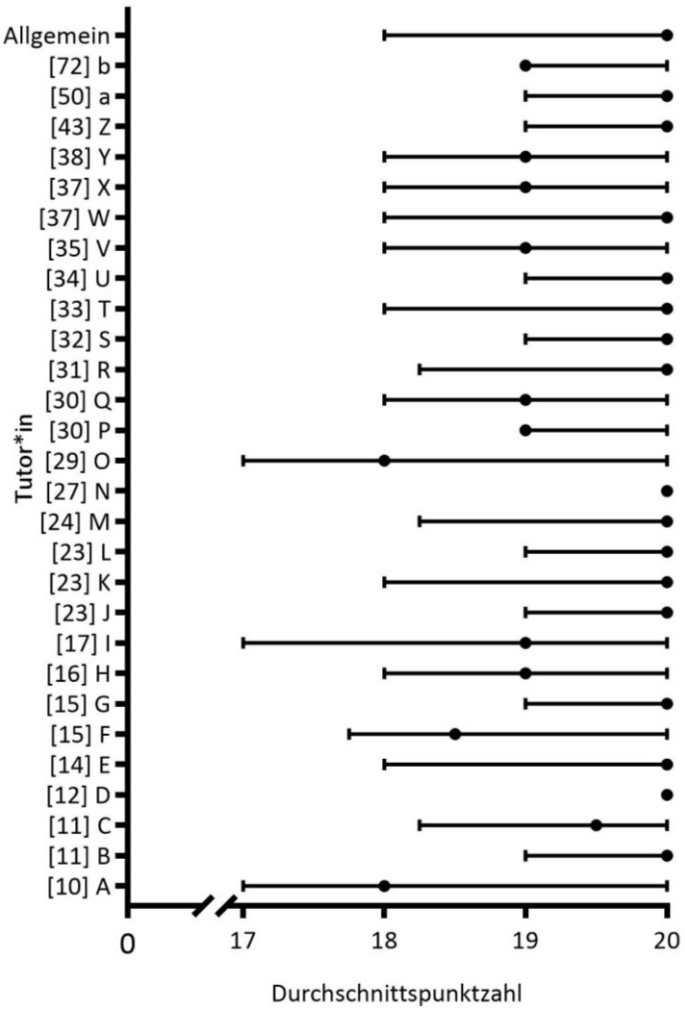


Abbildung 7: vergebene Durchschnittspunktzahl je Tutor*in (mindestens 10 Praxis-Checks geprüft) im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; Allgemein = allgemein vergebener Punktedurchschnitt; n = 747; [n] = Anzahl Praxis-Checks der Tutor*in; A-b = kodierte Tutor*innen

Zusätzlich wurde die Durchschnittspunktzahl der einzelnen Tutor*innen bezogen auf die jeweiligen PCs betrachtet (s. Anhang 7.1.1). In allen untersuchten PCs sind die medianen Punktewerte der Tutor*innen ebenfalls maximal 10 % vom Median entfernt. Tendenziell findet sich allerdings keine starke Abweichung der Tutor*innen zur allgemeinen Durchschnittspunktzahl. Des Weiteren konnte die Durchschnittspunktzahl nicht in Korrelation mit der Erfahrung gebracht werden (P (two-tailed) = 0,76; r = −0,06). Tutor*innen mit vielen PCs wiesen keinen Unterschied zu Unerfahreneren auf (s. Anhang 7.1.2). Im Zuge der Auswertung der acht Stationen mit Kontext- und Skillpunkten ergab sich, dass tendenziell mehr Fehler bei den Kontextpunkten gewertet wurden als bei Skillpunkten (Abbildung 8).

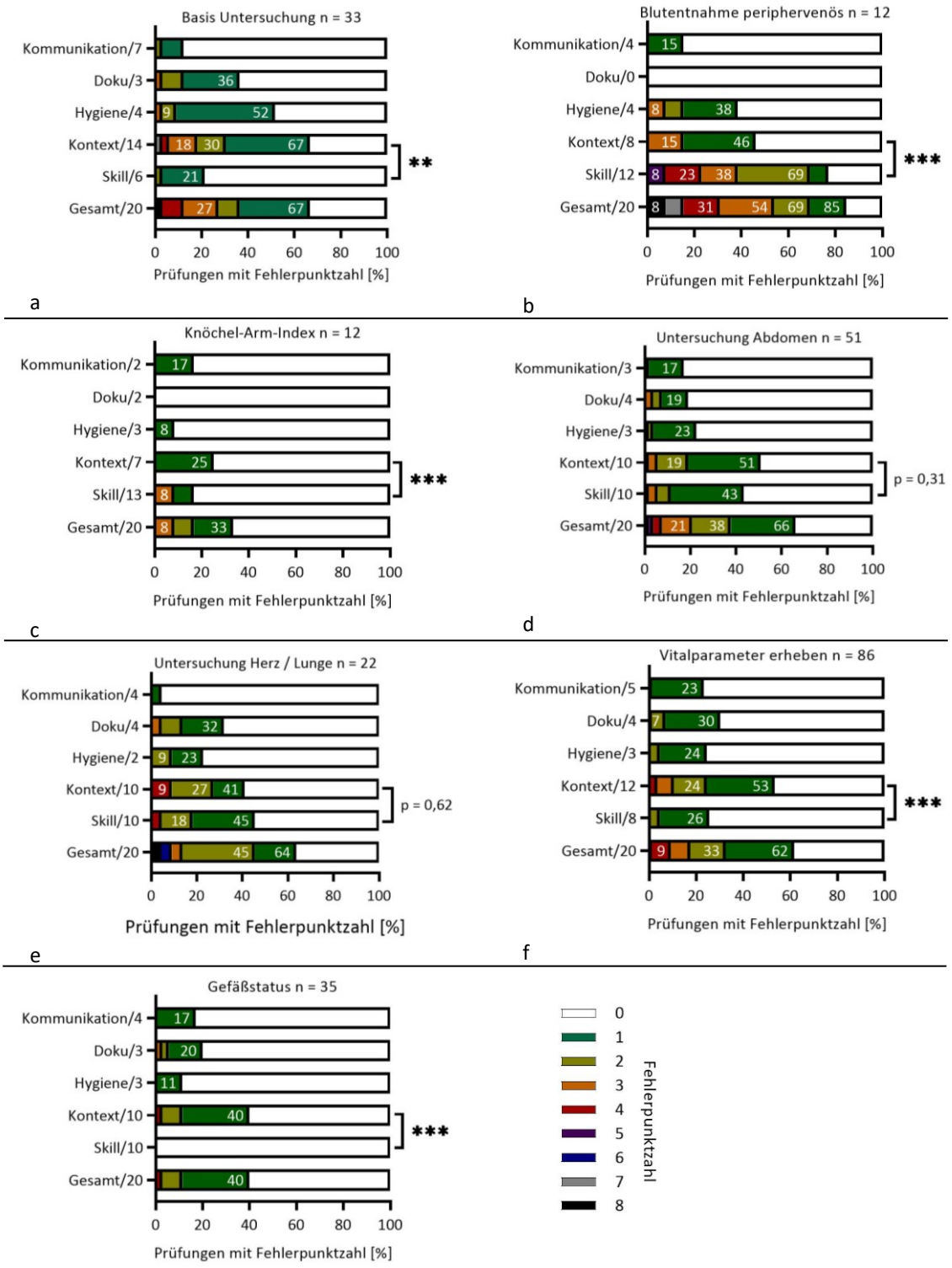


Abbildung 8a-g: Tendenziell werteten die Tutor*innen mehr Fehler in den Kontextstationen als in den Skillstationen. Verteilung Fehlerpunkte nach Skill, Kommunikation, Dokumentation (Doku) und Hygiene je Praxis-Check im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Kontext = Kommunikation, Hygiene und Dokumentation zusammengefasst; Lesebeispiel: Basis Untersuchung Kontext/14: 67 % haben mindestens einen Kontextfehler gemacht, 37 % haben genau einen Kontextfehler gemacht. Statistik: Wilcoxon-Test; *** ≙ $p < 0{,}001$, ** ≙ $p < 0{,}01$

3.1.4 **Einschätzung und Bewertung der Praxis-Checks von Studierenden und Tutor*innen**

Im Rahmen des PB wurden Studierende und Tutor*innen um Einschätzung zu verschiedenen Fragen gebeten. Die Fragen „Die Lernziele waren verständlich formuliert“ und „Die/der Tutorin/Tutor hat die praktische Fertigkeit gut vermittelt“ wurden mit „trifft völlig zu“ bewertet (Abbildung 9).

Sehr gut bewertet wurde auch, dass die PCs entsprechend der angekündigten Kriterien durchgeführt wurden (s. Anhang 7.1.11). Die Tutor*innen waren im Median mit dem Ablauf des PC völlig zufrieden (6(5;6), 6-stufige Likert-ähnliche Skala: 1 = trifft gar nicht zu, 6 = trifft völlig zu). Die PC-Stationen wurden von den Studierenden im Median 4 (3;5) Mal trainiert, bevor der PC abgelegt wurde. Die Studierenden waren im Median 91 - 120 (≥ 120 s; 31 - 60 s) Sekunden vorher fertig.

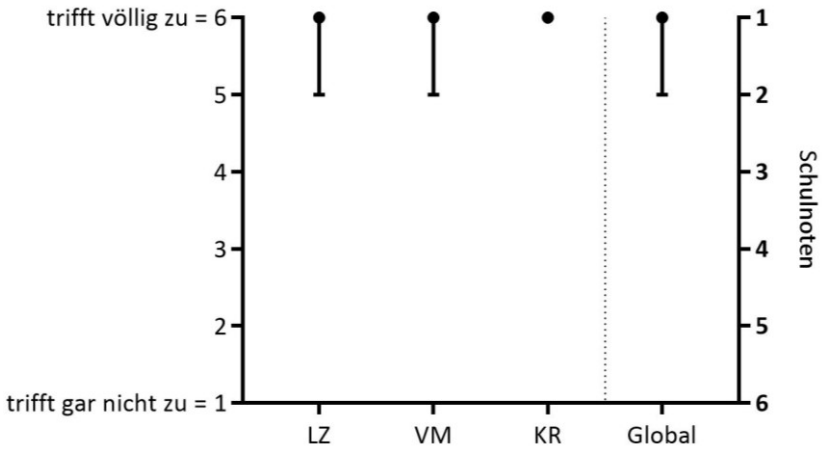


Abbildung 9: Darstellung der allgemeinen Evaluationen der Praxis-Checks. Auswertbare Prüfungsbögen im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; n = 775; Median ± Interquartilsabstand; LZ = Die Lernziele waren verständlich formuliert; VM = Die*der Tutor*in hat die praktische Fähigkeit gut vermittelt; KR = Der Praxis-Check wurde entsprechend der angekündigten Kriterien durchgeführt; Global = Insgesamt würde ich Übungsstation und Praxis-Check mit folgender Schulnote bewerten.

Die Eigeneinschätzung der abgefragten Fertigkeiten stimmte in der globalen Analyse mit der Fremdeinschätzung durch die Tutor*innen überein. Im Median wurde von Tutor*innen und Studierenden „selbstständig tun“ (8(7;8), Skala: 1 = beobachten und nachvollziehen, 9 = andere unterrichten) angegeben. Die subjektiv erbrachte Leistung wurde im Vergleich mit der objektiven Note von den Tutor*innen als „gleich“ (3(3;3), Skala: 1 = viel schlechter, 5 = viel besser) angegeben.

Die subjektive Bewertung der Tutor*innen stimmte in 96,2 % mit der objektiven Note überein (Tabelle 1). Allgemein betrachtet fielen die Ergebnisse der einzelnen PCs sehr ähnlich aus. Differenzen zwischen den einzelnen PCs fanden sich in der Prüfungsrestzeit, der Einschätzung der Fertigkeiten und in der medialen Punktzahl.

Tabelle 1: In 96,2 % der Prüfungen stimmt die subjektive Einschätzung der Tutor*innen mit der Note überein. Übereinstimmung zwischen subjektivem Empfinden der Tutor*innen und dem objektiven Bestehen/Nichtbestehen im Skills Lab der Universität Ulm; Zeitraum Wintersemester 18/19 bis 19/20; n = 753; Objektiv + = bestanden; Objektiv − = nicht bestanden; Subjektiv + = nach Meinung des*der Tutor*in bestanden; Subjektiv − = nach Meinung des*der Tutor*in nicht bestanden

Allgemein	Subjektiv +	Subjektiv –	Gesamt
Objektiv +	695 (92,2%)	25 (3%)	720
Objektiv –	4 (0,5%)	29 (4%)	33
Gesamt	699	54	753

3.2 Bewertungen von Haupt- und Co-Prüfer*innen bei realen Praxis-Checks

Da die Tutor*innen in der vorherigen Analyse nur anhand der Durchschnittspunktzahl verglichen werden konnten, wurde im nächsten Schritt ein PC von mehreren Prüfer*innen gleichzeitig bewertet. Die CP wurden in diesem Fall zu reinen Studienzwecken eingesetzt. Dies ermöglichte einen direkten Vergleich zwischen den Prüfer*innen, egal ob ärztlich oder studentisch. Allerdings konnten mit dieser Variante nur die Prüfer*innen des gleichen PC verglichen werden.

3.2.1 **Datensatz**

Insgesamt wurden im SoSe 19 und WiSe 19/20 ein Rohdatensatz von 305 Prüfungen mit Haupt- und Co-Prüfer*innen erstellt (Abbildung 10). 28 Prüfungen mussten entfernt werden, da die CP nicht zugelassen waren oder der PB nicht ausgefüllt war. Dadurch ergab sich ein auswertbarer Datensatz von 277 zusammenhängenden HP- und CP-Bögen. Insgesamt waren 30 Tutor*innen als HP und 25 Tutor*innen (Co-Prüfer*in Tutor*in; CP-T) und Ärzt*innen (Co-Prüfer*in Ärzt*in; CP-A) als CP tätig. CP wurden in den PCs Händehygiene, Untersuchung Abdomen, Nähen, Gefäßstatus und Thoraxdrainage eingesetzt.

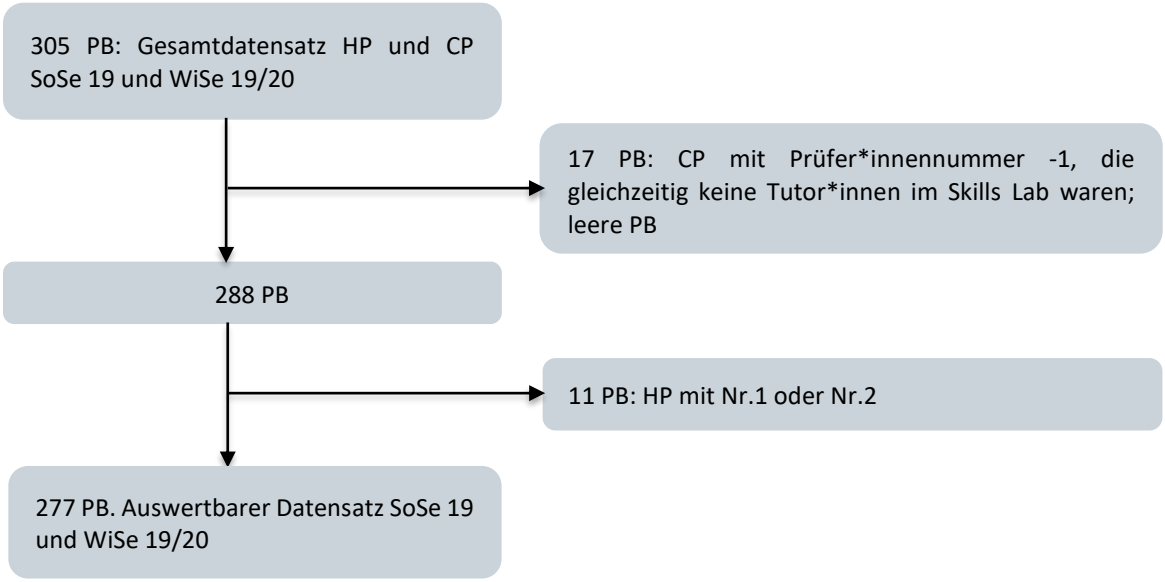


Abbildung 10: Ausschlusskriterien und Anzahl eingeschlossener Prüfungsbögen (PB) im Skills Lab der Universität Ulm. Insgesamt wurden 277 von 305 (entspricht 91 %) der PB ausgewertet. HP = Hauptprüfer*in, CP = Co-Prüfer*in; Wintersemester (WiSe), Sommersemester (SoSe)

3.2.2 **Vergleich der Bewertungen von Haupt- und Co-Prüfer*innen bei realen Praxis-Checks**

Im ersten Schritt wurden, wie zuvor, die erreichten Durchschnittspunktzahlen verglichen (Abbildung 11). Betrachtet man den allgemeinen Punktedurchschnitt, stimmten HP, CP-T und CP-A mit 19 Punkten überein. Bei detaillierterer Betrachtung der einzelnen PCs ergaben sich kleine Unterschiede. Die mediane Punktedifferenz zwischen CP-T und HP betrug nur 0,5 Punkte beziehungsweise 2,5 %. Der Unterschied zwischen CP-A und HP betrug maximal 1 Punkt beziehungsweise 5 %. Interessanterweise fand sich die größte Differenz zwischen CP-T und CP-A mit 1,5 Punkten. Der PC Nähen erreichte bei allen

Gruppen im Median die gleiche Durchschnittspunktzahl (18 Punkte). Gleichzeitig wies dieser PC mit 90 % Übereinstimmung zwischen HP und CP den niedrigsten Wert auf. Händehygiene erreichte eine mediane prozentuale Übereinstimmung von 91 %, während Thoraxdrainage (95 %) und Untersuchung Abdomen (97 %) bessere Werte erreichten (s. Anhang 7.1.4).

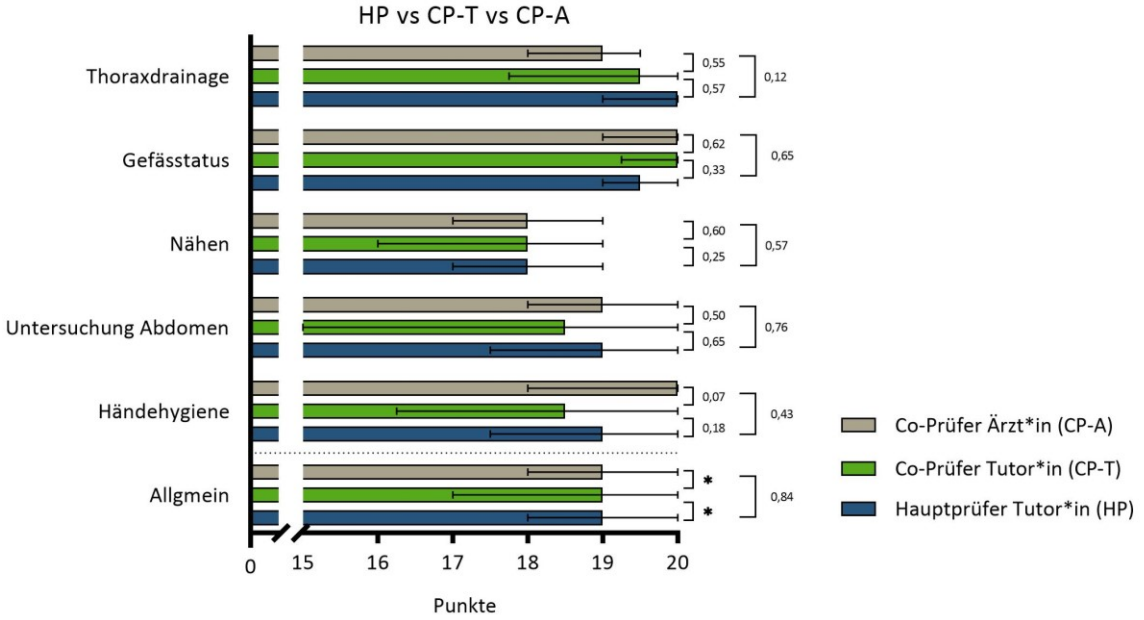


Abbildung 11: Mediane Punktzahl der HP und CP (Tutor*in/ Ärzt*in) im Skills Lab der Universität Ulm. Zeitraum: Sommersemester 19 und Wintersemester 19/20; Median ± Interquartilsabstand; Allgemein: n(HP) = 277, n(CP-T) = 124, n(CP-A) = 153, Median = 19; Hygiene n(HP) = 89, n(CP-T) = 28, n(CP-A) = 61, Median HP = 19, Median CP-T = 18,5, Median CP-A = 20; Untersuchung Abdomen n(HP) = 33, n(CP-T) = 14, n(CP-A) = 19, Median(HP/CP-A) = 19, Median CP-T = 18,5; Nähen n(HP) = 38, n(CP-T) = 19, n(CP-A) = 19, Median = 18; Gefäßstatus n(HP) = 22, n(CP-T) = 8, n(CP-A) = 10, Median HP = 19,5, Median (CP-T/CP-A) = 20; Thoraxdrainage n(HP) = 19, n(CP-T) = 6, n(CP-A) = 13, Median HP = 20, Median CP-T = 19,5, Median CP-A = 19;Statistik: ✱ $p < 0,05$; Kruskal-Wallis-Test mit unkorrigiertem Dunn-Test

Um zu überprüfen, ob eine Prüfer*innengruppe tendenziell mehr oder weniger Punkte vergab als eine andere, wurden Punktedifferenzen zwischen HP und CP erstellt (Abbildung 12). Dort zeigte sich eine mediale Abweichung von 0 Punkten. Ein signifikanter Unterschied zwischen CP-T und CP-A konnte nicht ermittelt werden.

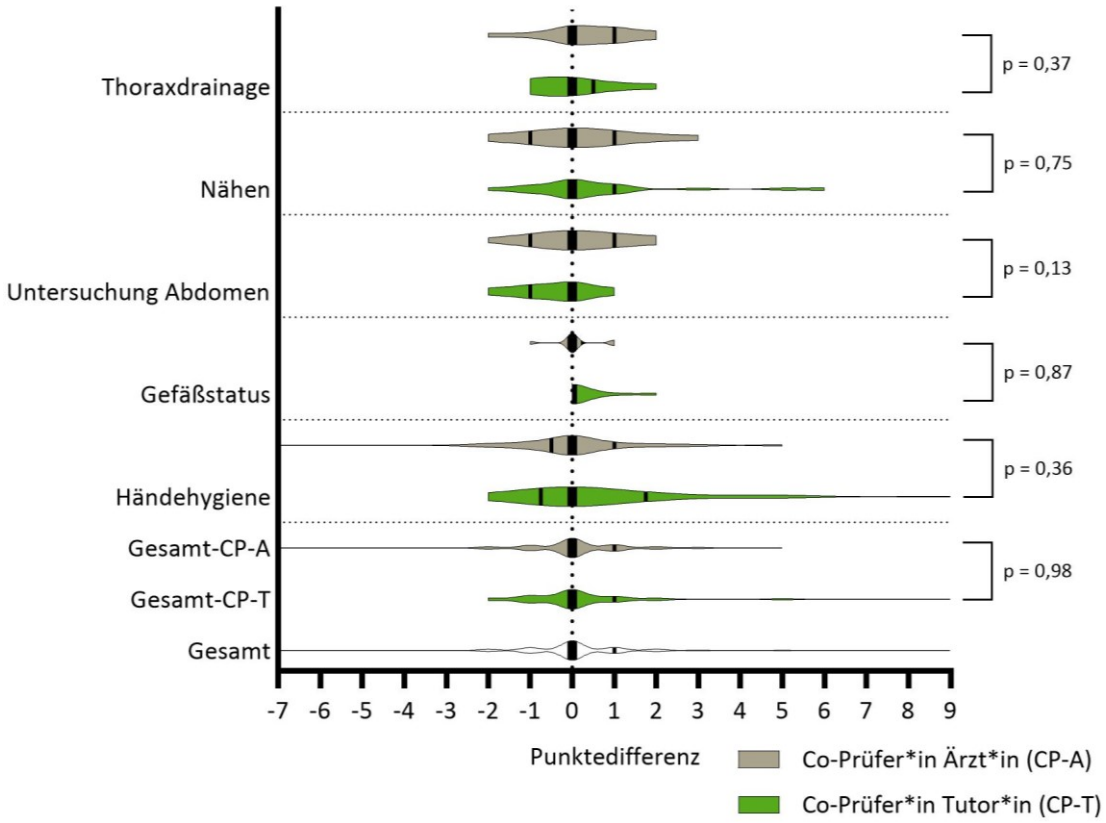


Abbildung 12: Tendenziell stimmen HP und CP mit der gegebenen Punktzahl überein. Durchschnittliche Differenz der vergebenen Punktzahl von Hauptprüfer*in (HP) und Co-Prüfer*in (CP) bei Praxis-Checks im Skills Lab der Universität Ulm im Sommersemester 19 und Wintersemester 19/20. Lesebeispiel: Rechnung: HP − CP = negativ, CP hat mehr Punkte gegeben als HP. Median ± Interquartilsabstand; Gesamt n = 277, Median = 0; Gesamt-CP-A n = 198, Median = 0, Gesamt-CP-T n = 79, Median = 0; Hygiene CP-A n = 61, Median=0, CP-T n = 28, Median = 0; Gefäßstatus CP-A n = 10, Median = 0, CP-T n = 12, Median = 0; Untersuchung Abdomen CP-A n = 19, Median = 0, CP-T n = 19, Median = 0; Nähen CP-A n = 19, Median = 0, CP-T n = 19, Median = 0; Thoraxdrainage CP-A n = 13, Median = 0, CP-T n = 6, Median = 0; Statistik: Mann-Whitney-Test

Die bisherigen Ergebnisse zeigten, dass die HP und CP unter anderem in der Durchschnittspunktzahl übereinstimmten. Als nächster Schritt wurde die prozentuale Übereinstimmung zwischen den Prüfenden für jeden Punkt ausgewertet, um mögliche Abweichungen zu ermitteln (Abbildung 13). Bereits im Median erreichten die Prüfenden Übereinstimmungswerte von 90 % und besser. Dies entspricht 18 von 20 gleich gewerteten Punkten. Die Übereinstimmung zwischen den Prüfer*innen lag für die einzelnen Punkte 1 – 20 meist bei Werten über 80 %. In 75 % der Fälle wurden sogar Werte über 90 % erreicht. Auffällig gut waren die Übereinstimmungswerte des PC Gefäßstatus (n = 22). Nur die Punkte 3, 5, 13 und 16 erreichten keine Übereinstimmungswerte von 100 %.

Auffällig unterschiedlich wurde Punkt 13 vom PC Nähen „Einstich und Ausstich im 90°-Winkel zum Wundverlauf, je ca. 3 – 10 mm Abstand von der Wunde“ mit 55 % Übereinstimmung gewertet. Punkt 16, 17 und 19 des PC Nähen wiesen ebenfalls Abweichungen auf, hier ergab sich eine Übereinstimmung von 70 %.

Daraufhin stellte sich die Frage, ob die Unstimmigkeiten darauf zurückzuführen waren, dass eher HP oder CP die Punkte vergeben hatten (s. Anhang 7.1.5). In den meisten Fällen konnte nicht geklärt werden, wer eher welche Punkte vergab.

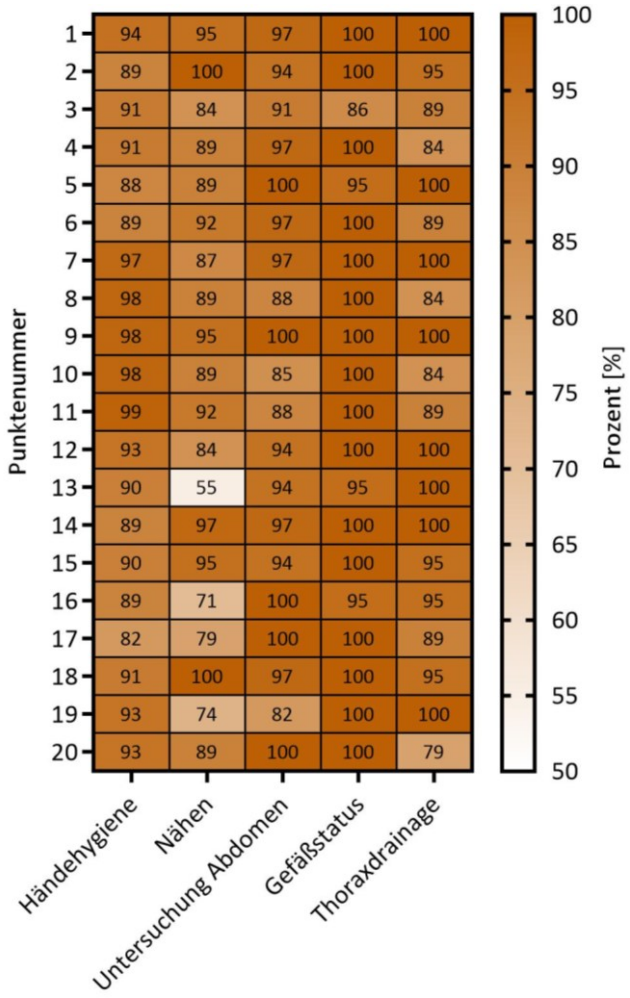


Abbildung 13: Prozentuale Übereinstimmung von Hauptprüfer*in und Co-Prüfer*in je Punkt für die 5 häufigsten Praxis-Checks im Skills Lab der Universität Ulm im Sommersemester 19 und Wintersemester 19/20. Lesebeispiel: Händehygiene Punkt 1: 94 % der Hauptprüfer*innen und Co-Prüfer*innen stimmten in ihrer Bewertung des Punktes überein. Hygiene n = 88; Nähen n = 38; Untersuchung Abdomen n = 33; Gefäßstatus n = 22; Thoraxdrainage n = 19

3.3 Vergleich der Bewertung mehrerer Prüfer*innengruppen durch Online-Praxis-Check-Simulation

In den vorausgegangenen Untersuchungen konnte jeweils nur ein Teil der Prüfenden direkt miteinander verglichen werden. Alle miteinander zu vergleichen, ließ sich nicht umsetzen, da die Prüfer*innen unterschiedliche PCs bewertet hatten. Zusätzlich war es nicht möglich, eine Aussage zur Qualität der Übereinstimmung zu treffen, da nicht analysiert werden konnte, ob HP oder CP oder keiner von beiden korrekt gewertet hatte. Mit der Online-PC-Simulation wurde versucht, die Probleme der vorhergegangenen Interventionen zu vermeiden und die aufgekommenen Fragen zu beantworten. In diesem Studienteil wurde es mittels Verwendung von Filmen als einheitliches Prüfmaterial möglich gemacht, alle Prüfer*innen miteinander zu vergleichen, da alle Prüfer*innen die identischen PCs zu bewerten hatten. Um mögliche Unterschiede in Skill- und Kontextpunkten zu finden, wurden PCs mit gezielten Fehlerschwerpunkten erstellt. Zusätzlich wurden Studierende als weitere Vergleichsgruppe hinzugefügt, um einen besseren Vergleich zu ermöglichen.

In der Analyse der Ergebnisse wurde sich immer auf die korrekt gewerteten Punkte der bewerteten PCs bezogen. 20 von 20 erreichte Punkte bedeutete, dass der*die Prüfer*in alle, sowohl fehlerhaft ausgeführte als auch richtig ausgeführte, Punkte korrekt und zutreffend bewertet hatte.

3.3.1 Datensatz

Insgesamt nahmen an der Online-Simulation 74 Proband*innen teil. Einige Prüfer*innen mussten aus der Studiengruppe entfernt werden. Vier Prüfer*innen brachen nach dem Eingangsfragebogen ab, vier Prüfer*innen bearbeiteten nur den ersten Film, ein Prüfer war zum Bearbeitungszeitpunkt bereits im PJ und ein Prüfer spulte die Filme vor und musste aus diesen Gründen entfernt werden. An der Studie nahmen 24 Studierende, 23 Tutor*innen und 18 Ärzt*innen teil.

In der Gruppe der Studierenden fanden sich 10 männliche und 14 weibliche Prüfer*innen (s Anhang 7.1.6). Diese hatten keine Berufserfahrung (1(1;1), Skala: 1 = keine Erfahrung; 2 = 1 Jahr; 3 = 2 Jahre; 4 = 3 Jahre; 5 = +3 Jahre) und befanden sich im Median im 10. (9.;10.) Semester. 58 % der Studierenden gaben an, Erfahrungen als Tutor*in vor allem im Präparierkurs (20 %), in der Physiologie (15 %) und in Pharmakologie-POL (15 %) zu haben

(s. Anhang 7.1.7). Erfahrungen als Prüfer*in gaben 50 % und Erfahrungen mit dem PC Untersuchung Abdomen 71 % an.

Die Studiengruppe der Tutor*innen setzte sich aus 9 Tutoren und 14 Tutorinnen zusammen. Im Median befanden sich diese im 8. (6.;9.) Semester, hatten ebenfalls keine Berufserfahrung (1(1;1), Skala: 1 = keine Erfahrung; 2 = 1 Jahr; 3 = 2 Jahre; 4 = 3 Jahre; 5 = +3 Jahre) und waren 23 (22;24) Jahre alt. 96 % (n = 22) gaben Erfahrungen als Prüfer*in und 91 % Erfahrungen mit dem PC Untersuchung Abdomen an.

Die ärztliche Gruppe bestand aus 8 männlichen und 10 weiblichen Ärzt*innen. Im Median waren diese 30 (28;55) Jahre alt und hatten 3-5 (3(1,75;5), Skala: 1 = 1 Jahr; 2 = 2 Jahre; 3 = 3-5 Jahre; 4 = 6-10 Jahre; 5 = 10+ Jahre) Jahre Berufserfahrung. Erfahrungen als Prüfer*in wiesen 89 % vor, allerdings hatten nur 22 % Erfahrung mit dem PC Untersuchung Abdomen. Am häufigsten vertreten war der Fachbereich Chirurgie (29 %) gefolgt von der Inneren Medizin (24 %) und Gynäkologie (18 %) (s. Anhang 7.1.8).

3.3.2 **Analyse der Fragebögen**

Der Eingangsfragebogen enthielt einige Fragen zur Selbsteinschätzung der Teilnehmenden (6-stufige Likert-ähnliche Skala: (−−−) = trifft gar nicht zu, (+++) = trifft völlig zu) (s. Anhang 7.1.9). Interessanterweise schätzen sich alle 3 Prüfer*innengruppen ähnlich ein. Die Fragen „Ich halte mich für eine*n objektive*n Prüfer*in“, „Ich fühle mich grundsätzlich sicher darin, studentische Leistungen in praktischen Fertigkeiten einzuschätzen“ und „Ich fühle mich dazu befähigt, die Qualität einer durchgeführten Abdomen Untersuchung zu beurteilen“ wurden alle mit „trifft zu“ bewertet. Nur die Frage „Ich fühle mich grundsätzlich dazu befähigt, eine strukturierte Abdomen Untersuchung durchzuführen“ wurde von den Ärzt*innen, im Gegensatz zu den anderen Gruppen („trifft zu“), mit „trifft völlig zu“ bewertet.

In der Abschlussbefragung wurde „Das Werkzeug "Praxis-Check" halte ich für geeignet, um klinisch-praktische Leistungen zu bewerten“ mit mindestens „trifft zu“ bewertet (s. Anhang 7.1.10). Die Frage „Das Werkzeug der Praxis-Checks halte ich für geeignet, geschulten studentischen Peer Teachern das Prüfen zu ermöglichen“ wurde ebenfalls mit „trifft zu“ bewertet. Die Frage „Die Bewertung der einzelnen Prüfungsfilme fiel mir leicht“ wurde von den Gruppen mit überwiegend „trifft zu“ bewertet. Die Übereinstimmung innerhalb der

Gruppe der Tutor*innen und Ärzt*innen wurde außer von den Tutor*innen mit „trifft zu“ bewertet. Die Tutor*innen bewerteten ihre gruppeninterne Übereinstimmung als sehr hoch („trifft völlig zu“) und die der Ärzt*innen nur als mäßig gut („trifft zu“).

3.3.3 Vergleich der Prüfer*innengruppen

Um eine erste Einschätzung der Ergebnisse zu erhalten, wurden von den drei Gruppen die Prozentwerte der maximal erreichbaren korrekt gewerteten Punkte ermittelt (Abbildung 14). Mit im Median 94 % (95 %;90 %) korrekt gewerteten Punkten erreichten die Studierenden den besten Wert gefolgt von den Tutor*innen mit 93 % (97 %;89 %). Die größte Differenz zeigte sich zu der ärztlichen Gruppe, welche einen medialen Punktewert von 91 % (94 %;88 %) erreichte.

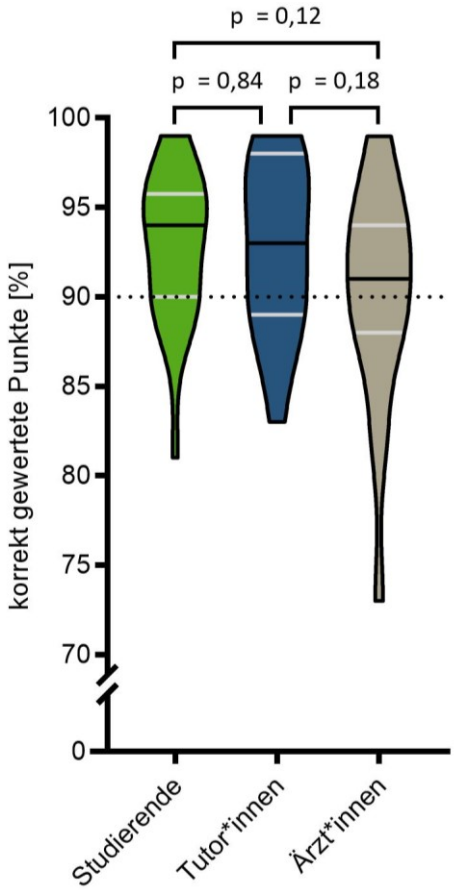


Abbildung 14: Im Median wertet alle Prüfer*innengruppen über 90 % der Punkte korrekt. Korrekt gewertete Punkte der Prüfer*innengruppen in Prozent in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm; Zeitraum: Sommersemester 20; Lesebeispiel: Im Median bewerteten die Studierenden in vier Filmen 94 % der Punkte korrekt. Median ± Interquartilsabstand; gestrichelte Linie: gute*r Prüfer*in bei > 90 % korrekt gewertete Punkte; Statistik: Kurskal-Wallis-Test mit unkorrigiertem Dunn-Test

Bei genauerer Betrachtung der einzelnen PCs fand sich bei allen Gruppen ein Median von 19 Punkten (Abbildung 15). Vor allem die Studierenden und Tutor*innen erreichten für jeden Film den gleichen Punktewert. Film 2 (Fehlerfrei-Variante) wurde von allen Gruppen im Median mit 20 korrekt gewerteten Punkten am besten bewertet. Dies entsprach 100 % korrekt gewerteten Punkten. Bei Film 4 (Mischfehler-Variante) wurden von Studierenden und Tutor*innen im Median 18 Punkte korrekt vergeben. Die ärztlichen Prüfer*innen erreichten nur 17 korrekt gewertete Punkte. In Film 1 (Skillfehler-Variante) erreichten alle Gruppen 18 korrekt gewertete Punkte. Film 3 (Kontextfehler-Variante) wurde von Tutor*innen und Studierenden (19 Punkte) besser bewertet als von den Ärzt*innen (18 Punkte).

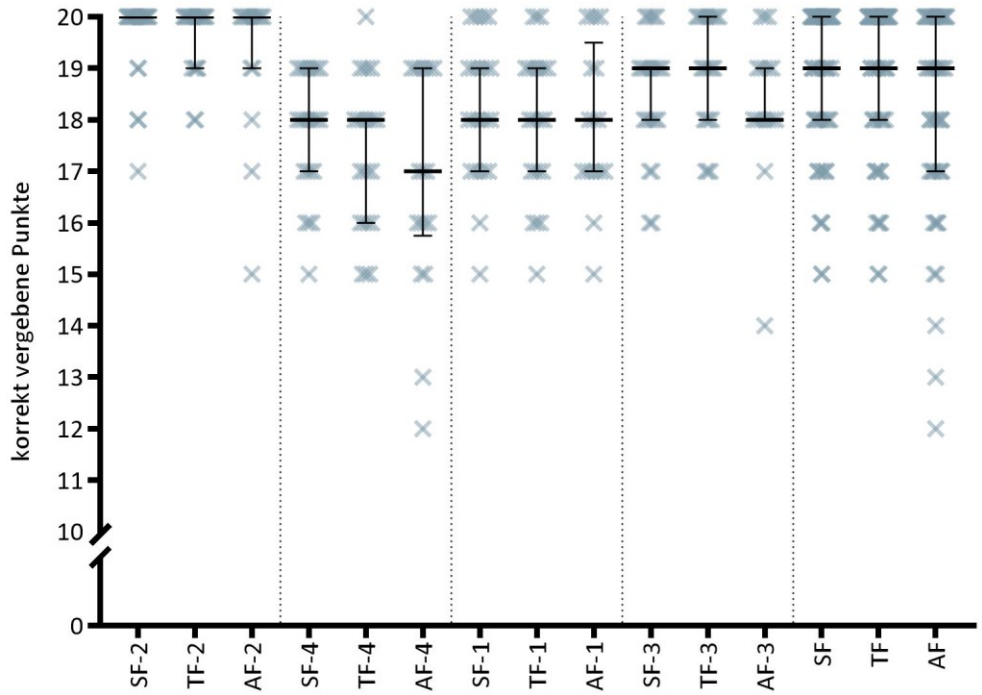


Abbildung 15: Die Prüfergruppen erreichen vergleichbare Anzahlen korrekt vergebener Punkte je Film. Performanz-Vergleich von Tutor*innen, Studierenden und Ärzte*innen. Anzahl korrekt vergebener Punkte je Film in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm; Zeitraum: Sommersemester 20; Median ± Interquartilsabstand; SF = Studierende, TF = Tutor*innen, AF = Ärzte*innen, F = Film 1-4

Interessant war, dass die Gruppen bei der Kontextfehler-Variante einen besseren Wert erreichten als bei der Skillfehler-Variante. Kontrollierte man die Gruppen auf einen statistisch signifikanten Unterschied, konnte nichts gefunden werden. Sowohl die Prozentwerte der Gesamtpunktzahl als auch die medianen Punktewerte der PCs wiesen keinen signifikanten Unterschied auf.

Alle drei Gruppen konnten im Median als gute Prüfer*innen eingestuft werden, da die erforderlichen 90 % oder 18 von 20 Punkten von allen erreicht wurden. Nur bei der Mischfehler-Variante hatten die Ärzt*innen etwas mehr Schwierigkeiten, die Prüfung zutreffend zu bewerten.

3.3.4 **Vergleichbarkeit innerhalb der Gruppen**

Betrachtete man die Gruppen einzeln für sich, so fielen tendenziell homogene Ergebnisse auf (Anhang 7.1.11). Die homogenste Gruppe bildeten die Studierenden. Die Differenz, bezogen auf korrekt gewertete Punkte, war innerhalb der ärztlichen Gruppe am größten.

Ein Großteil der Studierenden wertete über 90 % der Punkte korrekt. Insgesamt schwankten die Werte der korrekt ermittelten Punkte vor allem zwischen 88 % und 99 %. Von den 24 Studierenden erreichten fünf nicht die 90 % Grenze, womit diese nicht als gute Prüfer*innen definiert werden konnten.

Die Tutor*innen zeigten sich ebenfalls relativ homogen. Allerdings etwas stärker different als die Studierenden. Ebenfalls werteten die meisten Tutor*innen über 90 % der Punkte korrekt. Die Bandbreite schwankte hier zwischen 85 % und 99 % korrekt gewerteter Punkte. Allerdings erreichten von den 23 Tutor*innen acht nicht die 90 % Grenze. Im direkten Vergleich zu den Studierenden waren diese etwas besser. Die Tutor*innen wurden untereinander noch nach Erfahrung aufgeteilt. Tutor*innen mit mehr als zehn PCs wurden als “Erfahren” definiert. Tendenziell zeigte sich ein Zusammenhang von Erfahrung und der korrekt gewerteten Punktezahl (P (two-tailed) = 0,06; r = 0,15). Unerfahrenere Tutor*innen (< 10 PCs) erreichten im Median 90 % korrekt gewertete Punkte, während erfahrenere Tutor*innen 93 % korrekt werteten und somit ein minimal besseres Prüfer*innenergebnis erzielten (s. Anhang 7.1.12).

Die differenteste Gruppe bildeten die Ärzt*innen. Werte zwischen 82 % und 99 % wurden erreicht. Allerdings erreichte nur knapp mehr als die Hälfte der 18 Ärzt*innen die 90 % Marke. Des Weiteren auffällig war das Verhältnis des Alters zur korrekt gewerteten Prozentpunktzahl je Ärzt*in: Niedrigere Werte wurden tendenziell von älteren Ärzt*innen erreicht (P (two-tailed) = 0,01; r = 0,35). Der mediane Punktewert der Ärzt*innen bis 30 Jahren betrug 94 %. Die älteren Ärzte*innen hingegen erreichten im Median nur einen Punktewert von 88 % (Anhang 7.1.13).

Zusammenfassend kann man sagen, dass unter den Gruppen und innerhalb der Gruppen kleine Unterschiede zu finden waren. Allgemein betrachtet war es für die geprüfte Person irrelevant, wer sie prüfte, da alle ein ähnliches Ergebnis erzielten.

3.3.5 Bewertung der vier Filmvarianten im Detail

In diesem Kapitel wird die Bewertung der einzelnen Punkte des PC Untersuchung Abdomen genauer betrachtet (Abbildung 16). Es stellte sich die Frage, welche Punkte eher falsch gewertet wurden und wie gut Fehler von den Prüfer*innen erkannt wurden. Die Werte zeigen den Prozentsatz des jeweiligen Punktes, welcher von der Gruppe korrekt gewertet wurde.

Film 2: Fehlerfrei-Variante

Fast alle Punkte wurden von den drei Vergleichsgruppen in den meisten Fällen zu 100 % korrekt bewertet. Der einzige Punkt, bei dem die Studierenden kleine Schwierigkeiten hatten, war Punkt 8 Loslass-Schmerz. Dieser wurde nur von 88 % korrekt gewertet. Die Ärzt*innen erreichten bei Punkt 8 nur 82 %. Zusätzlich wurden bei den Punkten 18 und 19 nur 88 % korrekt gewertet. Die Tutor*innen erreichten, ähnlich zu den Studierenden, sehr gute Werte. Nur Punkt 17 (83 %) wurde schlechter bewertet als die restlichen Punkte. Generell war Film 2 der von allen Gruppen am korrektesten bewertete Film.

Film 4: Mischfehler-Variante

Richtig durchgeführte Punkte wurden mehrheitlich von allen Gruppen korrekt gewertet. Bei fehlerhaften Punkten waren die Werte, verglichen zu den richtigen Punkten, oft etwas erniedrigt. Die richtigen Punkte 2, 4, 5, 6, 7, 9, 11, 13, 14 und 17 wurden von allen Gruppen zu > 90 % korrekt erkannt. Von den fehlerhaften Punkten wurden die Punkte 1, 15, 16 und 20 ebenfalls korrekt erkannt. Bei Punkt 1 (Begrüßung des Patienten) hatten nur die Ärzt*innen Probleme, den Fehler zu erkennen und werteten diesen nur in 60 % der Fälle korrekt. Der am schlechtesten bewertete Punkt war Punkt 3 (hygienische Händedesinfektion). Der Schauspieler trug am linken Handgelenk ein gut sichtbares Armband (Abbildung 3b), wodurch die hygienische Händedesinfektion fehlerhaft war. Dieser Fehler wurde von den Studierenden nur in 29 %, bei den Tutor*innen und Ärzt*innen sogar nur in 17 % der Fälle korrekt gewertet. Der fehlerhafte Punkt Untersuchung Leber (Punkt 10) erreicht mit ca. 60 % von allen Gruppen ebenfalls nur einen

schlechten Wert. Die Punkte 8, 12 und 18 sind fehlerhafte Punkte, die aber meistens (ca. zu 80 %) korrekt erkannt wurden. Punkt 19 (Dokumentation Palpation) ist der einzige Punkt in diesem Film, der, obwohl er richtig durchgeführt wurde, nur in ca. 80 % erkannt wurde. Vor allem die Ärzt*innen schienen sich mit den Punkten 1, 18 und 19 schwer zu tun. Diese fehlerhaften Punkte wurden von ihnen in nur 67 % der Fälle erkannt.

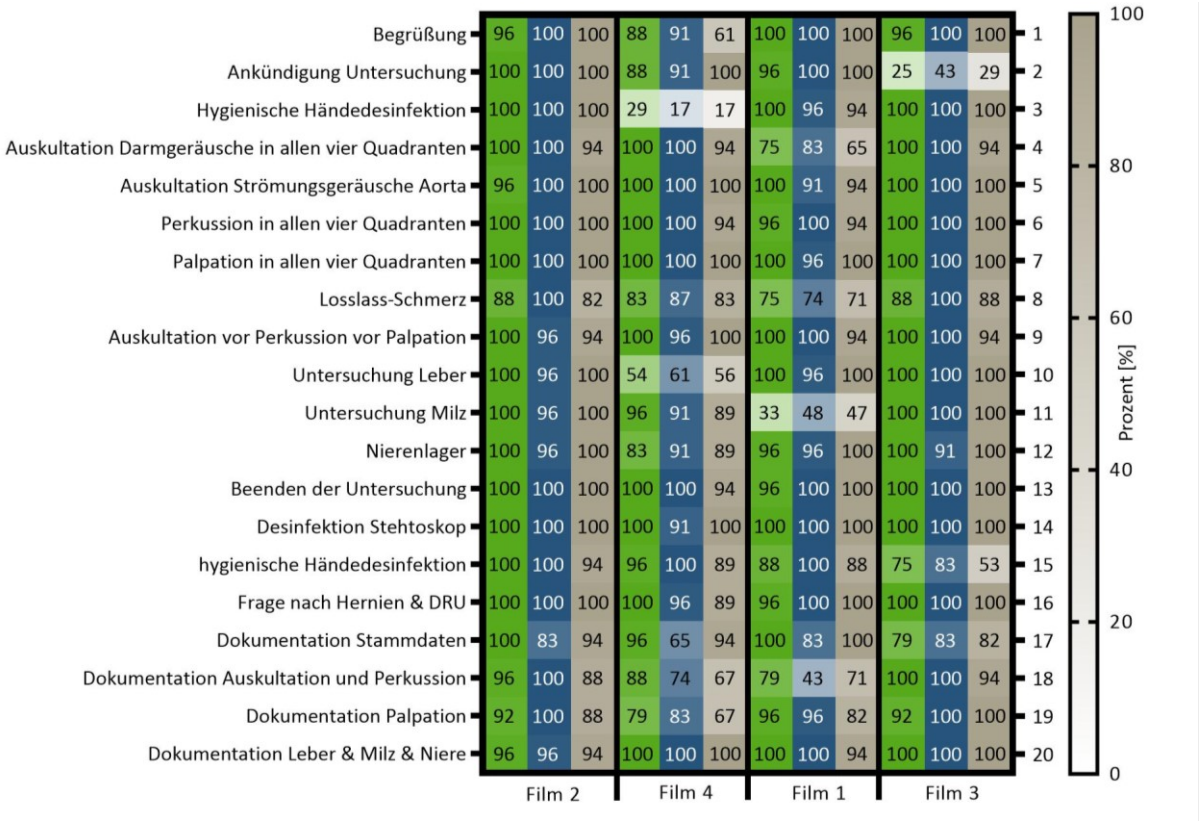


Abbildung 16: Prozent korrekt gewerteter Punkte je Gruppe in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. grün = Studierende, blau = Tutor*innen, braun = Ärzt*innen; Zeitraum: Sommersemester 20; Lesebeispiel: Begrüßung Film 2: 96 % der Gruppe der Studierenden werteten diesen Punkt korrekt.

Tabelle 2: richtig und falsch demonstrierte Punkte in den Filmen in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm:
grün = richtig durchgeführter Punkt, rot = fehlerhafter durchgeführter Punkt; Film 1 Punkt 18 nur bei Tutor*innen fehlerhaft; Zeitraum: Sommersemester 20

Punkt	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Film 2																				
Film 4																				
Film 1																				
Film 3																				

Film 1: Skillfehler-Variante

Wieder fällt auf, dass richtig durchgeführte Punkte von allen Gruppen meistens korrekt erkannt und bewertet wurden. Von den fehlerhaft durchgeführten Punkten wurde nur Punkt 9 korrekt erkannt. Die Punkte 4 und 11 bereiteten den Gruppen mehr Probleme. Die Untersuchung der Milz (Punkt 11) wurde von den Tutor*innen und Ärzt*innen in ca. 47 % der Fälle korrekt bewertet, von den Studierenden allerdings nur in 33 % der Fälle. Des Weiteren hatten die Gruppen Probleme, Punkt 8 Loslass-Schmerz als richtig durchgeführt zu erkennen. Die hygienische Händedesinfektion wurde von einigen Studierenden und Ärzt*innen ebenfalls als fehlerhaft gewertet, obwohl diese richtig durchgeführt wurde. Punkt 18 (Dokumentation Auskultation und Perkussion) wurde auffallend schlecht von den Tutor*innen (43 %) bewertet. Im Gegensatz zu den anderen Gruppen war bei den Tutor*innen im Befundbogen tatsächlich ein Fehler zu finden, der allerdings für die anderen Gruppen verbessert wurde. Trotzdem werteten auch die anderen Gruppen diesen Punkt auffällig selten korrekt.

Film 3: Kontextfehler-Variante

Die richtigen Punkte wurden gruppenübergreifend mehrheitlich korrekt gewertet. Punkt 8 (Loslass-Schmerz) bereitete den Studierenden und Ärzt*innen wieder Probleme. Nur 88 % werteten diesen Punkt korrekt. Punkt 2 (Ankündigung der Untersuchung) wurde selten korrekt bewertet. Studierende werteten 25 %, Ärzt*innen 29 % und Tutor*innen nur 43 % der Punkte korrekt. Punkt 15 (hygienische Händedesinfektion) und 17 (Dokumentation Stammdaten) erreichten Werte um 80 %. Die Ärzt*innen werteten Punkt 15 nur in 53 % der Fälle korrekt.

Im Gesamten betrachtet fällt auf, dass einige Fehler, unabhängig davon, ob der Punkt richtig oder falsch ausgeführt wurde, schlechter erkannt wurden als andere. Punkt 8 (Loslass-Schmerz) wurde vor allem von den Studierenden und Ärzt*innen jedes Mal schlechter bewertet. Es wurden immer Werte um die 80 % erreicht. Punkt 17 (Dokumentation der Stammdaten) wurde von den Tutor*innen in jedem Film, öfters als von den anderen beiden Gruppen, fehlerhaft gewertet.

Die Ergebnisse der Studierenden, Tutor*innen und Ärzt*innen waren allerdings in den meisten Fällen sehr ähnlich. In den meisten Punkten fand sich kein relevanter Unterschied

zwischen den Gruppen. Vor allem die Studierenden und Tutor*innen stimmten meist stark miteinander überein.

3.3.6 Analyse der Skill- und Kontextpunkte

Insgesamt wurden Skill- und Kontextfehler von den Gruppen gleich häufig erkannt (s. Anhang 7.1.3). Auffällig war, dass alle Gruppen richtig gemachte Punkte deutlich häufiger korrekt werteten als fehlerhafte durchgeführte Punkte. Die Gruppen hatten also eher Schwierigkeiten fehlerhafte Punkte korrekt zu bewerten als richtig durchgeführte Punkte. Tendenziell wurden Skillfehler besser erkannt als Kontextfehler. Bei der Mischfehler-Variante hatten alle Gruppen ebenfalls tendenziell eher Problem mit den Kontextfehlern. Die ärztliche Gruppe hatte im Vergleich zu den Studierenden und Tutor*innen zum Teil noch mehr Probleme, Kontextfehler korrekt zu bewerten. Bei richtig durchgeführten Punkten fand sich kein Unterschied zwischen Skill- und Kontextpunkten.

Dokumentationspunkte

Punkt 18 in Film 1 der Tutor*innen konnte nicht mit den anderen Gruppen verglichen werden, da die Tutor*innen einen organisationstechnisch fehlerhaften Anamnesebogen hatten (Abbildung 16). Allerdings zeigte das Ergebnis, dass die Tutor*innen in diesem Punkt nicht homogen auftraten. Auch die Dokumentation der Stammdaten schien vor allem bei den Tutor*innen, verglichen zu den anderen Gruppen, ein Schwachpunkt zu sein. Studierende und Tutor*innen erreichten ähnlichere Prozentwerte korrekt gewerteter Punkte als im Vergleich zur ärztlichen Gruppe.

Kommunikationspunkte

Prinzipiell fand sich unter den Gruppen eine homogene Übereinstimmung in allen Kommunikationspunkten. Lediglich die Begrüßung (Punkt 1) in Film 4 wurde von den Ärzt*innen (61 %) deutlich schlechter gewertet als von den anderen beiden Gruppen (Studierende 88 %; Tutor*innen 91 %). Des Weiteren wurde die Ankündigung der Untersuchung (Punkt 2) in Film 3 deutlich besser von den Tutor*innen erkannt und bewertet als von den Vergleichsgruppen. Die Tutor*innen werteten in 43 % den Punkt korrekt, während Studierende 25 % und Ärzt*innen 29 % erreichten. Insgesamt prüften die Tutor*innen präziser und genauer als die Studierenden und Ärzt*innen.

Hygienepunkte

Ähnlich zu den anderen Kontextpunkten wurden die hygienisch richtig durchgeführten Punkte von allen Gruppen meist korrekt erkannt. Interessant war die Bewertung der richtig ausgeführten hygienischen Händedesinfektion (Punkt 15) in Film 1. Sowohl Studierende als auch Ärzt*innen werteten dort Fehler. Wieder fällt auf, dass fehlerhafte Punkte allgemein schlechter korrekt gewertet wurden. Insgesamt hatten die Ärzt*innen mehr Hygienefehler übersehen als die anderen Gruppen.

3.3.7 Übereinstimmung der Prüfer*innen

Die Übereinstimmung zwischen den Prüfer*innen wurde mit Hilfe der ICC berechnet. Es wurden die Übereinstimmungswerte allgemein und innerhalb der Prüfer*innengruppen berechnet.

Tabelle 3: Berechnete Intraklassen-Korrelation (ICC) in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm allgemein und je Prüfer*innengruppe. Zeitraum: Sommersemester 20; ICC (KI = 95 % Konfidenzintervall (Untergrenze; Obergrenze)

	Allgemein	Studierende	Tutor*innen	Ärzt*innen
ICC	0,97 (KI 0,90;1)	0,91 (KI 0,73;0,99)	0,94 (KI 0,82;1)	0,81 (KI 0,38; 0,97)

3.3.8 **Relevanz für die Geprüften**

Um die Relevanz für die Geprüften einschätzen zu können, wurde die vergebene Punktzahl der Prüfer*innen von der Zielpunktzahl abgezogen (Abbildung 17). In Film 1 und 2 zeigte sich im Median keine Abweichung zur Zielpunktzahl. Die Filme 3 und 4 wurden im Gegensatz dazu im Median einen Punkt besser bewertet als die anderen Filme. Dem Prüfling wurde also eher ein Punkt zu viel gegeben. Ob und welche Punkte im PC fehlerhaft gewertet wurden, spielt in dieser Analyse keine Rolle. Allerdings kann man sagen, dass die von den Prüfer*innen vergebenen Punktzahlen gleich oder zumindest nah an der Zielpunktzahl lagen und in den meisten Fällen kein Nachteil für die geprüfte Person entstanden wäre.

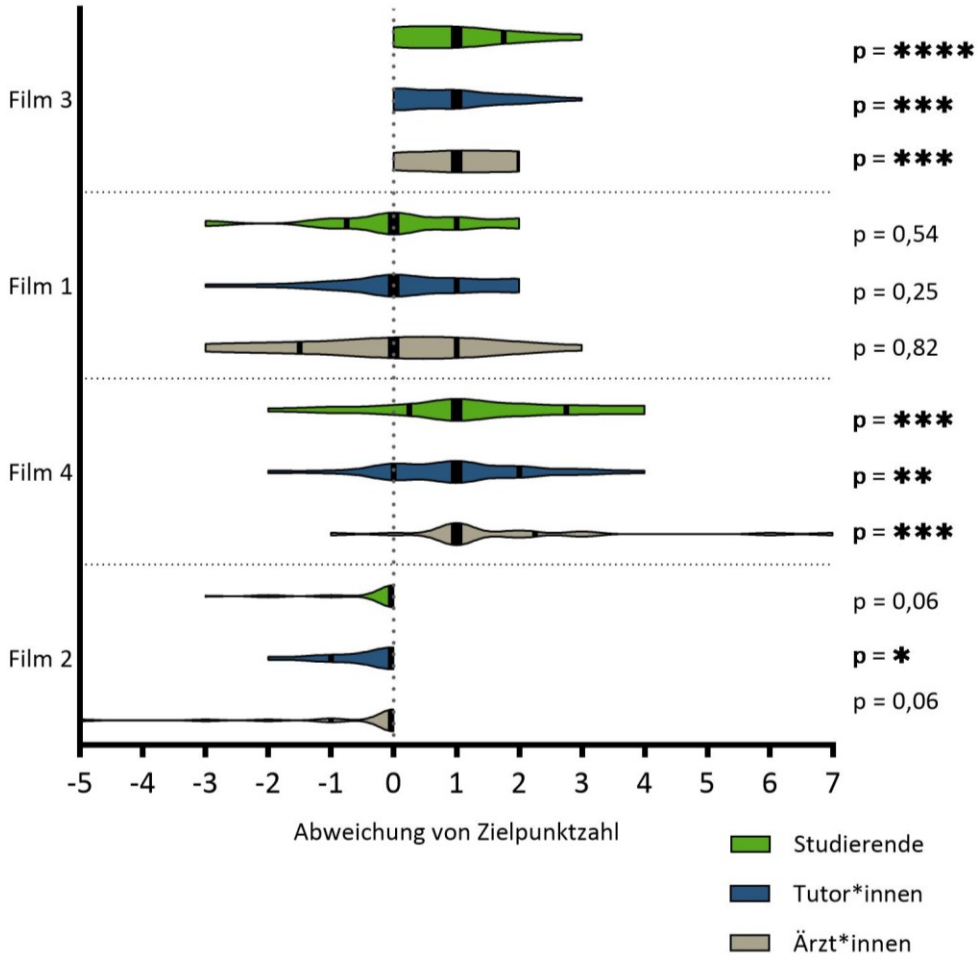


Abbildung 17: Die maximale mediane Differenz zur Zielpunktzahl liegt bei 1 Punkt. Abweichung der vergebenen Punktzahl von der Zielpunktzahl (gegebene Punktzahl - Zielpunktzahl) in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm; Zeitraum: Sommersemester 20; Median ± Interquartilsabstand; ✱✱✱✱ ≙ p < 0,0001, ✱✱✱ ≙ p < 0,001, ✱✱ ≙ p < 0,01, ✱ < p 0,05; Statistik: One sample Wilcoxon-Test

4 Diskussion

4.1 Urteiler*innenübereinstimmung bei Praxis-Checks mit einem*r Prüfer*in

In der Gesamtauswertung der erhoben PC-Daten zeigten die prüfenden Tutor*innen tendenziell ein, auf die Durchschnittspunktzahl bezogenes, ähnliches Prüfverhalten. Die medianen Durchschnittspunktzahlen der Tutor*innen schwankten zwischen 19 und 20 Punkten. Die Anzahl der abgenommen PCs (> 10 PCs) hatte keinen erkennbaren Effekt auf eine breitere Streuung der Durchschnittspunktzahlen. Das gleiche Ergebnis zeigte sich bei der Betrachtung der einzelnen PCs. Die vergebenen Durchschnittspunktzahlen von den Tutor*innen befanden sich tendenziell maximal 5 % von der allgemeinen Durchschnittspunktzahl entfernt. Mit der Annahme, dass die Tutor*innen korrekt gewertet hatten und eine Prüfer*innenleistung von 90 % korrekt gewerteten Punkten als gut definiert wurde, könnte die Abweichung von 5 % vom allgemeinen Durchschnitt als gute Übereinstimmung gewertet werden. Eine mögliche Erklärung für die übereinstimmenden Werte könnte die einheitliche Ausbildung der Tutor*innen sein. Offenbar wurde durch die regelmäßige Weiterbildung die zu beachtenden Punkte von den Tutor*innen gut und einheitlich erkannt. Durch Erfahrung mehrerer PC-Prüfungen wurden die Tutor*innen auch auf seltene Fehler aufmerksam. Durch Feedback und Erfahrung von anderen Tutor*innen und der medizinischen Leitung des Skills Lab in Kombination mit einem klaren Kriterienkatalog war es erklärbar, dass die Tutor*innen ein weitgehend übereinstimmendes Prüfverhalten präsentierten.

Diese Annahme hat allerdings mehrere und durchaus starke Limitationen. Die Leistung der bewerteten Studierenden war nicht bekannt. Somit konnte nicht geklärt werden, ob die Punkte, welche vergeben wurden, korrekt oder inkorrekt gewertet wurden. Tutor*innen konnten so nur anhand der Durchschnittspunktzahl verglichen werden in der Annahme, dass durch die größere Anzahl an Prüfungen fehlerhaft gewertete Punkte die Aussage nicht zu stark verzerren würden. Des Weiteren durften nicht alle angebotenen PCs von allen Tutor*innen geprüft werden. Häufig abgelegte und von vielen Tutor*innen geprüfte PCs hatten somit eine bessere Aussagekraft bezogen auf die Übereinstimmung der Durchschnittspunktzahlen als seltenere geprüfte PCs.

Um die angesprochenen Limitationen auf die allgemeine Durchschnittspunktzahl der Tutor*innen etwas zu minimieren, wurde die Auswertung auch nach einzelnen PCs erstellt. Unterschiedliche schlecht vergleichbare Schwierigkeiten und Aufgabenstellungen der PCs konnten somit eliminiert werden. Dann zeigte sich: Auch bei gleichen PCs wiesen die Tutor*innen eine gute Übereinstimmung der Durchschnittspunktzahlen auf. Allerdings erreichte man durch die oft geringe Anzahl der gleichen PCs pro Prüfer*in nur eingeschränkt repräsentative Ergebnisse.

Die allgemeine Auswertung der PCs über mehrere Semester ermöglichte es, einen ersten Überblick über die Situation im Skills Lab der Universität Ulm zu erhalten und die folgenden Studienteile besser zu planen.

4.2 Urteiler*innenübereinstimmung von Haupt- und Co-Prüfer*innen bei realen Praxis-Checks

4.2.1 Zusammenfassung der Ergebnisse

Um das Prüfverhalten der Tutor*innen als HP besser analysieren zu können, wurde ein Teil der Prüfungen von einem*r weiteren Prüfer*in parallel bewertet. Als CP wurden andere Tutor*innen und Ärzt*innen eingesetzt. So konnten bei jedem PC zwei Prüfer*innen (HP und CP) miteinander verglichen werden. Ein Vergleich aller Tutor*innen war wie in der vorangegangenen Analyse wieder nur bedingt möglich, allerdings war die Vergleichbarkeit der Prüfer*innen im Vergleich zum ersten Studienteil höher.

Im allgemeinen Vergleich stimmten die Bewertungen der Tutor*innen, CP-T und CP-A hinsichtlich der Durchschnittspunktzahl überein. Die rohe Prüfer*innenübereinstimmung zwischen HP und CP lag bei den betrachteten PCs bei ≥ 90 %. Dies bedeutete, dass tendenziell 18 von 20 Punkten identisch von HP und CP bewertet wurden. In der Analyse der einzelnen Punkte fiel das Ergebnis ähnlich aus. Der Großteil der Punkte wurde von HP und CP identisch bewertet. Die direkten Übereinstimmungswerte befanden sich für die meisten Punkte bei ≥ 90 %.

4.2.2 Bewertung der Ergebnisse

Bezogen auf die Gesamtpunktzahl des PC zeigte sich, dass HP und CP tendenziell miteinander übereinstimmten. Es schien sich so eher die Annahme aus der Analyse von Teil

1 zu bestätigen, dass die Bewertungen der PCs innerhalb der Tutor*innen vergleichbar waren. Zusätzlich konnte auch im Vergleich mit den Ärzt*innen eine hohe Übereinstimmung der Bewertungen gefunden werden.

Da approbierte Ärzt*innen in OSCEs als Prüfer*innen eingesetzt werden, wurden diese als Sollwert für die Tutor*innen bei den PCs verwendet [50; 51]. Die Übereinstimmung von HP und CP-A wies darauf hin, dass die Qualität der Bewertung der Tutor*innen gut, beziehungsweise dem Niveau der Ärzt*innen entsprach. Interessanterweise schienen HP tendenziell mehr Punkte zu vergeben als CP. Dies könnte unter anderem daran liegen, dass HP sich um mehrere Dinge gleichzeitig kümmern mussten. Anweisungen mussten vorgelesen werden, Stoppuhren gestartet und der Ablauf koordiniert werden. Dem CP hingegen fiel eine reine beobachtende Rolle zu, wodurch penibler und genauer auf die praktischen Fertigkeiten geachtet werden konnte. Die relative Übereinstimmung zwischen den Prüfer*innen betrug über 90 %. Ähnliche Ergebnisse fanden sich in vergleichbaren Studien [23; 28; 42; 51; 57].

In der Studie von Chenot et al. 2007 wurden, vergleichbar mit dieser Dissertation, reale Prüfungen von zwei Prüfer*innen bewertet [13]. Hierzu bewerteten ärztliche und studentische Prüfer*innen parallel insgesamt 4 Stationen einer OSCE. Bezogen auf die bewerteten Kriterien zeigte sich ebenfalls eine hohe mediane Prüfer*innenübereinstimmung. Die ermittelte rohe Übereinstimmung in der Studie war mit 63-72 % niedriger als in der vorliegenden Dissertation (≥ 90 %)[13].

In einer weiteren Studie von Koch 2008 wurde untersucht, ob studentische und ärztliche Prüfer*innen in ihrer Bewertung übereinstimmten [28]. Hierzu übten Simulationspatient*innen (SP) im Voraus definierte Anamnesen ein, um für die Prüfer*innen vergleichbare Situationen zu schaffen. Die Bewertenden erreichten bei der Beurteilung der Prüfungen gute Übereinstimmungswerte. In dieser Studie zeigte sich, vergleichbar mit unseren Ergebnissen, dass unerfahrene Prüfer*innen kaum Probleme hatten, Prüfungen objektiv zu bewerten. Eine Limitation dieser Studie war, dass die SP die Prüfung nicht jedes Mal identisch vorspielen konnten, sodass sich kleine Unterschiede zwischen den Prüfungen nicht vermeiden ließen. Allerdings konnte von den SP nach der Prüfung eingeschätzt werden, wie vergleichbar die einzelnen Prüfungen waren [28]. In unserer Studie konnte die tatsächlich erbrachte Leistung von den Prüflingen allerdings

nicht analysiert werden. HP und CP bewerteten die identische Prüfung gleichzeitig, allerdings war es nicht bekannt, ob die vergebenen Punkte korrekt von den Prüfer*innen gewertet wurden. Somit waren HP und CP zwar vergleichbar, aber ein Rückschluss auf die Qualität der Bewertung war nicht möglich. Auch wurden die Prüfungen jeweils nur von einem Teil der Prüfer*innen bewertet, wodurch ein allgemeiner Vergleich nur bedingt möglich war. Weder in der Studie von Koch 2008 oder Chenot et al. 2007 noch in dieser konnten alle Prüfer*innen direkt miteinander verglichen werden. Zusammenfassend zeigten sich bei PCs mit HP und CP ausgeprägte Übereinstimmungen der Bewertungen. Allerdings konnte auch hier, wie in Teil 1, nur eine Tendenz angegeben werden, da nicht alle PCs direkt miteinander verglichen werden konnten.

4.3 Urteiler*innenübereinstimmung mehrerer Prüfer*innengruppen durch Online-Praxis-Check-Simulation

4.3.1 Zusammenfassung der Ergebnisse

Um die beschriebenen Limitierungen der vorangegangenen Studienteile zu verringern, wurden PC-Filme erstellt. So war ein detaillierterer Vergleich zwischen den Prüfer*innen möglich. Ein weiterer Vorteil zu den vorangegangenen Analysen, bei denen PCs jeweils nur von einem kleinen Teil der Prüfer*innen gleichzeitig bewertet werden konnten, war, dass alle Prüfer*innen alle PCs bewerteten. So konnten sowohl die 3 Prüfer*innengruppen als auch alle Prüfer*innen direkt miteinander verglichen werden.

In den Prüfungsfilmen war es möglich, je Film maximal 20 Punkte korrekt zu bewerten. Es wurde nicht unterschieden, ob der Punkt vom Schauspieler richtig oder falsch durchgeführt wurde, sondern nur, ob der entweder richtig oder fehlerhaft ausgeführte Punkt korrekt erkannt und gewertet wurde. Als Ergebnis wiesen die Prüfer*innen in allen analysierten Punkten übereinstimmende Ergebnisse auf. Innerhalb der Gruppen erreichten die Tutor*innen (ICC = 0,94) die höchste Übereinstimmung gefolgt von den Studierenden (ICC = 0,92) und den Ärzt*innen (ICC = 0,81). Die allgemeine Übereinstimmung der Prüfer*innen über alle Gruppen hinweg lag bei ICC = 0,97. Nach Koo et al. erreichten die Ärzt*innen eine gute Übereinstimmung [29]. Von Tutor*innen, Studierenden und über alle 3 Gruppen hinweg wurde eine exzellente Übereinstimmung erreicht. Insgesamt war die Übereinstimmung der Prüfer*innen hinsichtlich der korrekt gewerteten Gesamtpunktzahl

(Abbildung 14), der korrekt gewerteten Gesamtpunkte je PC (Abbildung 15), der Skill- und Kontextpunkte (Abbildung 16) und der Zielpunktzahl (Abbildung 17) hoch.

4.3.2 **Bewertung der Ergebnisse**

Die Ergebnisse wiesen in allen Punkten darauf hin, dass die Gruppen in ihrem Prüfverhalten übereinstimmten. Die Gruppen erreichten alle ähnliche Prozentwerte an korrekt gewerteten Gesamtpunkten. Der größte mediale Unterschied fand sich mit 3 % Differenz zwischen Studierenden und Ärzt*innen. Deutlich übereinstimmender bewerteten Tutor*innen und Studierende mit einem Unterschied von 1 %.

Überträgt man die Prozentwerte korrekt gewerteter Punkte auf einen PC, wurden tendenziell über 90 % der Punkte von den Prüfer*innen korrekt bewertet. Dies entsprach 18 von 20 korrekt gewerteten Punkten. Eine Differenz von maximal 3 % zwischen den Gruppen entsprach, bezogen auf einen PC, weniger als einem Punkt Unterschied. Im Schnitt bewerteten die Prüfer*innen unterschiedlicher Gruppen also einen PC Untersuchung Abdomen korrekt und vergleichbar ähnlich.

Betrachtet man die einzelnen PC-Filme, bewerteten Ärzt*innen, Studierende und Tutor*innen wieder vergleichbar gut: Im Median wurden gleich viele Punkte pro PC-Film korrekt gewertet. Nur die ärztliche Gruppe wertete in zwei Filmen einen Punkt weniger korrekt als die anderen beiden. Außerdem wurden bei fehlerhaften PCs weniger Punkte korrekt bewertet als bei fehlerfreien PCs [10]: Prüfer*innen haben sich mit der Bewertung korrekt durchgeführter Fertigkeiten leichter getan.

Die Übereinstimmung der Prüfer*innen wurde ebenfalls in der Auswertung der einzelnen Punkte deutlich. Für jeden Punkt aus jedem Film wurde für jede Gruppe der Prozentsatz korrekt gewerteter Punkte berechnet. In den meisten Punkten erreichten die Gruppen sehr ähnliche Werte. Richtig demonstrierte Punkte wurden meistens von allen korrekt erkannt. Fehlerhafte Punkte hingegen wurden schlechter erkannt, allerdings tendenziell von allen Gruppen ähnlich schlecht. Die ärztliche Gruppe bewertete dabei etwas schlechter als die Vergleichsgruppen. Dies könnte daran liegen, dass sowohl die Tutor*innen als auch die Studierenden eher mit der PC-Situation und dem PB vertraut waren. Auch erhielten alle Prüfer*innen nur eine sehr kurze Einführung zum PC und PB. Dadurch erhielten Tutor*innen tendenziell einen Vorteil, da diesen sowohl der PC als auch PB bekannt war.

Obwohl eine Prüfer*innenschulung für Ärzt*innen und Studierende fehlte, gaben diese qualitativ gute Prüfungsbewertungen ab. In realen Prüfungssituationen kam es immer wieder vor, dass ärztliche Prüfer*innen ohne Schulung als Prüfer*innen bei OSCEs mitwirken sollten. Obwohl von den Ärzt*innen 78 % keine Erfahrung mit dem PC Untersuchung Abdomen angaben, zeigte sich eine zu den anderen Prüfer*innengruppen vergleichbar gute Bewertung der PC-Filme. Dies könnte durch die präzisen Punktekriterien des PB und der klinischen Erfahrung der Ärzt*innen erklärt werden.

Ein weiterer Grund für die Differenzen in den Bewertungen könnte sein, dass Ärzt*innen gewisse Verhaltensweisen in der Klinik als korrekt ansahen, die eventuell nicht mehr dem aktuellen Stand der Lehre entsprachen. Studierende und Tutor*innen waren möglicherweise, genauso wie junge Ärzte*innen, eher mit der aktuellen Lehrmeinung vertraut als ältere Ärzt*innen. Dies könnte in Kombination mit den technischen Schwierigkeiten die minimal schlechtere Bewertungsleistung der älteren Ärzt*innen erklären. Beispielsweise achteten Tutor*innen und Studierende eher darauf, dass keine Suggestivfragen gestellt wurden. Ärzt*innen in der Klinik oder Praxis sind die Patient*innen oft bekannt und vom Untersuchenden gestellte Suggestivfragen wurden möglicherweise nicht als Fehler wahrgenommen (Kapitel 4.4.4 Abbildung 16 Film 4 Punkt 1).

4.3.2.1 Skill- und Kontextpunkte

Bei der Einzelbetrachtung der Skill- und Kontextpunkte zeigte sich wieder eine gute Übereinstimmung. Keine der Gruppen bewertete eine der Punktevarianten deutlich schlechter oder besser als die Vergleichsgruppen. Tendenziell wurden die Punkte von allen gleich gut oder gleich schlecht bewertet. Es bestätigte sich, dass den Prüfer*innen in den PCs klar definiert Kriterien halfen, klare und objektive Bewertungen hinsichtlich der definierten Kriterien zu treffen [19]. Tendenziell schienen die Ärzt*innen Kommunikationspunkte, die fehlerhaft demonstriert wurden, etwas schlechter zu bewerten als die Vergleichsgruppen. Dies könnte wieder an der bereits erwähnten Art der gewohnten Kommunikation in der Klinik bzw. Praxis liegen. Des Weiteren setzten Ärzt*innen möglicherweise ihren Fokus auf Punkte, die sie aufgrund ihrer Erfahrung im Alltag für relevant halten. Fehlerhafte Punkte in den PCs wurden dadurch eventuell eher übersehen. Die gleiche Begründung konnte bei den Hygienepunkten aufgeführt werden. Durch Hektik oder Zeitstress der Ärzt*innen wurde bei der hygienischen Händedesinfektion

oft nicht auf die Zeit geachtet. Deswegen achteten Ärzt*innen möglicherweise nicht in dem Maße wie Studierende oder Tutor*innen auf die erforderlichen 30 s der hygienischen Händedesinfektion. Tutor*innen und Studierende hingegen wurden im Studium und vor allem im Skills Lab darauf trainiert, die 30 s penibel einzuhalten und zu beachten. Dies war eine Möglichkeit, die hohe Übereinstimmung zwischen Studierenden und Tutor*innen, vor allem PC Film 1-4 Punkt 15 (Abbildung 16), im Gegensatz zu den Ärzt*innen zu erklären.

Nicht beurteilbar war die Schwierigkeit der eingebauten Fehler. Die dargestellten Fehler entsprachen Fehlern, die von Studierenden im Skills Lab bei PCs gemacht wurden und stellten somit reale Fehler nach. Ob in einem der Prüfungsfilme die gezeigten Skill- oder Kontextfehler schwerer zu finden waren als in einem anderen, kann nicht geklärt werden. So ist die Aussage, dass Kontextfehler seltener gefunden wurden als Skillfehler, eventuell durch die unterschiedlichen Schwierigkeiten der eingebauten Fehler verzerrt (s. Kapitel 3.3.6). Um die Prüfer*innen hinsichtlich Ihrer Übereinstimmung zu beurteilen war dies allerdings nicht von Relevanz.

4.3.2.2 Gruppe der Ärzt*innen

In der ärztlichen Gruppe fiel auf, dass Ärzt*innen, die älter als 30 Jahre sind, am wenigsten korrekt innerhalb der ärztlichen Gruppen bewerteten. Eine Erklärung für den Unterschied zu jüngeren Ärzt*innen war, dass die älteren dreimal häufiger technische Schwierigkeiten bei der Durchführung der Online-Simulation angaben. Die Überforderung mit der Technik könnte somit Einfluss auf die Bewertungsleistung der Ärzt*innen genommen haben. Zusätzlich mussten sich die Ärzt*innen auf ein ihnen zum Großteil unbekanntes Prüfungssystem einstellen. Die Kombination aus technischer Überforderung, neuem Prüfsystem und einer Prüfsituation könnte der Auslöser dieser Tendenz gewesen sein. Möglicherweise hätten die Ärzt*innen von einem ausführlichen Prüfer*innentraining zu PC, PB und Bewertung am stärksten profitiert [48].

4.3.2.3 Gruppe der Tutor*innen

Bei der Gruppenanalyse der Tutor*innen zeigte sich die Tendenz, dass mit mehr PC-Prüferfahrung die Anzahl korrekt gewerteter Punkte zunahm. Je mehr PCs geprüft wurden, desto mehr Varianten an möglichen Fehlern wurden von den Tutor*innen gesehen. Eventuell fielen den Tutor*innen mit zunehmender Erfahrung mehr Fehler auf. Auch könnte der Umgang mit dem PB erfahreneren Tutor*innen leichter gefallen sein. Die

Konzentration konnte somit eher dem zu bewertenden PC gewidmet werden als dem PB. Möglicherweise hatten unerfahrene Tutor*innen zudem auch weniger klinische Erfahrung und Probleme, gewisse Untersuchungstechniken als korrekt oder inkorrekt zu bewerten.

4.3.2.4 Übereinstimmung der Prüfer*innen – Vergleich mit weiteren Studien

Um die Übereinstimmung zwischen den Prüfer*innen zu ermitteln, wurde der ICC berechnet [35]. Es zeigte sich auf alle Prüfer*innen betrachtet eine exzellente Übereinstimmung der Bewertungen. Bei den einzelnen Gruppen erreichten die Tutor*innen die beste Übereinstimmung, gefolgt von den Studierenden mit einer ebenfalls exzellenten Übereinstimmung. Die Ärzt*innen erreichten eine gute Übereinstimmung. Die Ergebnisse wiesen auf hohe Urteiler*innenübereinstimmung der Prüfer*innen hin. Damit konnte davon ausgegangen werden, dass studentische Tutor*innen als Prüfer*innen für PCs geeignet sind. Im Vergleich zu Ärzt*innen zeigten sich die Bewertung der Tutor*innen gleich gut oder zum Teil sogar besser.

Die hohe Übereinstimmung bei den Tutor*innen konnte durch die hohe Erfahrung im Umgang mit dem PC erklärt werden. Im Skills Lab war der PC Untersuchung Abdomen der fünfthäufigste PC. Die meisten Tutor*innen prüften diesen dementsprechend oft. 70 % der Studierenden wiesen ebenfalls Vorkenntnisse mit dem PC auf. Obwohl die Ärzt*innen nur in 22 % Erfahrung mit dem PC angaben, werteten sie diesen im Vergleich zu den Tutor*innen ähnlich gut. Ein möglicher Grund könnte gewesen sein, dass die Kriterien des PB präzise genug waren, sodass selbst unerfahrene Prüfer*innen keine Probleme mit der Bewertung der Leistung hatten. Ein weiterer Grund könnte gewesen sein, dass die meisten abgefragten Kriterien so praxisnah ausgewählt wurden, dass Ärzt*innen diese standardmäßig beachteten.

Andere Studien zeigten ebenfalls eine hohe Urteiler*innenübereinstimmung. In der Studie von Beck et al. 2016 wurden Ärzt*innen, studentische Tutor*innen und Schüler in ihren Bewertungen einer Laienreanimation verglichen. Die Reanimation wurde jeweils von zwei Prüfer*innen anhand einer Kriterien-Checkliste bewertet. Die Prüfer*innen erreichten in der allgemeine Analyse eine gute Übereinstimmung, im Vergleich zwischen studentischen Tutor*innen und Ärzt*innen eine moderate Übereinstimmung und zwischen Tutor*innen eine gute Übereinstimmung [4].

Im Vergleich mit den Werten der Studie von Beck et al. 2016 wurden in der vorliegenden Studie deutlich bessere Übereinstimmungswerte erzielt. In beiden Studien bewerteten Prüfer*innen praktische Prüfungen mit Hilfe eines Kriterienkatalogs. Allerdings bewerteten bei Beck et al. 2016 nur jeweils 2 Personen die gleiche Prüfung, während in der vorliegenden Studie alle PC-Filme von allen Prüfer*innen bewertet werden konnten. Ein weiterer deutlicher Unterschied der Studien ist, dass bei Beck et al. 2016 die Prüfungen real stattfanden, während in dieser Studie nur vier Varianten eines PCs gezeigt wurden. Wie in der Studie von Beck et al. 2016 konnte bestätigt werden, dass die Bewertungen von Tutor*innen und Ärzt*innen als Prüfer*innen in praktischen Prüfungen übereinstimmten [4].

In der Studie von Vivekananda-Schmidt, Lewis et al. 2007 wurde, vergleichbar zu einem PC, eine sechs Minuten lange praktische Fertigkeit mit einer Checkliste geprüft [61]. Allerdings wurden reale praktische Prüfungen gefilmt und nachträglich von einem*r anderen Prüfer*in die gefilmte OSCE online evaluiert. Ein Vorteil der Studienvariante von Vivekananda-Schmidt, Lewis et al. 2007 war, im Vergleich zu dieser Studie, dass die Urteiler*innenübereinstimmung besser auf die Realität übertragbar war, da eine von beiden Bewertungen zeitgleich zur Prüfung stattfand. Allerdings bewerteten nicht alle Prüfer*innen alle Prüfungen, womit die Vergleichbarkeit der Bewertungen nicht so gut war wie in dieser Studie. Die Bewertenden erreichten moderate Übereinstimmungswerte zwischen realer Bewertung und Videobewertung. Zusammenfassend war die Urteiler*innenübereinstimmung zwischen den Prüfer*innen in der vorliegenden Online-PC-Simulation besser und aussagekräftiger.

4.3.2.5 Übertragbarkeit der Ergebnisse der Online-Simulation auf reale Praxis-Checks

Bei den gefilmten PCs stellte sich die Frage, ob die gewonnen Ergebnisse auf reale PCs übertragbar sind. In der Studie von Sturpe et al. 2010 wurde gezeigt, dass sich die Bewertungen von gefilmten und realen Prüfungen deckten [57]. Dazu wurden live übertragene Prüfungen bewertet und diese nach einem Monat nochmals wiederbewertet. Des Weiteren war es für den Vergleich von Bewertungsleistungen vorteilhaft, dass alle die identische, beispielsweise gefilmte, Situation gezeigt bekamen, da unterschiedliche Fälle die Vergleichbarkeit zwischen den Bewertenden verringern [43]. In der Studie von Kiehl 2018 wurde nachgewiesen, dass eine gefilmte OSCE praktikabel, objektiv und reliabel war

und als Alternative zu einer realen OSCE angesehen werden konnte [27]. Da die Prüfungsform der PCs in vielen Teilen mit der einer OSCE übereinstimmte, konnte davon ausgegangen werden, dass die Ergebnisse der gefilmten PCs ebenso auf die realen PCs übertragen werden konnten. Allerdings wurde in der Online-Simulation nur der PC Untersuchung Abdomen analysiert. Aufgrund der gleichen Struktur und präzisen Kriterien des PB hätten die erlangten Ergebnisse sehr wahrscheinlich auf die anderen PCs des Skills Lab übertragen werden können.

4.3.2.6 Online-Simulation - Vorteile

Der entscheidende Vorteil der PC-Filme war, dass der Vergleich der Bewertungen der Prüfer*innen auf identischen PCs beruhte. In vielen anderen Studien mit praktischen Prüfungen wurden die Prüfungen nur von 2 Prüfer*innen bewertet [13; 21; 22]. Im vorliegenden Studienteil konnten alle PCs von allen Prüfer*innen bewertet werden. Durch die gefilmten PCs sah jeder die identischen Prüfungen und Prüfer*innen konnten valider und reliabler miteinander verglichen werden.

Ein weiterer Vorteil der Online-Simulation der vorliegenden Studie war, dass sie auch unter den Infektionsschutzmaßnahmen gegen Covid-19 durchgeführt werden konnte. Der Ursprungsplan war, teilnehmende Prüfer*innen in einem Raum mit einem, der Realsituation entsprechenden, papierbasierten PB die gefilmten PCs bewerten zu lassen. Dies wurde in eine kontaktfreie, zeit- und ortsunabhängige Online-Studie umgeplant.

Außerdem konnten ärztliche und studentische Teilnehmende einfacher rekrutiert werden. In Zeiten von Kontaktbeschränkungen und Homeoffice hätte sich die Rekrutierung von teilnehmenden Studierenden und Ärzt*innen bei einer ortsabhängigen Studie komplizierter gestaltet. Durch die Online-Simulation war es möglich, Ärzt*innen von verschiedenen Standorten aus an der Studie teilnehmen zu lassen. Ein weiterer Vorteil: Sowohl Studierende als auch Ärzt*innen konnten die Studie zu einem Zeitpunkt ihrer Wahl bearbeiten, was in einer Präsenzstudie nicht so möglich gewesen wäre und zu zusätzlichen organisatorischen Schwierigkeiten geführt hätte.

4.3.2.7 Online-Simulation - Nachteile

Die Nachteile der Online-Simulation im Vergleich zu einer Präsenzstudie ergaben sich überwiegend aus den erhöhten technischen Schwierigkeiten. Ein allgemeines Problem von

Filmen ist, dass die Prüfer*innen einen vorgegebenen Blickwinkel haben und diesen nicht wie in realen Prüfungssituationen anpassen können. Durch den statischen Bildausschnitt waren die Prüfer*innen darauf angewiesen, dass in diesem Ausschnitt alles sichtbar durchgeführt wurde. Die Filme wurden deshalb so erstellt, dass Fehler tatsächlich sichtbar waren und beim aufmerksamen Betrachten auch gefunden werden konnten (Abbildung 3).

Ein kleiner Nachteil der Onlinesimulation war, dass der PB nicht komplett im Original dargestellt werden konnte. Allerdings wurden die Kriterien und der Text identisch übernommen (s. Kapitel 2.3.5). Der Unterschied zwischen der Online-Variante und dem realen PB bestand also nur im Layout, wodurch angenommen werden konnte, dass Prüfer*innen mit beiden Varianten des PB die gleichen Ergebnisse erzielt hätten.

Des Weiteren konnte nicht ausgeschlossen werden, dass die Prüfer*innen während der Bearbeitung der Studie abgelenkt wurden und somit Punkte inkorrekt bewertet wurden. Allerdings sprachen die vergleichbaren Bewertungen der PC-Filme dafür, dass dies eher ein geringes Problem darstellte.

4.3.2.8 Relevanz für den Prüfling

Die Ergebnisse zeigten, dass es für die geprüften Studierenden irrelevant war, von wem sie geprüft wurden. Dazu wurde analysiert, wie stark die von den Prüfer*innen vergebene Punktzahl von der Zielpunktzahl abwich. Die Gruppen vergaben tendenziell die gleiche Punktzahl oder einen Punkt mehr als die Zielpunktzahl. Betrachtete man dieses Ergebnis mit der Information, dass „nur“ zwischen 90 - 93 % der Punkte korrekt gewertet wurden und entweder die gleich Punktzahl oder ein Punkt zu viel vergeben wurde, dann zeigte sich, dass die inkorrekt gewerteten Punkte der Prüfer*innen tendenziell eher zu Gunsten der Prüflinge sind: Fehlerhaft demonstrierte Punkte wurden eher fälschlicherweise mit einem Punkt bewertet als richtige Punkte fehlerhaft gewertet wurden. Des Weiteren wiesen die Prüfer*innen in allen Punkten eine exzellente Übereinstimmung auf, wodurch gezeigt werden konnte, dass die Bewertung des Prüflings nicht von der Prüfer*innengruppe abhängig war.

4.4 Prüfsystem Praxis-Check

Das Prüfungssystem PC der Universität Ulm ist, soweit uns bekannt, ein einzigartiges System in Deutschland. Praktische Fertigkeiten werden sonst oft in OSCE abgefragt [33]. In vielen Punkten sind sich beide Prüfsysteme allerdings sehr ähnlich. Beide verwenden klar definierte Kriterienkataloge, anhand deren ein oder zwei Prüfer*innen die Leistung bewerten [12; 14; 30; 56]. Der Kriterienkatalog hilft, eine vergleichbare Bewertung der Leistung zu erzielen [12]. Oft werden die Fertigkeiten an Schauspielpatient*innen oder Puppen demonstriert.

Im Gegensatz zu den PCs besteht die OSCE meist aus mehreren Stationen, welche jeweils 5 bis 10 Minuten dauern. Die PCs sind daher am ehesten mit einer 10-minütigen OSCE-Teilstation vergleichbar. Ein weiterer Unterschied ist, dass die abgefragten Fertigkeiten beim PC bekannt sind und die einzelnen Kriterien auf der Lernplattform Moodle abgerufen werden können.

Der in den PCs verwendete PB wurde auch in Teil 1, 2 und 3 der vorliegenden Dissertation verwendet. Allerdings ist der PB und der PC nicht validiert. Eine Limitation dieser Studie war also, dass PB und PC als nicht validierte Werkzeuge zum Bewerten von Prüfungen eingesetzt wurden. Dafür, dass PB und PC als valides System angesehen werden konnten, sprachen die Bewertungen der Gruppen. Alle Gruppen bewerteten den PB als sehr gutes Werkzeug, um praktische Fertigkeiten von Prüflingen zu bewerten. Dadurch konnte davon ausgegangen werden, dass der PB für Tutor*innen bzw. Prüfer*innen ein gutes Werkzeug zur Beurteilung von Prüfungen war. Des Weiteren schienen auch unerfahrene Prüfer*innen keine Probleme im Umgang mit dem PB zu haben, da im Median die Bewertung der PCs als leicht empfunden wurde. Dies sprach für einen klaren und übersichtlichen Aufbau des PB und den dort definierten Kriterien. Eine erhöhte Reliabilität wurde durch klar definierte Kriterienkataloge und standardisierte Fragen, wie sie im PB zu finden waren, erreicht [38; 64]. Ähnlich zu den OSCEs richtete sich die Vergabe der Punkte nach einheitlich formulierten Kriterien, an denen sich die Prüfer*innen orientierten [23]. In beiden Prüfsystemen absolvierte der Prüfling eine praktische Fertigkeit in einer limitierten Zeit, welche mit einem standardisierten Kriterienkatalog von den Prüfer*innen bewertet wurde. Aufgrund der Ähnlichkeit zur validierten OSCE konnte angenommen werden, dass die Prüfvariante PC ebenso valide war wie die OSCE.

4.5 Evaluation der Fragestellungen

Zusammenfassend zeigen die Ergebnisse, dass geschulte studentische Prüfende in der Bewertung praktischer Fertigkeiten in PCs ähnlich treffsicher waren wie ärztliche Prüfende.

Die Ergebnisse der Online-PC-Simulation zeigen, dass die Prüfer*innengruppen in ihren Bewertungen tendenziell übereinstimmten. Die generell kleinen Unterschiede zwischen den Gruppen wurden im Kapitel 4.3 im Detail beschrieben, hatten allerdings keinen relevanten Einfluss auf die Bewertungen der PCs durch unterschiedliche Prüfer*innengruppen.

Studentische Tutor*innen können also als eigenständige Prüfer*innen für curriculare PC-Prüfungen eingesetzt werden. Um diese als eigenständige Prüfer*innen zuzulassen, werden folgende Kriterien empfohlen: Tutor*innen sollten mindestens 10 reale PCs als CP bewertet haben und/oder einen Wert von > 90 % korrekt gewerteten Punkten in der Online-PC-Simulation erreicht haben (s. Kapitel 3.3.4 und 2.4).

4.6 Weitere Fragestellungen

Um die Validität und Vergleichbarkeit der PCs mit der OSCE zu bestätigen, könnten beide Prüfsysteme in weiteren Studien miteinander verglichen werden. Vermutlich ist der PC ebenso valide wie die OSCE und könnte ebenso als Prüfungsalternative eingesetzt werden. Die Ergebnisse zeigen, dass die Tutor*innen vergleichbar bewerteten und als Prüfer*innen eingesetzt werden konnten. Da nur der PC Untersuchung Abdomen in der online Simulation analysiert wurde, wäre eine Folgestudie mit weiteren PCs denkbar.

Mit der Online-PC-Simulation besteht nun die Möglichkeit, unerfahrene Tutor*innen in ihrer Prüfungsfähigkeit besser einzuschätzen und so individueller entscheiden zu können, ob ein*e Tutor*in bereit ist, eigenständig Prüfungen zu bewerten. Ob die Online-Simulation dafür als valides Werkzeug eingesetzt werden kann, kann noch genauer untersucht und evaluiert werden.

5 Zusammenfassung

Hintergrund:

Im Skills Lab der Universität Ulm wurde in den letzten Jahren der Praxis-Check (PC) als neue Prüfungsform für praktische Fertigkeiten eingeführt. Studentische Tutor*innen wurden hierfür als Prüfer*innen eingesetzt. Zur Überprüfung der Qualität wurden die Bewertungen der Tutor*innen untereinander und mit Ärzt*innen verglichen. Diese Untersuchung von Urteiler*innenübereinstimmungen war sinnvoll, da aufgrund der neuen Approbationsordnung für Ärzt*innen die Anzahl curricularer praktischer Prüfungen anstieg und studentische Tutor*innen bei vergleichbaren Ergebnissen mit Ärzt*innen als valide studentische Prüfer*innen eingesetzt werden könnten.

Methoden:

Um die Urteiler*innenübereinstimmung zu ermitteln, wurde die Studie in 3 Abschnitte untergliedert: Als erstes wurden die Daten der Prüfer*innen von PCs mit einem*r Tutor*in als Prüfer*in analysiert. Als zweites wurden die Bewertungen von Haupt- und Co-Prüfer*innen bei PCs analysiert. Als drittes wurde eine Online-PC-Simulation entwickelt und durchgeführt. Hier bewerteten alle Prüfer*innen die gleichen vier PC-Filme. Anhand der Daten wurden die Urteiler*innenübereinstimmung der Prüfer*innen evaluiert.

Ergebnisse:

Es zeigte sich, dass studentische Tutor*innen und Ärzt*innen in ihren Bewertungen praktischer Fertigkeiten ähnlich treffsicher übereinstimmten. In allen Teilen der Studie wiesen die Tutor*innen ein homogenes Prüfverhalten auf. In realen PCs mit Haupt- und Co-Prüfer*innen wurden relative Übereinstimmungswerte von ≥ 90 % erreicht. Die mit der Intraklassen-Korrelation berechnete Urteiler*innenübereinstimmung in der Online-PC-Simulation zeigte exzellente Übereinstimmungswerte der Bewertungen aller Prüfer*innen.

Schlussfolgerung:

Studentische Tutor*innen bewerteten PCs vergleichbar gut, tendenziell sogar korrekter als Ärzt*innen. Studentische Tutor*innen können aufgrund vergleichbarer Ergebnisse mit Ärzt*innen als valide studentische Prüfer*innen bei PCs eingesetzt werden. Aus den Untersuchungsergebnissen ergibt sich die Empfehlung, Tutor*innen mit mehr als 10 realen PCs als Co-Prüfer*in und/oder einem Wert von > 90 % korrekt gewerteten Punkten in der Online-Simulation zum eigenständigen Prüfen zuzulassen.

6 Literaturverzeichnis

1. Allen A R, Boraks N: Peer tutoring: putting it to the test. The Reading Teacher, 3: 274–278 (1978)

2. Arnold L: Assessing Professional Behavior: Yesterday, Today, and Tomorrow. Academic Medicine, 6: 502–515 (2002)

3. Bahlmann O: Illustrated versus non-illustrated anatomical test items in anatomy course tests and German Medical Licensing examinations (M1). GMS journal for medical education, 2: 1–11 (2018)

4. Beck S, Ruhnke B, Issleib M, Daubmann A, Harendza S, Zöllner C: Analyses of inter-rater reliability between professionals, medical students and trained school children as assessors of basic life support skills. BMC medical education, 1: 3–8 (2016)

5. Bene K L, Bergus G: When learners become teachers. Family medicine: 783–787 (2014)

6. Blackmagic Design: https://www.blackmagicdesign.com/de/products/davinciresolve/ (zuletzt aufgerufen am 03.05.2021)

7. Blohm M, Lauter J, Branchereau S, Krautter M, Köhl-Hackert N, Jünger J, Herzog W, Nikendei C: "Peer-assisted learning" (PAL) in the Skills-Lab--an inventory at the medical faculties of the Federal Republic of Germany. GMS Zeitschrift für medizinische Ausbildung, 1: 1–18 (2015)

8. Brueckner J K, MacPherson B R: Benefits from peer teaching in the dental gross anatomy laboratory. European journal of dental education : official journal of the Association for Dental Education in Europe, 2: 72–77 (2004)

9. Bundesgesundheitsministerium: https://www.bundesgesundheitsministerium.de/themen/gesundheitswesen/gesundheitsberufe/aerzte/approbationsordnung.html (zuletzt aufgerufen am 03.05.2021)

10. Burchard K W, Rowland-Morin P A, Coe N P, Garb J L: A surgery oral examination: interrater agreement and the influence of rater characteristics. Academic Medicine, 11: 1044–1046 (1995)

11. Campos-Outcalt D, Rutala P J, Witzke D B, Fulginiti J V: Performances of underrepresented-minority students at the University of Arizona College of Medicine, 1987-1991. Academic Medicine, 7: 577–582 (1994)

12. Chenot J-F, Ehrhardt M: Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. ZFA-Zeitschrift für Allgemeinmedizin, 09: 437–442 (2003)

13. Chenot J-F, Simmenroth-Nayda A, Koch A, Fischer T, Scherer M, Emmert B, Stanske B, Kochen M M, Himmel W: Can student tutors act as examiners in an objective structured clinical examination? Medical education, 11: 1032–1038 (2007)

14. Dent J A, Hesketh E A: Developing the teaching instinct, 13: How to teach in the clinical skills centre. Medical teacher, 3: 207–210 (2004)

15. Elshami W: Peer assessment in health professions education. JETHS, 1: 8–14 (2018)

16. Elton L RB, Laurillard D M: Trends in research on student learning. Studies in higher education, 1: 87–102 (1979)

17. Evasys: https://evasys.de/evasys/ (zuletzt aufgerufen am 06.05.2021)

18. Falchikov N, Blythman M: Learning together. Routledge Falmer, London (2001)

19. Finn G M, Garner J: Twelve tips for implementing a successful peer assessment. Medical teacher, 6: 443–446 (2011)

20. Goldschmid B, Goldschmid M L: Peer teaching in higher education: A review. High Educ, 1: 9–33 (1976)

21. Gupta P, Dewan P, Singh T: Objective structured clinical examination (OSCE) revisited. Indian pediatrics, 11: 911–920 (2010)

22. Harden R M, GLEESON F A: Assessment of clinical competence using an objective structured clinical examination (OSCE). Medical education, 1: 39–54 (1979)

23. Harden R M, Stevenson M, Downie W W, Wilson G M: Assessment of clinical competence using objective structured examination. British medical journal, 5955: 447–451 (1975)

24. IMPP: https://www.impp.de/informationen/kompetenzorientierte-gegenstandskataloge/masterplan-medizinstudium-2020-kopie.html (zuletzt aufgerufen am 05.05.2021)

25. Jünger J, Just I: Recommendations of the German Society for Medical Education and the German Association of Medical Faculties regarding university-specific assessments during the study of human, dental and veterinary medicine. GMS Zeitschrift für medizinische Ausbildung, 3: 1–23 (2014)

26. Kassab S, Abu-Hijleh M F, Al-Shboul Q, Hamdy H: Student-led tutorials in problem-based learning: educational outcomes and students' perceptions. Medical teacher, 6: 521–526 (2005)

27. Kiehl C: Standardisierte und qualitätsgesicherte Video-Prüfung: Aufklärungsgespräch vor der Operation. Niedersächsische Staats-und Universitätsbibliothek Göttingen, Göttingen (2018)

28. Koch A: Studentische Tutoren als Prüfer in einer''objective structured clinical examination''(OSCE): Evaluation ihrer Bewertungsleistungen. Niedersächsische Staats-und Universitätsbibliothek Göttingen, Göttingen (2008)

29. Koo T K, Li M Y: A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine, 2: 155–163 (2016)

30. Kruppa E, Jünger J, Nikendei C: Einsatz innovativer Lern-und Prüfungsmethoden an den Medizinischen Fakultäten der Bundesrepublik Deutschland–Eine aktuelle Bestandsaufnahme. Deutsche medizinische Wochenschrift (1946), 08: 371–372 (2009)

31. Kürten P, Messerer D A C, Denzinger M, Friz P, Böckers A, Grab C, Horneffer A: In 7 Stufen vom Tutoriat zum Peer Assessment: Das Ulmer Ausbildungscurriculum für Skills Lab-Tutor(inn)en. German Medical Science GMS Publishing House (2019)

32. Lai N M, Sivalingam N, Ramesh J C: Medical students in their final six months of training: progress in self-perceived clinical competence, and relationship between experience and confidence in practical skills. Singapore medical journal, 11: 1018–1027 (2007)

33. Laidlaw A, Salisbury H, Doherty E M, Wiskin C: National survey of clinical communication assessment in medical education in the United Kingdom (UK). BMC medical education, 1: 1–7 (2014)

34. Mavis B E, Cole B L, Hoppe R B: A survey of student assessment in U.S. medical schools: the balance of breadth versus fidelity. Teaching and learning in medicine, 2: 74–79 (2001)

35. Mehta S, Bastero-Caballero R F, Sun Y, Zhu R, Murphy D K, Hardas B,Koch G: Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. Statistics in medicine, 18: 2734–2752 (2018)

36. Miller G E: The assessment of clinical skills/competence/performance. Academic Medicine, 9 Suppl: 63–67 (1990)

37. Müller S, Settmacher U, Koch I, Dahmen U: A pilot survey of student perceptions on the benefit of the OSCE and MCQ modalities. GMS journal for medical education, 4: 1–13 (2018)

38. Newble D I, Hoare J, Sheldrake P F: The selection and training of examiners for clinical examinations. Medical education, 5: 345–349 (1980)

39. Nikendei C, Schilling T, Nawroth P, Hensel M, Ho A D, Schwenger V, Zeier M, Herzog W, Schellberg D, Katus H A, Dengler T, Stremmel W, Müller M, Jünger J: Integriertes Skills-Lab-Konzept für die studentische Ausbildung in der Inneren Medizin. Deutsche medizinische Wochenschrift (1946), 18: 1133–1138 (2005)

40. NKLM: https://nklm.de/zend/menu/index (zuletzt aufgerufen am 11.05.2021)

41. Norcini J J: Peer assessment of competence. Medical education, 6: 539–543 (2003)

42. O'Connor H M, McGraw R C: Clinical skills training: developing objective assessment instruments. Medical education, 5: 359–363 (1997)

43. P. Collins R M H J: AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. Medical teacher, 6: 508–521 (1998)

44. Pabst R: Medical education and reform initiatives in Germany. Academic Medicine, 11: 1006–1011 (1995)

45. Patrício M F, Julião M, Fareleira F, Carneiro A V: Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Medical teacher, 6: 503–514 (2013)

46. Prism - GraphPad: https://www.graphpad.com/scientific-software/prism/ (zuletzt aufgerufen am 03.05.2021)

47. Reschke K, Werwick K, Mersson L, Clasen K, Urbach D, Haß H J, Meyer F: Skills Lab aus chirurgischer Sicht. Erfahrungen des Universitätsklinikums Magdeburg. Der Chirurg; Zeitschrift für alle Gebiete der operativen Medizin, 10: 841–850 (2013)

48. Scheffer S: Validierung des „Berliner Global Rating“ (BGR), Berlin (2009)

49. Schleicher I, Leitner K, Juenger J, Moeltner A, Ruesseler M, Bender B, Sterz J, Stibane T, Koenig S, Frankenhauser S, Kreuder J G: Does quantity ensure quality? Standardized OSCE-stations for outcome-oriented evaluation of practical skills at different medical faculties. Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft: 55–60 (2017)

50. Schultz J-H, Nikendei C, Weyrich P, Möltner A, Fischer M R, Jünger J: Qualitätssicherung von Prüfungen am Beispiel des OSCE-Prüfungsformats: Erfahrungen der Medizinischen Fakultät der Universität Heidelberg. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 10: 668–672 (2008)

51. Schwerdtfeger K: Entwicklung, Reliabilität und Objektivität einer" Objective Structured Clinical Examination" in der Notfallmedizin. Niedersächsische Staats-und Universitätsbibliothek Göttingen, Göttingen (2010)

52. Schwill S, Fahrbach-Veeser J, Moeltner A, Eicher C, Kurczyk S, Pfisterer D, Szecsenyi J, Loukanova S: Peers as OSCE assessors for junior medical students - a review of routine use: a mixed methods study. BMC medical education, 1: 1–12 (2020)

53. Secomb J: A systematic review of peer teaching and learning in clinical education. Journal of clinical nursing, 6: 703–716 (2008)

54. Segarra L M, Schwedler A, Weih M, Hahn E G, Schmidt A: Der Einsatz von medizinischen Trainingszentren für die Ausbildung zum Arzt in Deutschland, Österreich und der deutschsprachigen Schweiz. GMS Zeitschrift für medizinische Ausbildung, 2: 1–7 (2008)

55. SPSS Statistics: https://www.ibm.com/products/spss-statistics?lnk=STW_US_STESCHP1_BLK&lnk2=trial_SPSSstat&lot=2&pexp=def&psrc=none (zuletzt aufgerufen am 26.05.2021)

56. Stillman P L: A Diagnostic Fourth-Year Performance Assessment. Arch Intern Med, 11: 1981–1985 (1987)

57. Sturpe D A, Huynh D, Haines S T: Scoring Objective Structured Clinical Examinations Using Video Monitors or Video Recordings. American journal of pharmaceutical education, 3: 1–5 (2010)

58. Topping K J: The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. High Educ, 3: 321–345 (1996)

59. Topping K J: Trends in Peer Learning. Educational Psychology, 6: 631–645 (2005)

60. Universität Ulm: https://www.uni-ulm.de/misc/uulm-pro-mint-med/themen-inhalte/massnahmen/methodisch-didaktische-weiterbildung-fuer-tutoren-und-mentoren/studientrack-train-the-tutor-in-der-medizin/ (zuletzt aufgerufen am 08.03.2021)

61. Vivekananda-Schmidt P, Lewis M, Coady D, Morley C, Kay L, Walker D, Hassell A B: Exploring the use of videotaped objective structured clinical examination in the assessment of joint examination skills of medical students. Arthritis and rheumatism, 5: 869–876 (2007)

62. Walker-Bartnick L A, Berger J H, Kappelman M M: A model for peer tutoring in the medical school setting. Journal of medical education, 4: 309–315 (1984)

63. Weitz G, Vinzentius C, Twesten C, Lehnert H, Bonnemeier H, König I R: Effects of a rater training on rating accuracy in a physical examination skills assessment. GMS Zeitschrift für medizinische Ausbildung, 4: 1–17 (2014)

64. Wilkinson T J, Frampton C M, Thompson-Fawcett M, Egan T: Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. Academic Medicine, 2: 219–223 (2003)

65. Yeager V L: Peer teaching in gross anatomy. Academic Medicine, 11: 922 (1981)

66. Youdas J W, Hoffarth B L, Kohlwey S R, Kramer C M, Petro J L: Peer teaching among physical therapy students during human gross anatomy: perceptions of peer teachers and students. Anatomical sciences education, 5: 199–206 (2008)

67. Zimmermann P, Kadmon M: Standardized examinees: development of a new tool to evaluate factors influencing OSCE scores and to train examiners. GMS journal for medical education, 4: 1–24 (2020)

7 Anhang

7.1 Abbildungen

7.1.1 Durchschnittspunktzahl je Tutor*in je PC

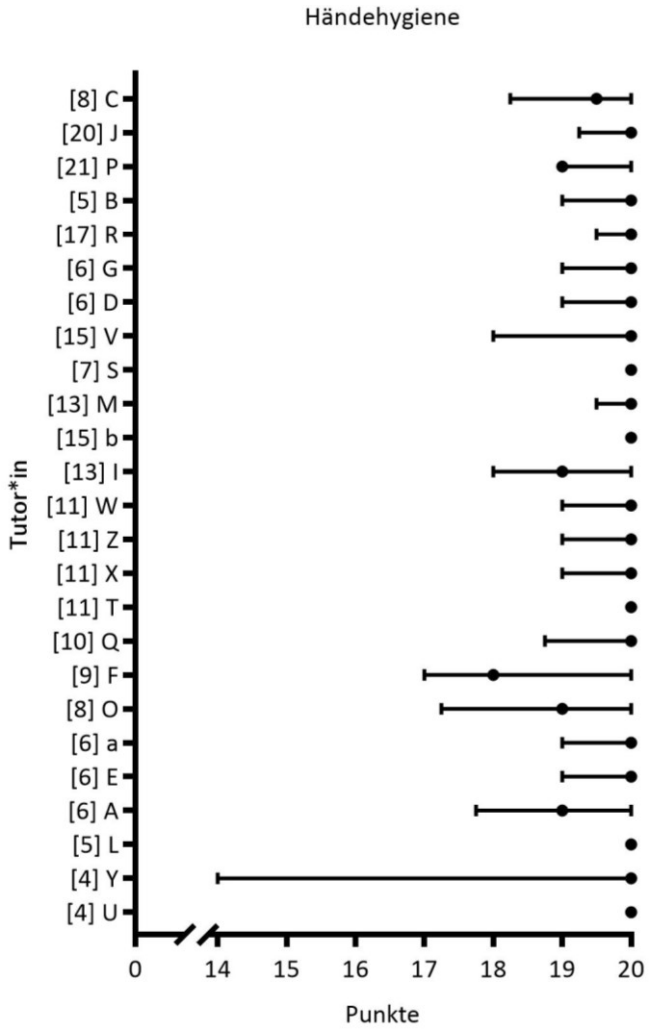


Abbildung 18: vergebene Durchschnittspunktzahl je Tutor*in (mindestens 10 Praxis-Checks geprüft) im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; Allgemein = allgemein vergebener Punktedurchschnitt; [n] = Anzahl Praxis-Checks der Tutor*in, A-b = codierte Prüfer*innen

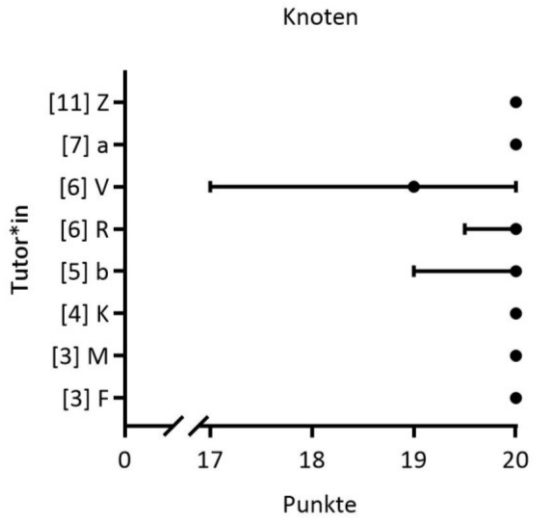


Abbildung 19: vergebene Durchschnittspunktzahl je Tutor*in (mindestens 10 Praxis-Checks geprüft) im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; allgemein = allgemein vergebener Punktedurchschnitt; [n] = Anzahl Praxis-Checks der Tutor*in, A-b = codierte Prüfer*innen

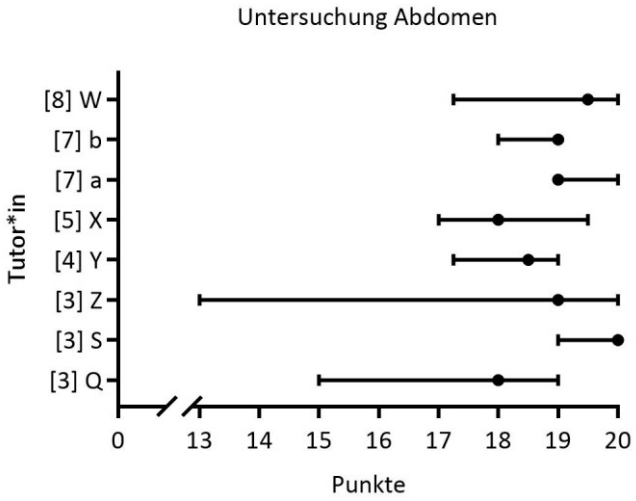


Abbildung 20: vergebene Durchschnittspunktzahl je Tutor*in (mindestens 10 Praxis-Checks geprüft) im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; Allgemein = allgemein vergebener Punktedurchschnitt; [n] = Anzahl Praxis-Checks der Tutor*in, A-b = codierte Prüfer*innen

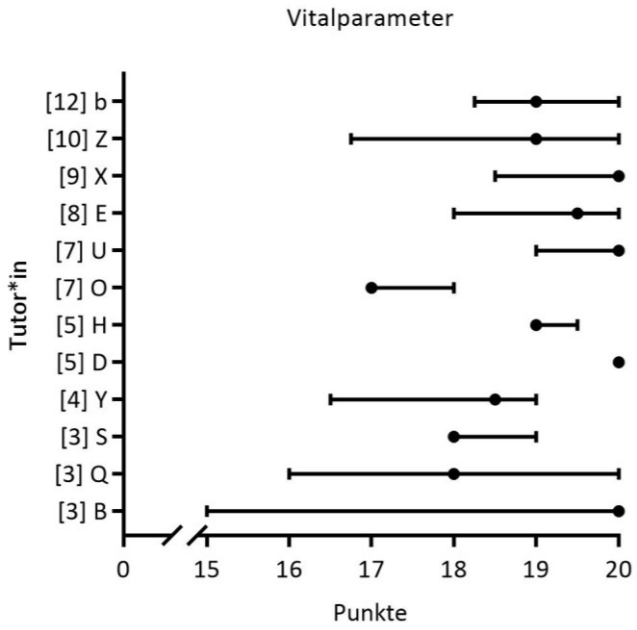


Abbildung 21: vergebene Durchschnittspunktzahl je Tutor*in (mindestens 10 Praxis-Checks geprüft) im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; Allgemein = allgemein vergebener Punktedurchschnitt; [n] = Anzahl Praxis-Checks der Tutor*in, A-b = codierte Prüfer*innen

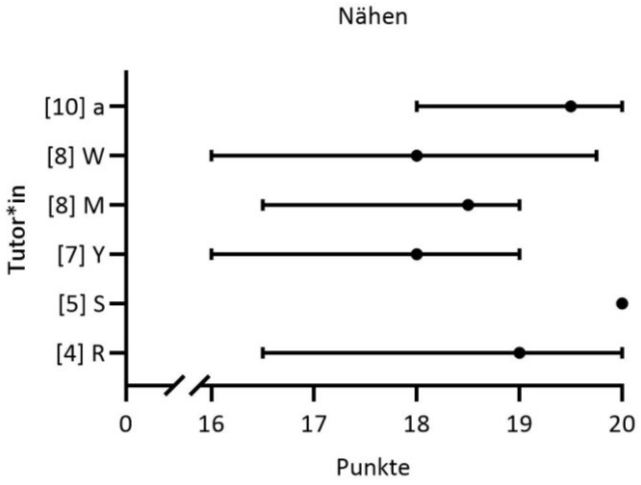


Abbildung 22: vergebene Durchschnittspunktzahl je Tutor*in (mindestens 10 Praxis-Checks geprüft im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; Allgemein = allgemein vergebener Punktedurchschnitt; [n] = Anzahl Praxis-Checks der Tutor*in, A-b = codierte Prüfer*innen

7.1.2 Korrelation Anzahl PC und Durchschnittspunktzahl

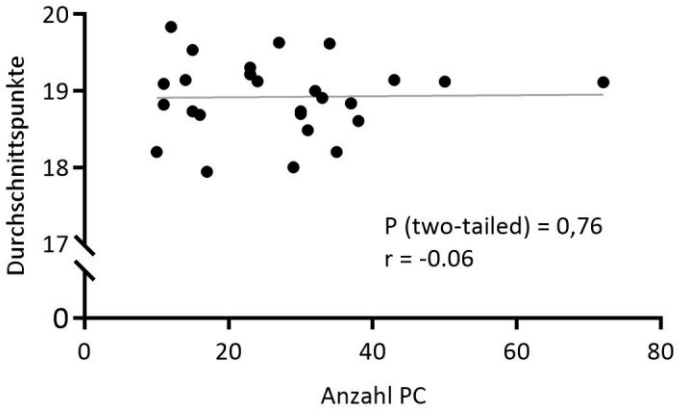


Abbildung 23: Korrelation Anzahl Praxis-Checks als Hauptprüfer*in und Durchschnittspunktzahl als Hauptprüfer*in im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Keine Korrelation feststellbar.

7.1.3 Skill und Kontextfehler

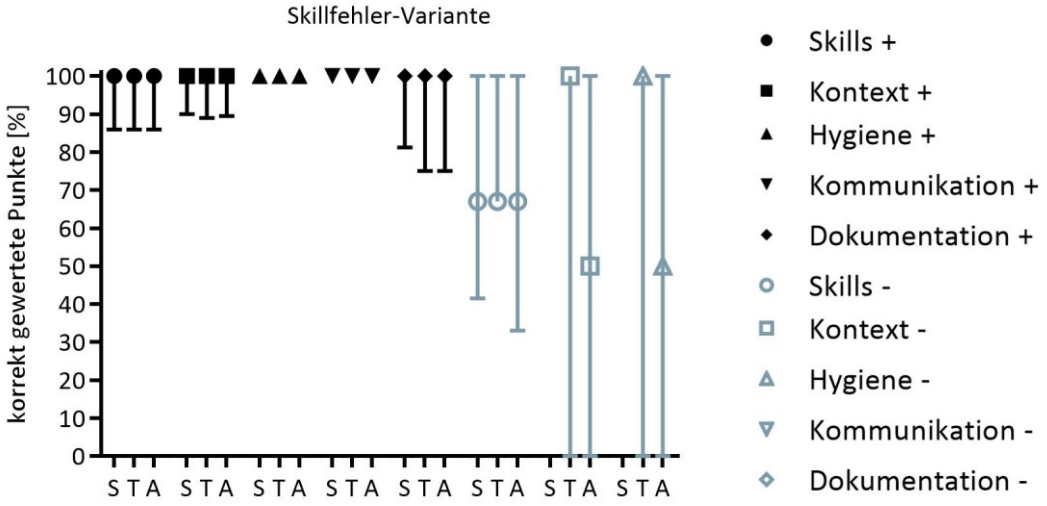


Abbildung 24: % korrekt gewerteter Punkte je Prüfer*innengruppe je Skill und Kontextpunkte in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Median ± Interquartilsabstand, Werte wurden auf 1 normiert, S = Studierende, A = Ärzt*innen, T = Tutor*innen, + ≙ richtige Punkte korrekt gewertet, – ≙ falsche Punkte korrekt gewertet

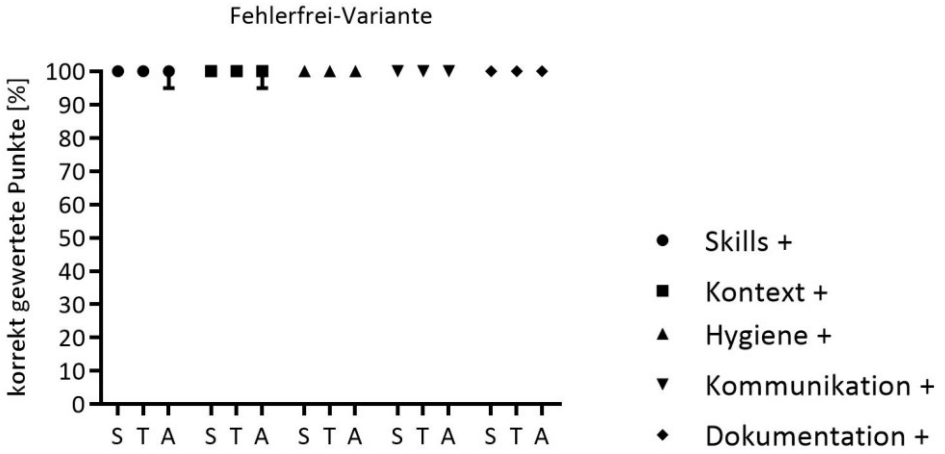


Abbildung 26: % korrekt gewerteter Punkte je Prüfer*innengruppe je Skill und Kontextpunkte in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Median ± Interquartilsabstand, Werte wurden auf 1 normiert, S = Studierende, A = Ärzt*innen, T = Tutor*innen, + ≙ richtige Punkte korrekt gewertet, − ≙ falsche Punkte korrekt gewertet

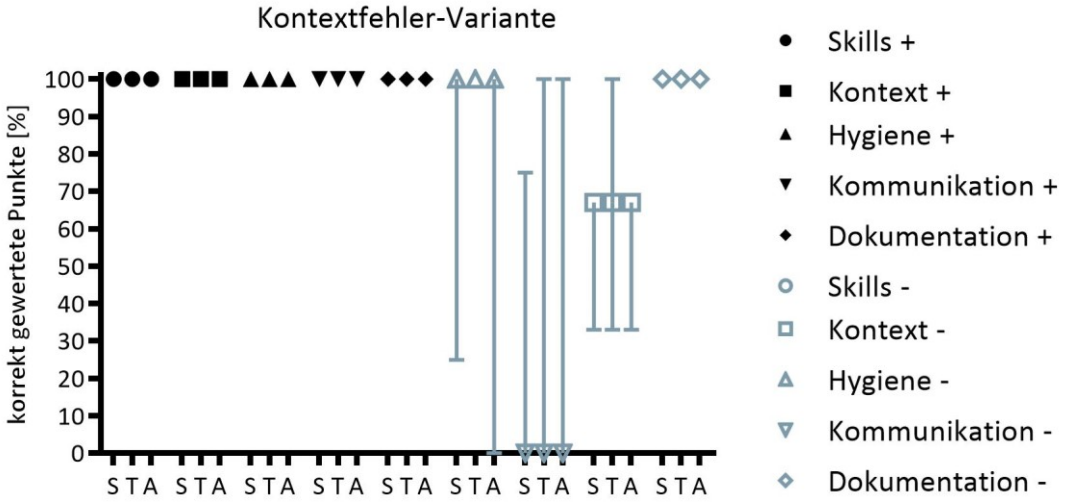


Abbildung 25: % korrekt gewerteter Punkte je Prüfer*innengruppe je Skill und Kontextpunkte in der Online-Praxis-Check-Simulation. Median ± Interquartilsabstand, Werte wurden auf 1 normiert, S = Studierende, A = Ärzt*innen, T = Tutor*innen, + ≙ richtige Punkte korrekt gewertet, − ≙ falsche Punkte korrekt gewertet

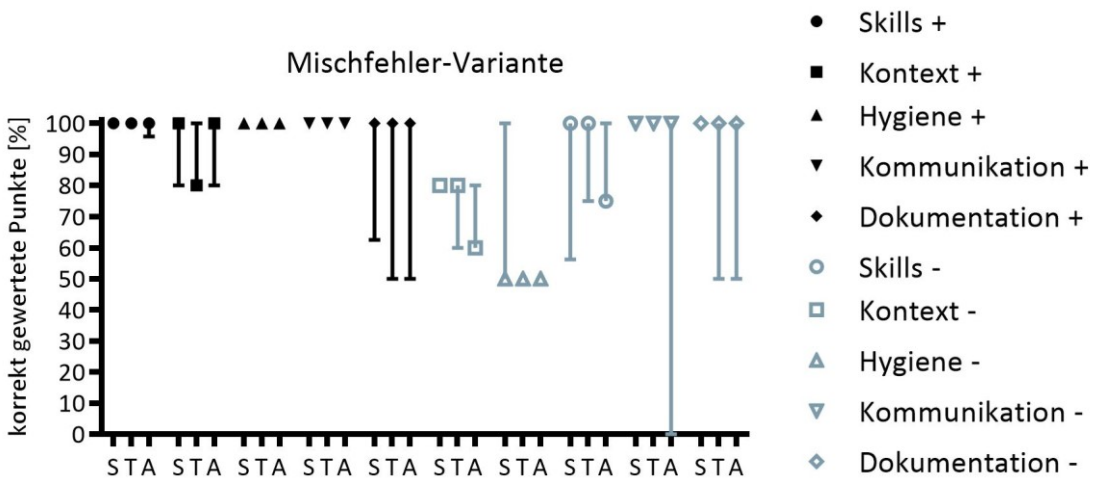


Abbildung 27: % korrekt gewerteter Punkte je Prüfer*innengruppe je Skill und Kontextpunkte in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Median ± Interquartilsabstand, Werte wurden auf 1 normiert, S = Studierende, A = Ärzt*innen, T = Tutor*innen, + ≙ richtige Punkte korrekt gewertet, – ≙ falsche Punkte korrekt gewertet

7.1.4 **Mediane Übereinstimmung zwischen HP und CP**

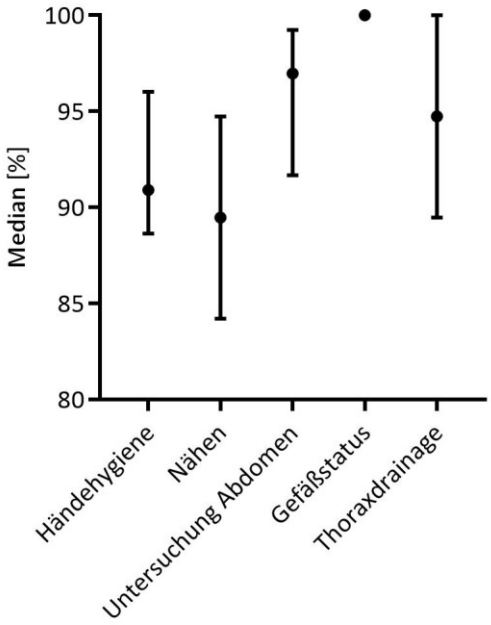


Abbildung 28: Vergleich der Mediane für die Prozentuale Übereinstimmung der Hauptprüfer*innen und Co-Prüfer*innen für die 5 häufigsten Praxis-Checks im Skills Lab der Universität Ulm im Sommersemester 19 und Wintersemester 19/20. Median ± Interquartilsabstand

7.1.5 **Ursache der Punktedifferenz**

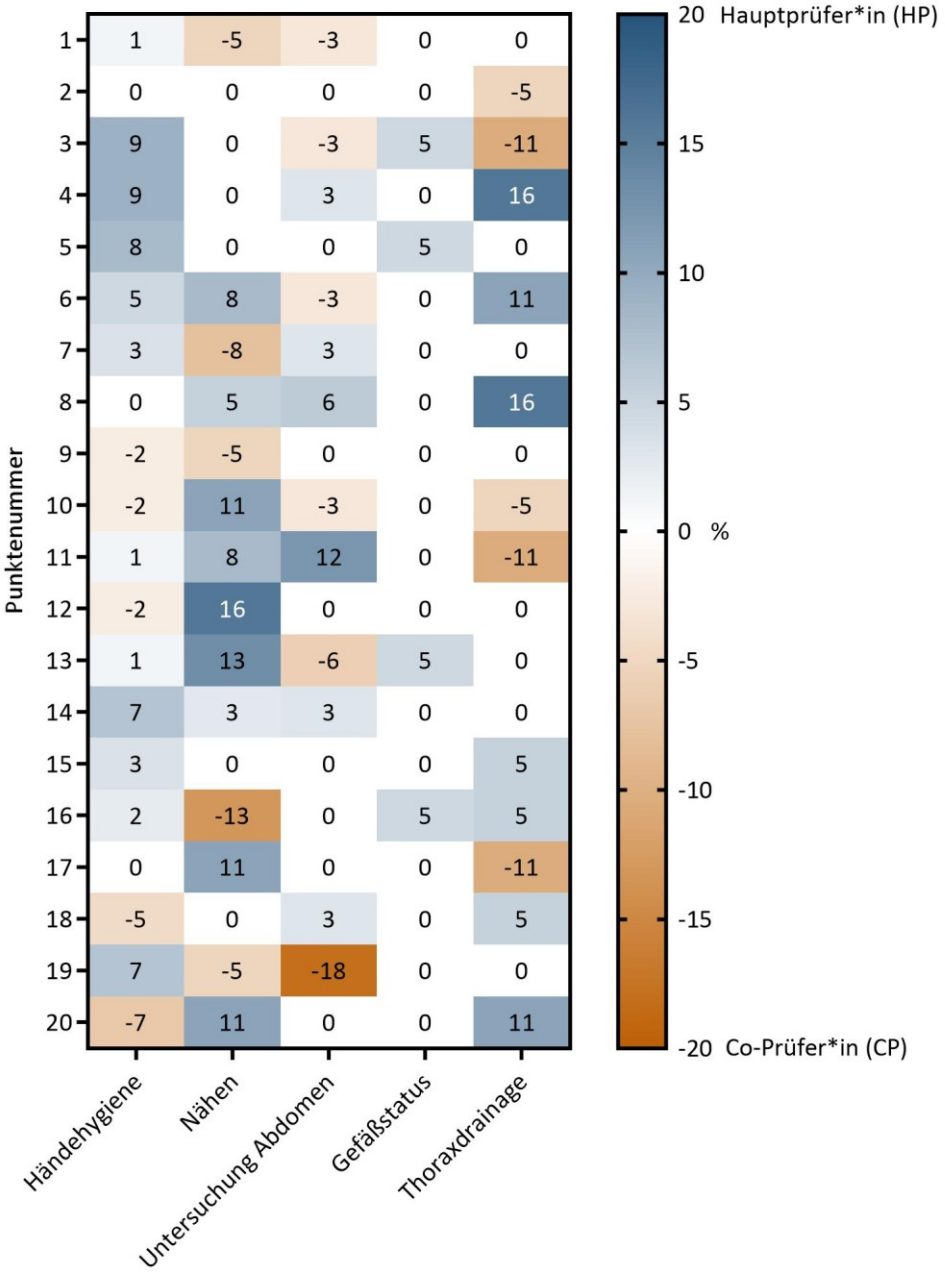


Abbildung 29: Bei einem Großteil der Punkte findet sich eine mindestens 90 %ige Übereinstimmung zwischen der Punktegabe von HP und CP. Ungleichheit ausgelöst durch HP (Tutor*in) oder CP (Tutor*in oder Ärzt*in) in % im Skills Lab der Universität Ulm im Sommersemester 19 und Wintersemester 19/20. Negative Zahlen: HP hat kein Punkt gegeben aber CP schon; Positive Zahlen: HP hat Punkt gegeben aber CP nicht. Hygiene n = 88; Nähen n = 38; Untersuchung Abdomen n = 33; Gefäßstatus n = 22; Thoraxdrainage n = 19

7.1.6 **Allgemeine Daten der Teilnehmende der Online-PC-Simulation**

Tabelle 4: Eingangsfragen der Teilnehmenden der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm; Zeitraum: Sommersemester 20; Median ± Interquartilsabstand

	Studierende	Skillslabtutor*innen	Ärzt*innen	Gesamt
Männlich	10 (42 %)	9 (39 %)	8 (44 %)	27 (42 %)
Weiblich	14 (58 %)	14 (61 %)	10 (56 %)	38 (58 %)
Skala	1=keine Erfahrung; 2=1 Jahr; 3=2 Jahre; 4=3 Jahre; 5=3+ Jahre			
Berufserfahrung	1(1;1)	1(1;1)		
Skala	1=1 Jahr; 2=2 Jahre; 3=3-5 Jahre; 4= 6-10 Jahre; 5=10+ Jahre			
Berufserfahrung			3(1,75;5)	
Fachsemester	10(9;10)	8 (6;9)		9(8;10)
Erfahrung als Tutor*in	14 (58 %)			
Keine Erfahrung als Tutor*in	10 (42 %)			
Erfahrung als Prüfer*in	12 (50 %)	22 (96 %)	16 (89 %)	50 (77 %)
Keine Erfahrung als Prüfer*in	12 (50 %)	1 (4 %)	2 (11 %)	15 (23 %)
Erfahrung mit PC Abdomen Untersuchung	17 (71 %)	21 (91 %)	4 (22 %)	43 (65 %)
Keine Erfahrung mit PC Abdomen Untersuchung	7 (29 %)	2 (9 %)	14 (78 %)	23 (35 %)

7.1.7 Erfahrung der Studierenden als Tutor*in der Online-PC-Simulation

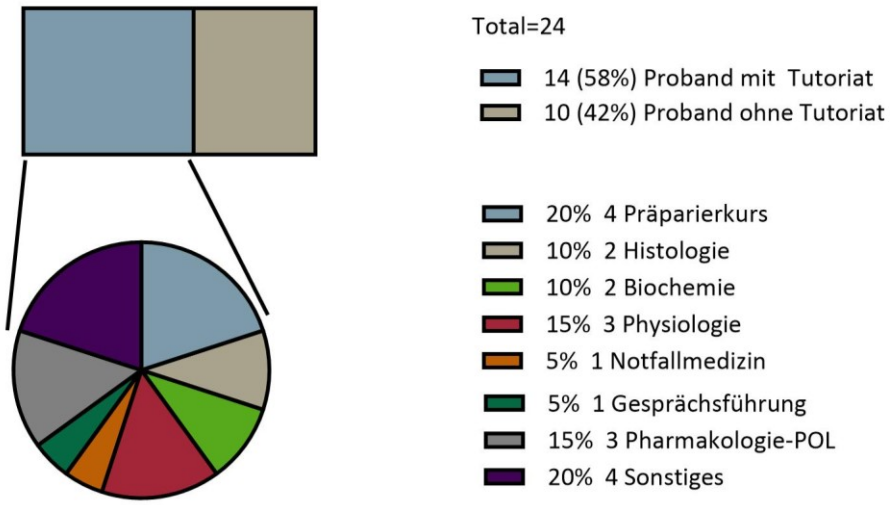


Abbildung 30: Erfahrung der Studierenden (nicht Skills Lab Tutor*innen) als Tutor*in nach Fachbereichen in % der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20

7.1.8 Fachbereiche der Ärzt*innen der Online-PC-Simulation

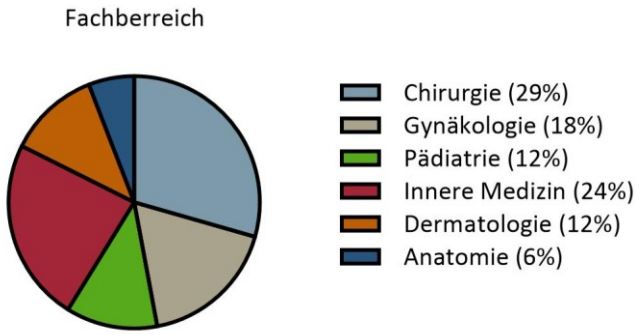


Abbildung 31: Fachbereiche der Ärzt*innen in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; n = 15

7.1.9 **Selbsteinschätzung der Teilnehmenden der Online-PC-Simulation**

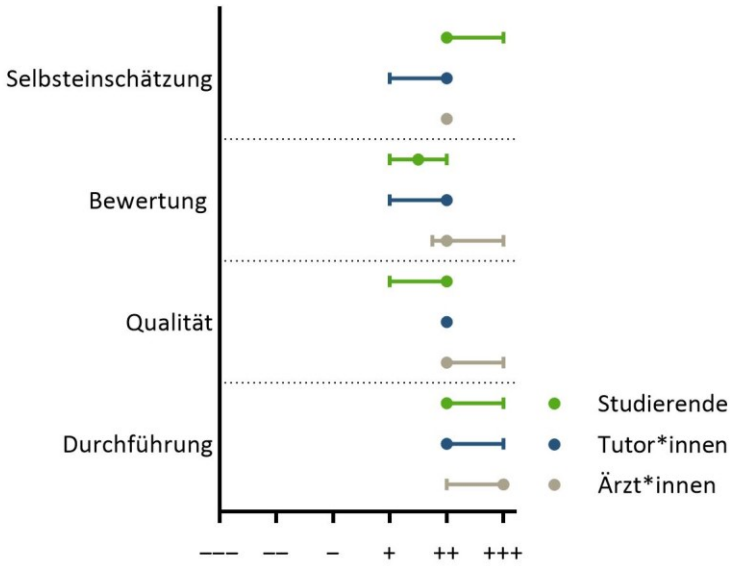


Abbildung 32: Selbsteinschätzung der Teilnehmenden der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; 6-stufige Likert-ähnliche Skala: −−− = trifft gar nicht zu, +++ = trifft völlig zu. Selbsteinschätzung: Ich halte mich für eine*n objektive*n Prüfer*in.; Bewertung: Ich fühle mich grundsätzlich sicher darin, studentische Leistungen in praktischen Fertigkeiten einzuschätzen.; Qualität: Ich fühle mich dazu befähigt, die Qualität einer durchgeführten Abdomen Untersuchung zu beurteilen.; Durchführung: Ich fühle mich grundsätzlich dazu befähigt, eine strukturierte Abdomen Untersuchung durchzuführen.

7.1.10 **Abschlussbefragung**

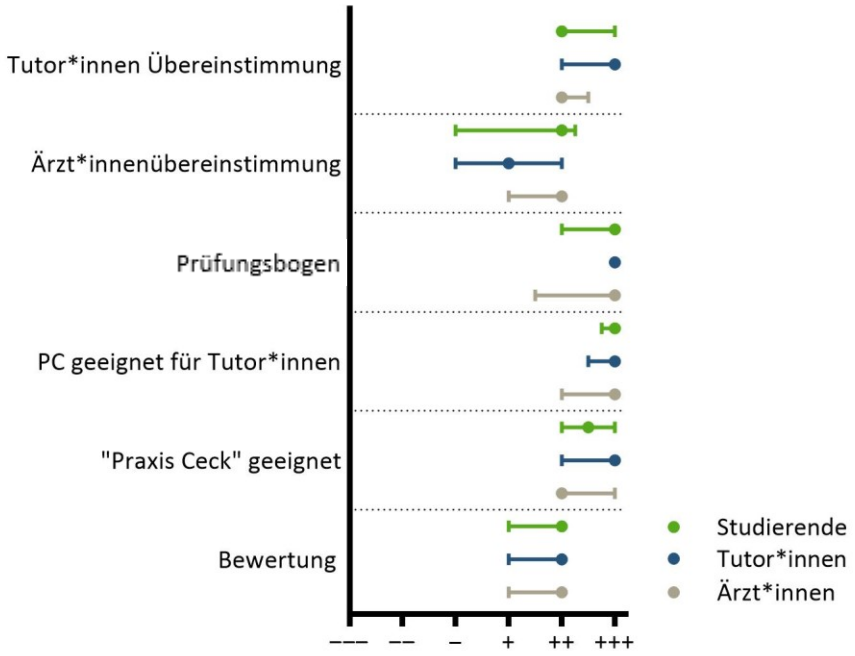


Abbildung 33: Abschlussbefragung der Teilnehmenden der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; 6-stufige Likert-ähnliche Skala: ––– = trifft gar nicht zu, +++ = trifft völlig zu. Abschlussbefragung: Abweichung = In welche Richtung vermuten Sie, dass ärztliche und studentische Bewertungen voneinander abweichen? Tutor*innen Übereinstimmung = Ich glaube, dass geschulte studentische Prüfende in den Praxis-Checks eine hohe Urteiler*innenübereinstimmung aufweisen. Ärzt*innen Übereinstimmung = Ich glaube, dass ärztliche Prüfende in den Praxis-Checks eine hohe Urteiler*innenübereinstimmung aufweisen. Prüfungsbogen = Ich empfand den Prüfungsbogen zu den Praxis-Checks als hilfreich für die Punktevergabe. PC geeignet für Tutor*innen = Das Werkzeug der Praxis-Checks halte ich für geeignet, geschulten studentischen Peer-Teachern das Prüfen zu ermöglichen. „Praxis Check“ geeignet = Das Werkzeug "Praxis-Check" halte ich für geeignet, um klinisch-praktische Leistungen zu bewerten.; Bewertung = Die Bewertung der gezeigten Prüfungen fiel mir leicht.

7.1.11 Prozent korrekt gewerteter Punkte nach Teilnehmenden der Online-PC-Simulation

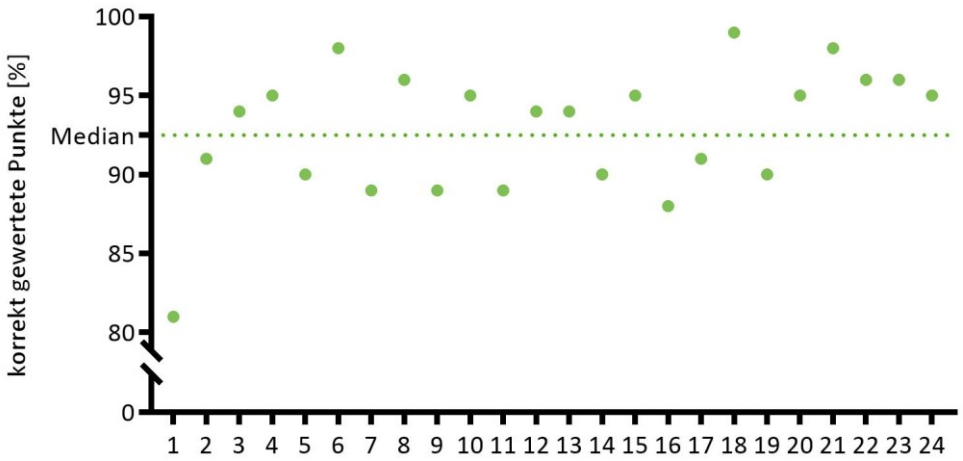


Abbildung 34: % korrekt vergebener Punkte über alle vier Prüfungsfilme der Studierenden der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; 1 – 24 = kodierte Prüfer*innen

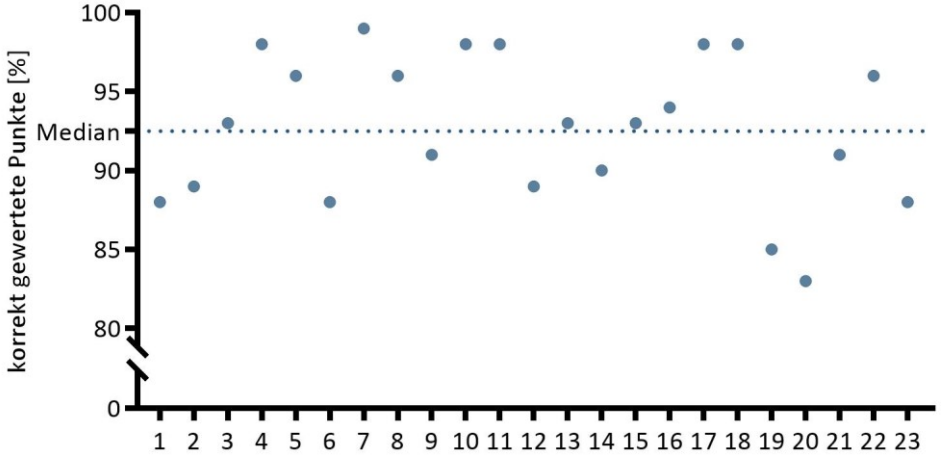


Abbildung 35: % korrekt vergebener Punkte über alle vier Prüfungsfilme der Tutor*innen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; 1 – 23 = kodierte Prüfer*innen

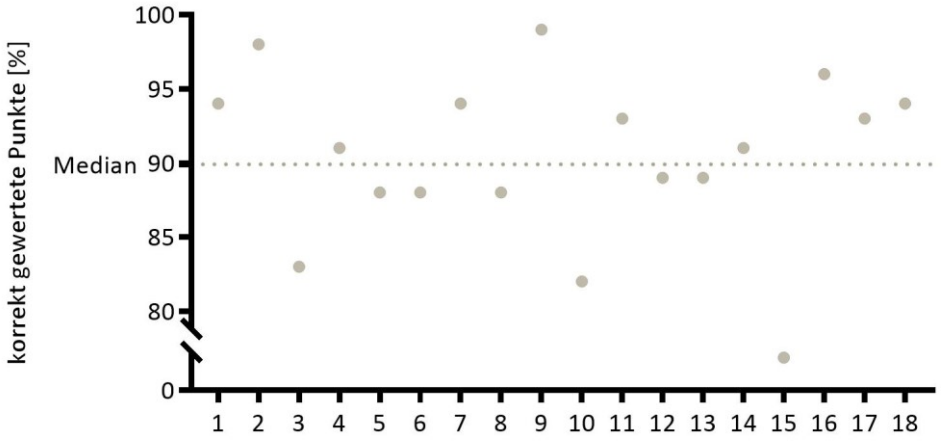


Abbildung 36: % korrekt vergebener Punkte über alle vier Prüfungsfilme der Ärzt*innen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; 1 – 18 = kodierte Prüfer*innen

7.1.12 Korrelation Erfahrung-Anzahl PCs Tutor*innen

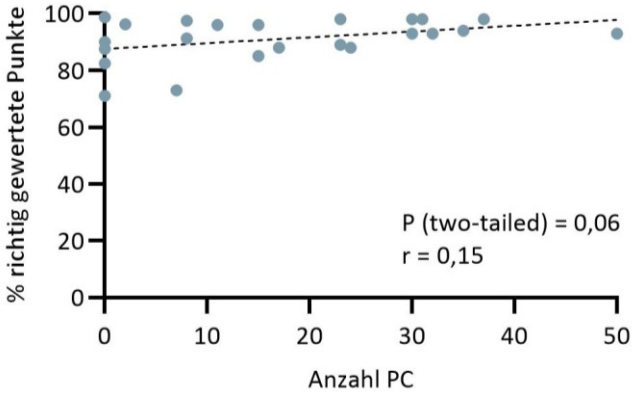


Abbildung 37: Korrelation der Anzahl korrekt gewerteter Punkte über alle vier Prüfungsfilme in % und der Anzahl Praxis-Checks (PC) von den Tutor*innen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20

7.1.13 Korrelation Alter-Punkte Ärzt*innen

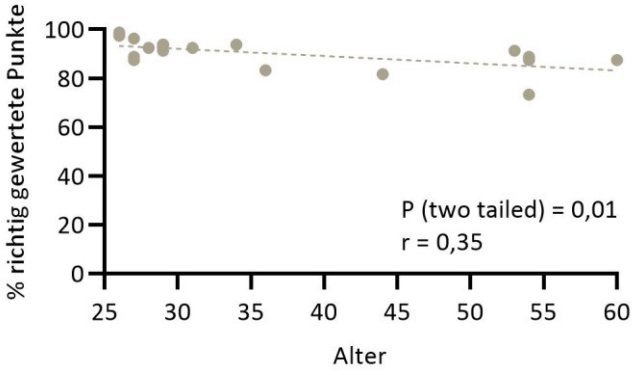


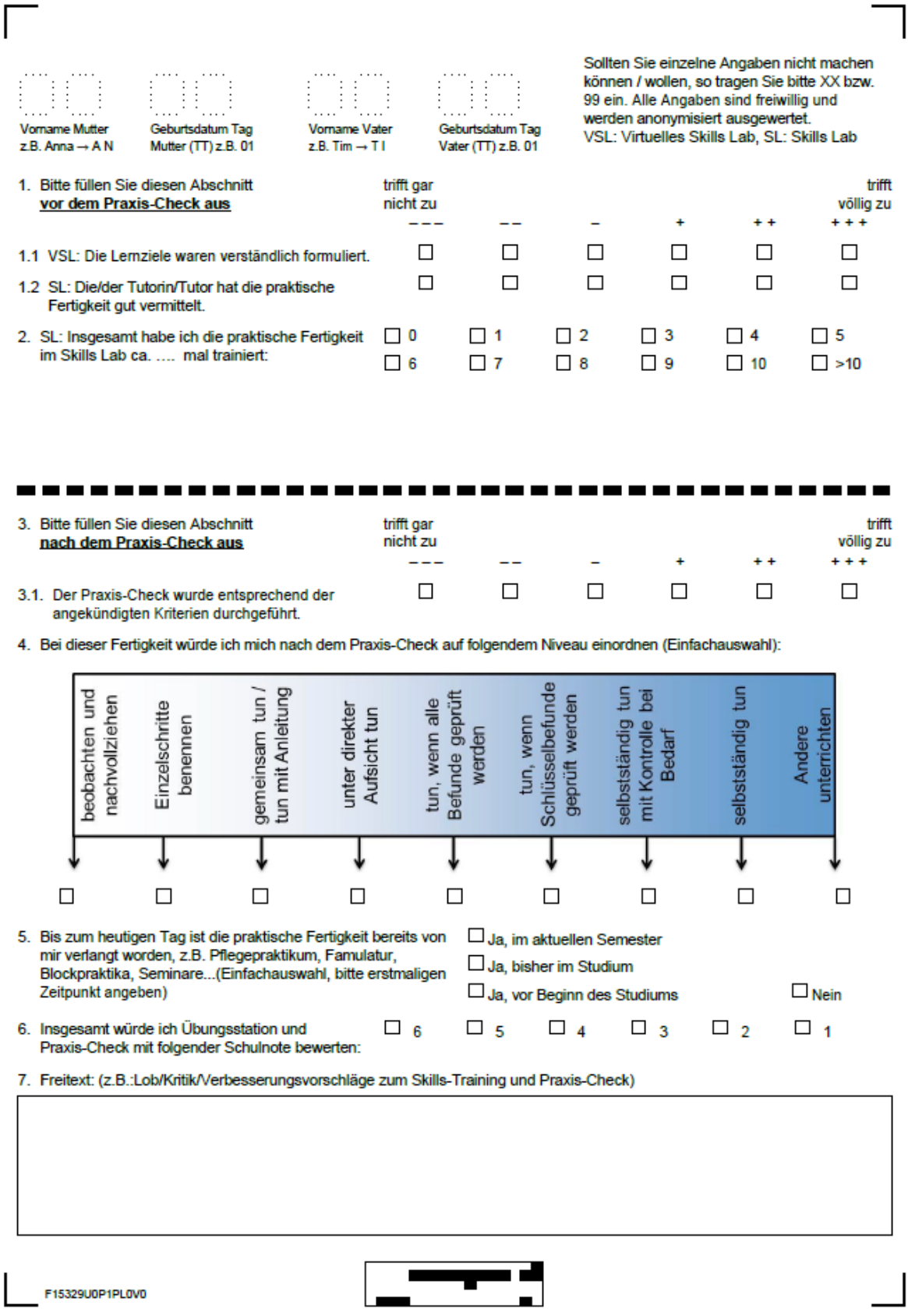
Abbildung 38: Korrelation der Anzahl korrekt gewerteter Punkte über alle vier Prüfungsfilme in % und dem Alter der Ärzt*innen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20

7.1.14 **Bewertung der PCs durch Studierende**

Tabelle 5: Übersicht der Evaluation der betrachteten Praxis-Checks (PC) im Skills Lab der Universität Ulm. Zeitraum Wintersemester 18/19 bis 19/20; Median ± Interquartilsabstand; Allgemein n = 775, Händehygiene n = 253, Vitalparameter n = 86, Nähen n = 52, Untersuchung Abdomen n = 51, Knoten n = 55, Gefäßstatus n = 35, Thoraxdrainage n = 11, AZ = Ich bin mit dem Ablauf des PC zufrieden, KR = Der PC wurde entsprechend der angekündigten Kriterien durchgeführt, LZ = Die Lernziele waren verständlich formuliert, VM = Die*der Tutor*in hat die praktische Fähigkeit gut vermittelt, GL = Insgesamt würde ich Übungsstation und PC mit folgender Schulnote bewerten, ZÜ = Von der gegebenen Prüfungszeit waren noch … Sekunden übrig, E-TU = Bei dieser Fertigkeit schätze ich den Prüfling nach dem PC auf folgendem Niveau ein (Einschätzung Tutor), SE-TN = Bei dieser Fertigkeit würde ich mich nach dem PC auf folgendem Niveau einordnen (Selbsteinschätzung Teilnehmer), SB = Verglichen mit der objektiven Note würde ich die Leistung subjektiv bewerten als… , DP = Durchschnitt Punkte; SL = Insgesamt habe ich die praktische Fertigkeit im Skills Lab ca. ….. mal trainiert.

PC	Allgemein	Hygiene	Vitalparameter	Nähen	Knoten	Untersuchung Abdomen	Gefäßstatus	Thorax-drainage
Skala	1 = trifft gar nicht zu; 2 = trifft nicht zu; 3 = trifft kaum zu, 4 = trifft etwas zu; 5 = trifft zu; 6 = trifft völlig zu							
AZ	6 (5;6)	5 (4;6)	6 (5;6)	5 (5;6)	6 (5;6)	5,5 (4;6)	6 (5;6)	6 (5;6)
KR	6 (6;6)	6 (6;6)	6 (6;6)	6 (6;6)	6 (6;6)	6 (6;6)	6 (6;6)	6 (5;6)
LZ	6 (5;6)	6 (5;6)	6 (5;6)	6 (6;6)	6 (6;6)	6 (5,75;6)	6 (6;6)	5 (4;5)
VM	6 (5;6)	6 (5;6)	6 (6;6)	6 (5;6)	6 (6;6)	6 (5;6)	6 (6;6)	6 (5;6)
Skala	1 = sehr gut; 2 = gut; 3 = befriedigend; 4 = ausreichend; 5 = mangelhaft; 6 = ungenügend							
GL	6 (5;6)	6 (5;6)	6 (5;6)	6 (5;6)	6 (6;6)	6 (5;6)	6 (6;6)	5 (5;6)
Skala	1 > 120s ; 2 = 91-120s ; 3 = 61-90s ; 4 = 31-60s ; 5 = 1-30s ; 6 = Zeit abgelaufen							
ZÜ	2 (1;4)	3 (2;4)	4 (2;5)	2 (1;5)	1 (1;1)	4 (2;5)	2 (1;4)	1 (1;1)
Skala	1=beobachten und nachvollziehen; 2=Einzelschritte benennen; 3=gemeinsam tun/ tun mit Anleitung; 4=unter direkter Aufsicht tun; 5=tun, wenn alle Befunde geprüft werden; 6=tun, wenn Schlüsselbefunde geprüft werden; 7= selbstständig tun mit Kontrolle bei Bedarf; 8=selbstständig tun; 9= Andere unterrichten							
E-TU	8 (7;8)	8 (7;8)	8 (8;8)	7 (6;8)	8 (7;9)	8 (7;8)	8 (8;8)	7 (7;9)
SE-TN	8 (7;8)	8 (7;8)	8 (8;8)	7 (6,25;8)	8 (7;9)	8 (7;8)	8 (7;8)	7 (4;8)
Skala	1=viel schlechter; 2=schlechter;3=gleich, 4=besser; 5=viel besser							
SB	3 (3;3)	3 (3;3)	3 (3;3,25)	3 (3;3)	3 (3;3)	3 (3;3)	3 (3;4)	3 (2;3)
Skala	Punkte 1-20							
DP	20(18;20)	20(19;20)	19(18;20)	19(18;20)	20(20;2)	19(18;20)	20(19;20)	20(19;20)
SL	4(3;5)	3(2;4)	3(2,75;5)	6(4;10,7)	5(3;11)	4(3;6)	4(3;5)	4(3;5)

7.2 Hauptprüfer*in Prüfungsbogen



Vorname Mutter z.B. Anna → A N | Geburtsdatum Tag Mutter (TT) z.B. 01 | Vorname Vater z.B. Tim → T I | Geburtsdatum Tag Vater (TT) z.B. 01

Sollten Sie einzelne Angaben nicht machen können / wollen, so tragen Sie bitte XX bzw. 99 ein. Alle Angaben sind freiwillig und werden anonymisiert ausgewertet.
VSL: Virtuelles Skills Lab, SL: Skills Lab

1. Bitte füllen Sie diesen Abschnitt **vor dem Praxis-Check aus** — trifft gar nicht zu (– – –) … trifft völlig zu (+ + +)

1.1 VSL: Die Lernziele waren verständlich formuliert. ☐ – – – ☐ – – ☐ – ☐ + ☐ + + ☐ + + +

1.2 SL: Die/der Tutorin/Tutor hat die praktische Fertigkeit gut vermittelt. ☐ – – – ☐ – – ☐ – ☐ + ☐ + + ☐ + + +

2. SL: Insgesamt habe ich die praktische Fertigkeit im Skills Lab ca. …. mal trainiert: ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ >10

3. Bitte füllen Sie diesen Abschnitt **nach dem Praxis-Check aus** — trifft gar nicht zu (– – –) … trifft völlig zu (+ + +)

3.1. Der Praxis-Check wurde entsprechend der angekündigten Kriterien durchgeführt. ☐ – – – ☐ – – ☐ – ☐ + ☐ + + ☐ + + +

4. Bei dieser Fertigkeit würde ich mich nach dem Praxis-Check auf folgendem Niveau einordnen (Einfachauswahl):

☐ beobachten und nachvollziehen
☐ Einzelschritte benennen
☐ gemeinsam tun / tun mit Anleitung
☐ unter direkter Aufsicht tun
☐ tun, wenn alle Befunde geprüft werden
☐ tun, wenn Schlüsselbefunde geprüft werden
☐ selbstständig tun mit Kontrolle bei Bedarf
☐ selbstständig tun
☐ Andere unterrichten

5. Bis zum heutigen Tag ist die praktische Fertigkeit bereits von mir verlangt worden, z.B. Pflegepraktikum, Famulatur, Blockpraktika, Seminare...(Einfachauswahl, bitte erstmaligen Zeitpunkt angeben)
☐ Ja, im aktuellen Semester
☐ Ja, bisher im Studium
☐ Ja, vor Beginn des Studiums
☐ Nein

6. Insgesamt würde ich Übungsstation und Praxis-Check mit folgender Schulnote bewerten: ☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

7. Freitext: (z.B.:Lob/Kritik/Verbesserungsvorschläge zum Skills-Training und Praxis-Check)

F15329U0P1PL0V0

Abbildung 39: Vorderseite des verwendeten Prüfungsbogens im Skills Lab der Universität Ulm. Zeitraum: Wintersemester 18/20 bis Wintersemester 19/20. Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

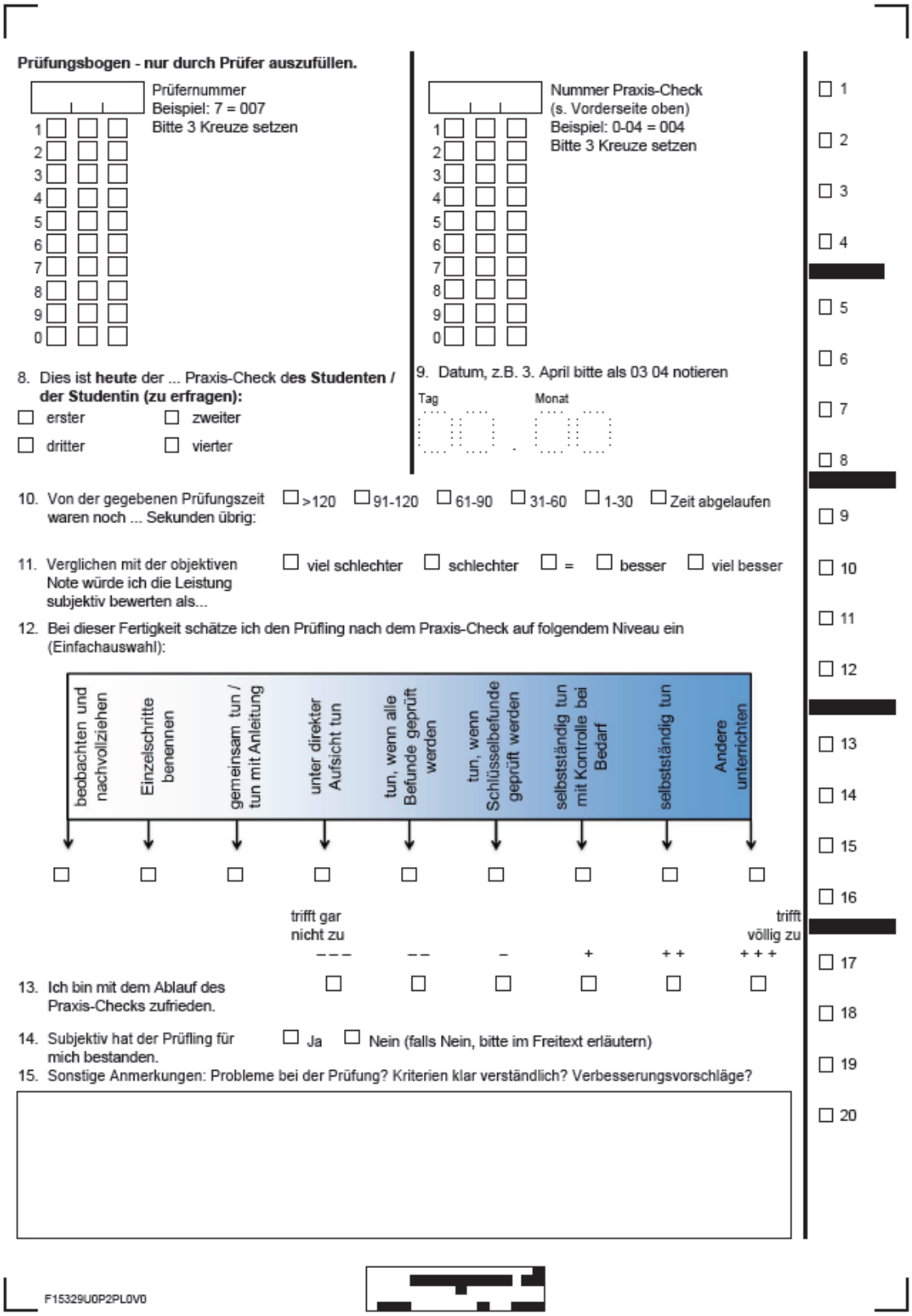


Abbildung 40: Rückseite des verwendeten Prüfungsbogens im Skills Lab der Universität Ulm. Zeitraum: Wintersemester 18/20 bis Wintersemester 19/20; Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

Untersuchung Abdomen – Praxis-Check 4-02 – Skills-in-Context: Kom./Hy./Dok.

20 Punkte – Zeitrahmen: **10 Minuten** – **Proband erforderlich** – Version Wintersemester 2020

A1: *„Sie sollen…“* **1 Punkt:** Begrüßung und Vorstellung mit Name und Funktion, eindeutige Identifikation des Probanden (Name, Geburtsdatum aktiv nennen lassen)	1	☐ 1
1 Punkt: Ankündigung der Untersuchung + Bitte, den Bauch freizumachen + Herstellen einer geschützten Untersuchungsumgebung	2	☐ 2
1 Punkt: hygienische Händedesinfektion (30 Sekunden) B1	3	☐ 3
1 Punkt: Auskultation Darmgeräusche in allen vier Quadranten	4	☐ 4
1 Punkt: Auskultation Strömungsgeräusch Aorta	5	☐ 5
1 Punkt: Perkussion in allen vier Quadranten	6	☐ 6
1 Punkt: Palpation oberflächlich und tief in allen vier Quadranten	7	☐ 7
1 Punkt: Prüfung Loslass-Schmerz	8	☐ 8
1 Punkt: Auskultation vor Perkussion vor Palpation	9	☐ 9
1 Punkt: Untersuchung Leber: tiefe Palpation unter Rippenbogen + Perkussion oder Kratzauskultation	10	☐ 10
1 Punkt: Untersuchung Milz: Palpation	11	☐ 11
1 Punkt: Prüfung Klopfschmerz Nierenlager beidseits	12	☐ 12
1 Punkt: verbales Beenden der Untersuchung + Proband auffordern, sich wieder anzukleiden + Verabschiedung / Ausblick	13	☐ 13
1 Punkt: Desinfektion des Stethoskops + ggfs. Maßband B2	14	☐ 14
1 Punkt: hygienische Händedesinfektion (30 Sekunden) B1	15	☐ 15
A2: *„Welche beiden Untersuchungen (…)?“* **1 Punkt:** Hernien, rektale Untersuchung	16	☐ 16
A3: *„Bitte dokumentieren Sie jetzt den Befund.“* **1 Punkt:** Stammdaten, Ernährungszustand + Leberhautzeichen + OP-Narben B3 + 4	17	☐ 17
1 Punkt: Auskultationsbefund + Perkussionsbefund B5	18	☐ 18
1 Punkt: Palpationsbefund B6	19	☐ 19
1 Punkt: Befund Leber + Befund Milz + Befund Nieren B7	20	☐ 20



uulm

… zur Vorbereitung (Praxis-Check, Famulatur, …) im Virtuellen Skills Lab abrufbar

Abbildung 41: Prüfungsbogen im Skills Lab der Universität Ulm im Zeitraum Wintersemester 18/20 bis Wintersemester 19/20, kombiniert mit Praxis-Check Bogen, während der Prüfung. Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

7.3 Co-Prüfer*in Prüfungsbogen

Code
Teilnehmer/-in

Vorname Mutter z.B. Anna → A N | Geburtsdatum Tag Mutter (TT) z.B. 01 | Vorname Vater z.B. Tim → T I | Geburtsdatum Tag Vater (TT) z.B. 01

Sollten Sie einzelne Angaben nicht machen können / wollen, so tragen Sie bitte XX bzw. 99 ein. Alle Angaben sind freiwillig und werden anonymisiert ausgewertet.
VSL: Virtuelles Skills Lab, SL: Skills Lab

Code **Co-Prüfer/in**

Vorname Mutter z.B. Anna → A N | Geburtsdatum Tag Mutter (TT) z.B. 01 | Vorname Vater z.B. Tim → T I | Geburtsdatum Tag Vater (TT) z.B. 01

Sollten Sie einzelne Angaben nicht machen können / wollen, so tragen Sie bitte XX bzw. 99 ein. Alle Angaben sind freiwillig und werden anonymisiert ausgewertet.
VSL: Virtuelles Skills Lab, SL: Skills Lab

1. Ich bin ... (Mehrfachnennungen sind möglich):

☐ Student/-in ☐ Skills-Lab-Tutor/-in ☐ Arzt/Ärztin

2. Als **Lernende/r** habe ich diesen Praxis-Check ... (Mehrfachnennungen sind möglich):

☐ noch nie trainiert ☐ bereits trainiert, aber ihn noch nicht erfolgreich abgelegt ☐ bei einem/r Tutor/-in im Skills Lab bereits erfolgreich abgelegt ☐ bei einem/r Arzt/Ärztin im Skills Lab bereits erfolgreich abgelegt

3. Als **Prüfende/r** habe ich diesen Praxis-Check ... (Mehrfachnennungen sind möglich):

☐ bereits als Coprüfer durchgeführt ☐ bereits als Hauptprüfer durchgeführt

F15331U0P1PL0V0

Abbildung 42: Vorderseite des verwendeten Co-Prüfer*innenbogens im Skills Lab der Universität Ulm im Sommersemester 19 und Wintersemester 19/20. Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

Prüfungsbogen - nur durch Prüfer auszufüllen.

Prüfernummer
Beispiel: 7 = 007
Bitte 3 Kreuze setzen

1	☐	☐	☐
2	☐	☐	☐
3	☐	☐	☐
4	☐	☐	☐
5	☐	☐	☐
6	☐	☐	☐
7	☐	☐	☐
8	☐	☐	☐
9	☐	☐	☐
0	☐	☐	☐

Nummer Praxis-Check
(s. Vorderseite oben)
Beispiel: 0-04 = 004
Bitte 3 Kreuze setzen

1	☐	☐	☐
2	☐	☐	☐
3	☐	☐	☐
4	☐	☐	☐
5	☐	☐	☐
6	☐	☐	☐
7	☐	☐	☐
8	☐	☐	☐
9	☐	☐	☐
0	☐	☐	☐

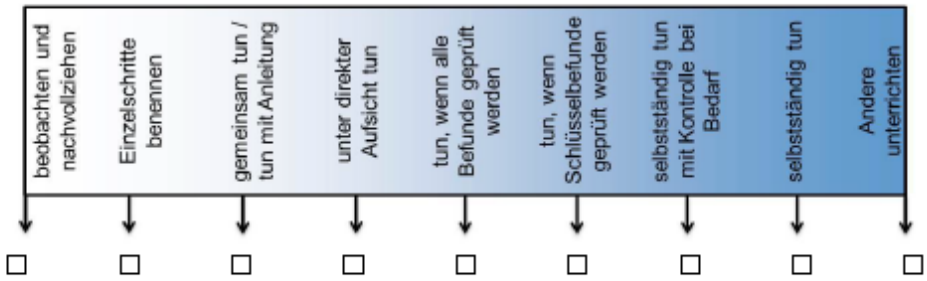
4. Datum, z.B. 3. April bitte als 03 04 notieren

Tag ..... Monat .....

5. Von der gegebenen Prüfungszeit waren noch ... Sekunden übrig: ☐ >120 ☐ 91-120 ☐ 61-90 ☐ 31-60 ☐ 1-30 ☐ Zeit abgelaufen

6. Verglichen mit der objektiven Note würde ich die Leistung subjektiv bewerten als... ☐ viel schlechter ☐ schlechter ☐ = ☐ besser ☐ viel besser

7. Bei dieser Fertigkeit schätze ich den Prüfling nach dem Praxis-Check auf folgendem Niveau ein (Einfachauswahl):



	trifft gar nicht zu					trifft völlig zu
	– – –	– –	–	+	+ +	+ + +
8. Ich bin mit dem Ablauf des Praxis-Checks zufrieden.	☐	☐	☐	☐	☐	☐

9. Subjektiv hat der Prüfling für mich bestanden. ☐ Ja ☐ Nein (falls Nein, bitte im Freitext erläutern)

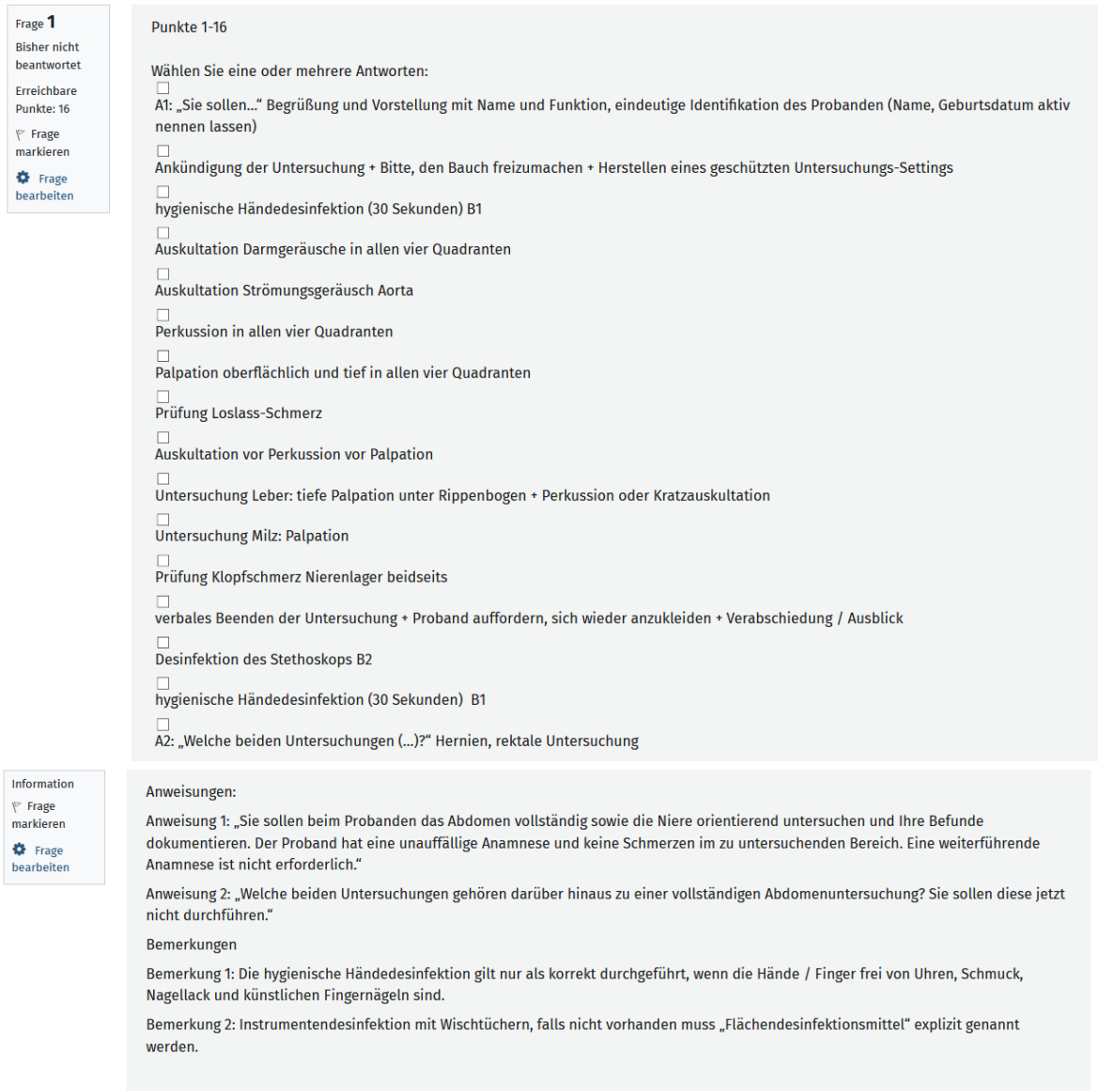
10. Sonstige Anmerkungen: Probleme bei der Prüfung? Kriterien klar verständlich? Verbesserungsvorschläge?

F15331U0P2PL0V0

☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6
☐ 7
☐ 8
☐ 9
☐ 10
☐ 11
☐ 12
☐ 13
☐ 14
☐ 15
☐ 16
☐ 17
☐ 18
☐ 19
☐ 20

Abbildung 43: Rückseite des verwendeten Co-Prüfer*innenbogens im Skills Lab der Universität Ulm im Sommersemester 19 und Wintersemester 19/20. Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

7.4 Digitaler Prüfungsbogen mit PC-Kriterien



Frage 1
Bisher nicht beantwortet
Erreichbare Punkte: 16
Frage markieren
Frage bearbeiten

Punkte 1-16

Wählen Sie eine oder mehrere Antworten:

- ☐ A1: „Sie sollen…“ Begrüßung und Vorstellung mit Name und Funktion, eindeutige Identifikation des Probanden (Name, Geburtsdatum aktiv nennen lassen)
- ☐ Ankündigung der Untersuchung + Bitte, den Bauch freizumachen + Herstellen eines geschützten Untersuchungs-Settings
- ☐ hygienische Händedesinfektion (30 Sekunden) B1
- ☐ Auskultation Darmgeräusche in allen vier Quadranten
- ☐ Auskultation Strömungsgeräusch Aorta
- ☐ Perkussion in allen vier Quadranten
- ☐ Palpation oberflächlich und tief in allen vier Quadranten
- ☐ Prüfung Loslass-Schmerz
- ☐ Auskultation vor Perkussion vor Palpation
- ☐ Untersuchung Leber: tiefe Palpation unter Rippenbogen + Perkussion oder Kratzauskultation
- ☐ Untersuchung Milz: Palpation
- ☐ Prüfung Klopfschmerz Nierenlager beidseits
- ☐ verbales Beenden der Untersuchung + Proband auffordern, sich wieder anzukleiden + Verabschiedung / Ausblick
- ☐ Desinfektion des Stethoskops B2
- ☐ hygienische Händedesinfektion (30 Sekunden) B1
- ☐ A2: „Welche beiden Untersuchungen (…)?“ Hernien, rektale Untersuchung

Information
Frage markieren
Frage bearbeiten

Anweisungen:

Anweisung 1: „Sie sollen beim Probanden das Abdomen vollständig sowie die Niere orientierend untersuchen und Ihre Befunde dokumentieren. Der Proband hat eine unauffällige Anamnese und keine Schmerzen im zu untersuchenden Bereich. Eine weiterführende Anamnese ist nicht erforderlich.“

Anweisung 2: „Welche beiden Untersuchungen gehören darüber hinaus zu einer vollständigen Abdomenuntersuchung? Sie sollen diese jetzt nicht durchführen.“

Bemerkungen

Bemerkung 1: Die hygienische Händedesinfektion gilt nur als korrekt durchgeführt, wenn die Hände / Finger frei von Uhren, Schmuck, Nagellack und künstlichen Fingernägeln sind.

Bemerkung 2: Instrumentendesinfektion mit Wischtüchern, falls nicht vorhanden muss „Flächendesinfektionsmittel“ explizit genannt werden.

Abbildung 44: Digitaler Prüfungsbogen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Seite 1; Dieser Text wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

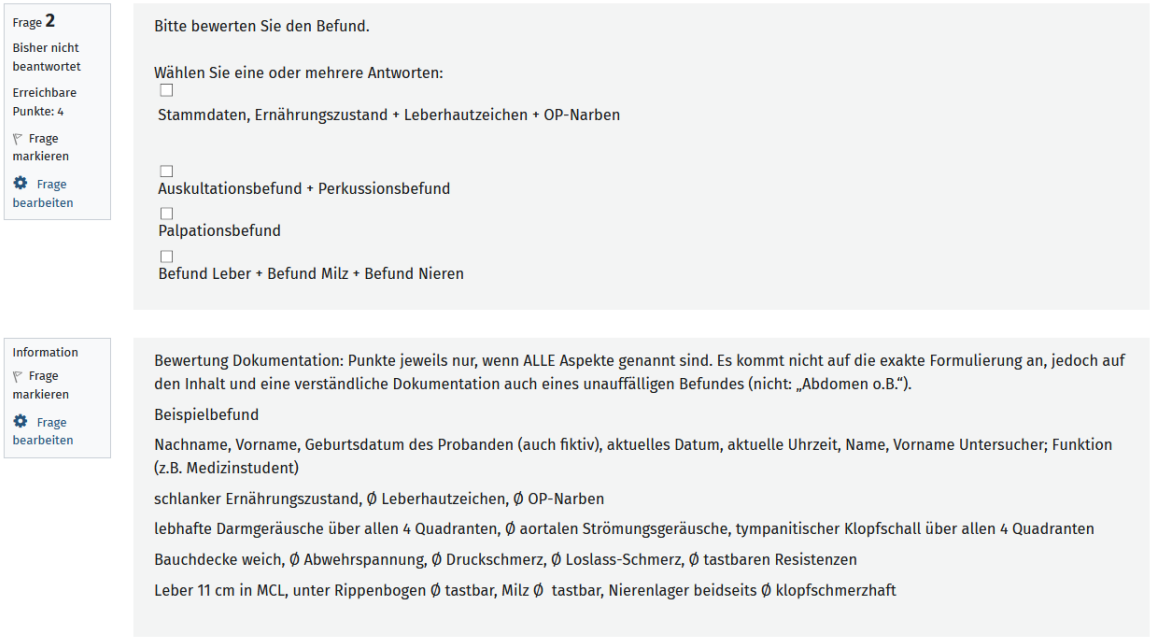


Abbildung 45: Digitaler Prüfungsbogen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Seite 2; Dieser Text wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

7.5 Impression der Online-PC-Simulation – geteilter Bildschirm

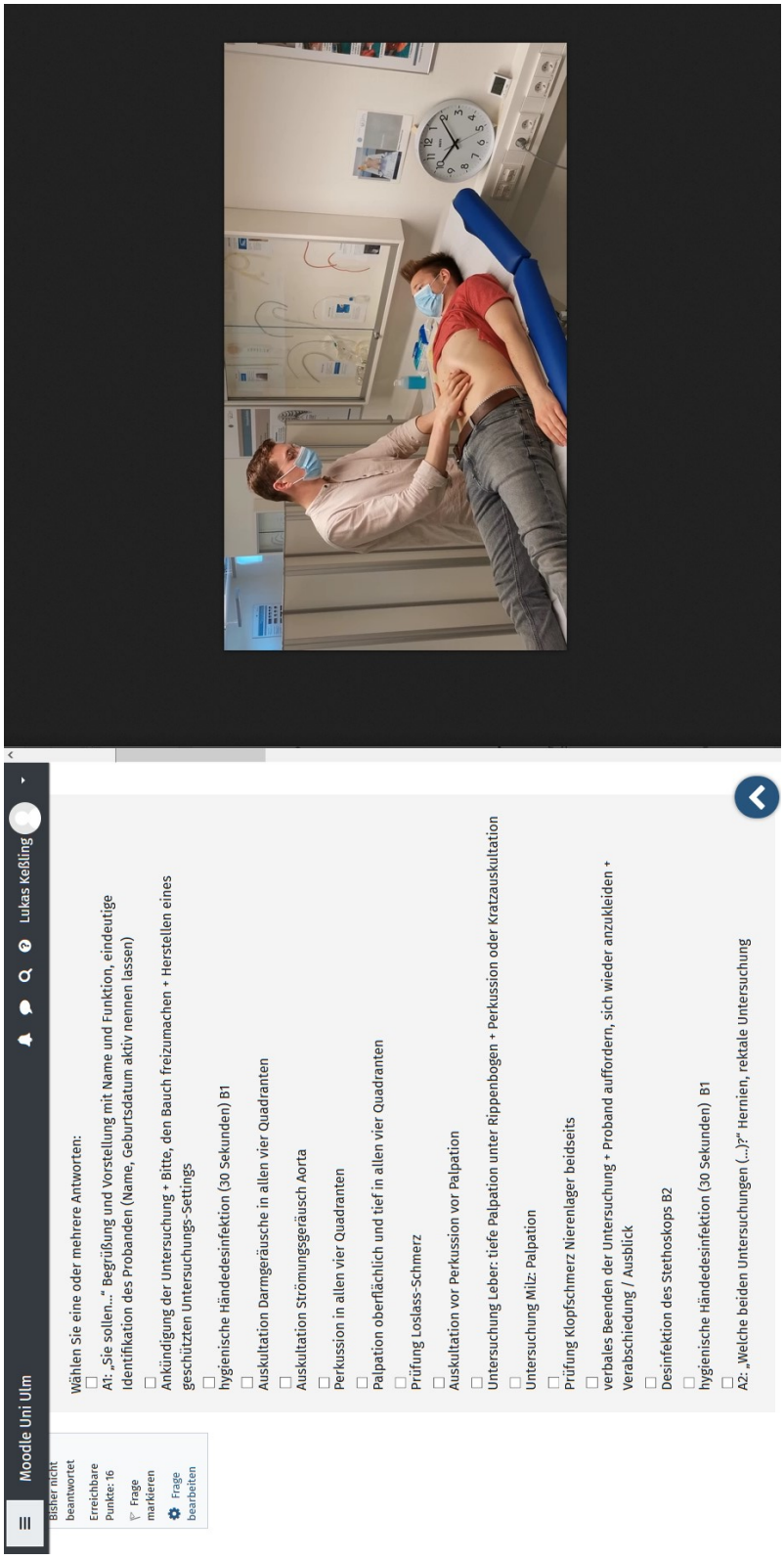


Abbildung 46: Impressionen der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm im Sommersemester 20. Beispiel geteilter Bildschirm (links) und Video (rechts) während der Online-Simulation

7.6 Praxis-Check Bogen Untersuchung Abdomen

Untersuchung Abdomen – Praxis-Check 4-02 – Skills-in-Context: Kom./Hy./Dok.

20 Punkte – Zeitrahmen: **10 Minuten** – **Proband erforderlich** – Version Wintersemester 2020

A1: *„Sie sollen…“* **1 Punkt:** Begrüßung und Vorstellung mit Name und Funktion, eindeutige Identifikation des Probanden (Name, Geburtsdatum aktiv nennen lassen)	1
1 Punkt: Ankündigung der Untersuchung + Bitte, den Bauch freizumachen + Herstellen einer geschützten Untersuchungsumgebung	2
1 Punkt: hygienische Händedesinfektion (30 Sekunden) B1	3
1 Punkt: Auskultation Darmgeräusche in allen vier Quadranten	4
1 Punkt: Auskultation Strömungsgeräusch Aorta	5
1 Punkt: Perkussion in allen vier Quadranten	6
1 Punkt: Palpation oberflächlich und tief in allen vier Quadranten	7
1 Punkt: Prüfung Loslass-Schmerz	8
1 Punkt: Auskultation vor Perkussion vor Palpation	9
1 Punkt: Untersuchung Leber: tiefe Palpation unter Rippenbogen + Perkussion oder Kratzauskultation	10
1 Punkt: Untersuchung Milz: Palpation	11
1 Punkt: Prüfung Klopfschmerz Nierenlager beidseits	12
1 Punkt: verbales Beenden der Untersuchung + Proband auffordern, sich wieder anzukleiden + Verabschiedung / Ausblick	13
1 Punkt: Desinfektion des Stethoskops + ggfs. Maßband B2	14
1 Punkt: hygienische Händedesinfektion (30 Sekunden) B1	15
A2: *„Welche beiden Untersuchungen (…)?“* **1 Punkt:** Hernien, rektale Untersuchung	16
A3: *„Bitte dokumentieren Sie jetzt den Befund.“* **1 Punkt:** Stammdaten, Ernährungszustand + Leberhautzeichen + OP-Narben B3 + 4	17
1 Punkt: Auskultationsbefund + Perkussionsbefund B5	18
1 Punkt: Palpationsbefund B6	19
1 Punkt: Befund Leber + Befund Milz + Befund Nieren B7	20



ulm university universität **uulm**

… zur Vorbereitung (Praxis-Check, Famulatur, …) im Virtuellen Skills Lab abrufbar

Abbildung 47: Vorderseite des verwendeten Praxis-Checks Untersuchung Abdomen im Wintersemester 18/19, Sommersemester 19 und Wintersemester 19/20 und in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm im Sommersemester 20. Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

Anweisungen

Anweisung 1: *„Sie sollen beim Probanden das Abdomen vollständig sowie die Niere orientierend untersuchen und Ihre Befunde dokumentieren. Der Proband hat eine unauffällige Anamnese und keine Schmerzen im zu untersuchenden Bereich. Eine weiterführende Anamnese ist nicht erforderlich.“*

Anweisung 2: *„Welche beiden Untersuchungen gehören darüber hinaus zu einer vollständigen Abdomenuntersuchung? (Sie sollen diese jetzt nicht durchführen.)“*

Bemerkungen

Bemerkung 1: Die hygienische Händedesinfektion gilt nur als korrekt durchgeführt, wenn die Hände / Finger frei von Uhren, Schmuck, Nagellack und künstlichen Fingernägeln sind.

Bemerkung 2: Instrumentendesinfektion mit Wischtüchern, falls nicht vorhanden **muss** „Flächendesinfektionsmittel“ explizit genannt werden

Bemerkung 3: Bewertung Dokumentation: Punkte jeweils nur, wenn ALLE Aspekte genannt sind. Es kommt nicht auf die exakte Formulierung an, jedoch auf den Inhalt und eine verständliche Dokumentation auch eines unauffälligen Befundes *(nicht: „Abdomen o.B.“)*

Beispielbefund
Nachname, Vorname, Geburtsdatum des Probanden (auch fiktiv) ***aktuelles Datum, aktuelle Uhrzeit*** ***Name, Vorname Untersucher; Funktion (z.B. Medizinstudent)*** (Bemerkung 4)
schlanker Ernährungszustand, Ø Leberhautzeichen, Ø OP-Narben (Bemerkung 4)
lebhafte Darmgeräusche über allen 4 Quadranten, Ø aortalen Strömungsgeräusche tympanitischer Klopfschall über allen 4 Quadranten (Bemerkung 5)
Bauchdecke weich, Ø Abwehrspannung, Ø Druckschmerz, Ø Loslass-Schmerz, Ø tastbaren Resistenzen (Bemerkung 6)
Leber 11 cm in MCL, unter Rippenbogen Ø tastbar, Milz Ø nicht tastbar, Nierenlager beidseits Ø klopfschmerzhaft (Bemerkung 7)

Stationsaufbau

Untersuchungsliege, zwei spanische Wände, Desinfektionsmittelspender, Stethoskop, Spatel, Lineal oder Maßband, Flächendesinfektionsmittel (Wischtücher), Stift, Befundbogen

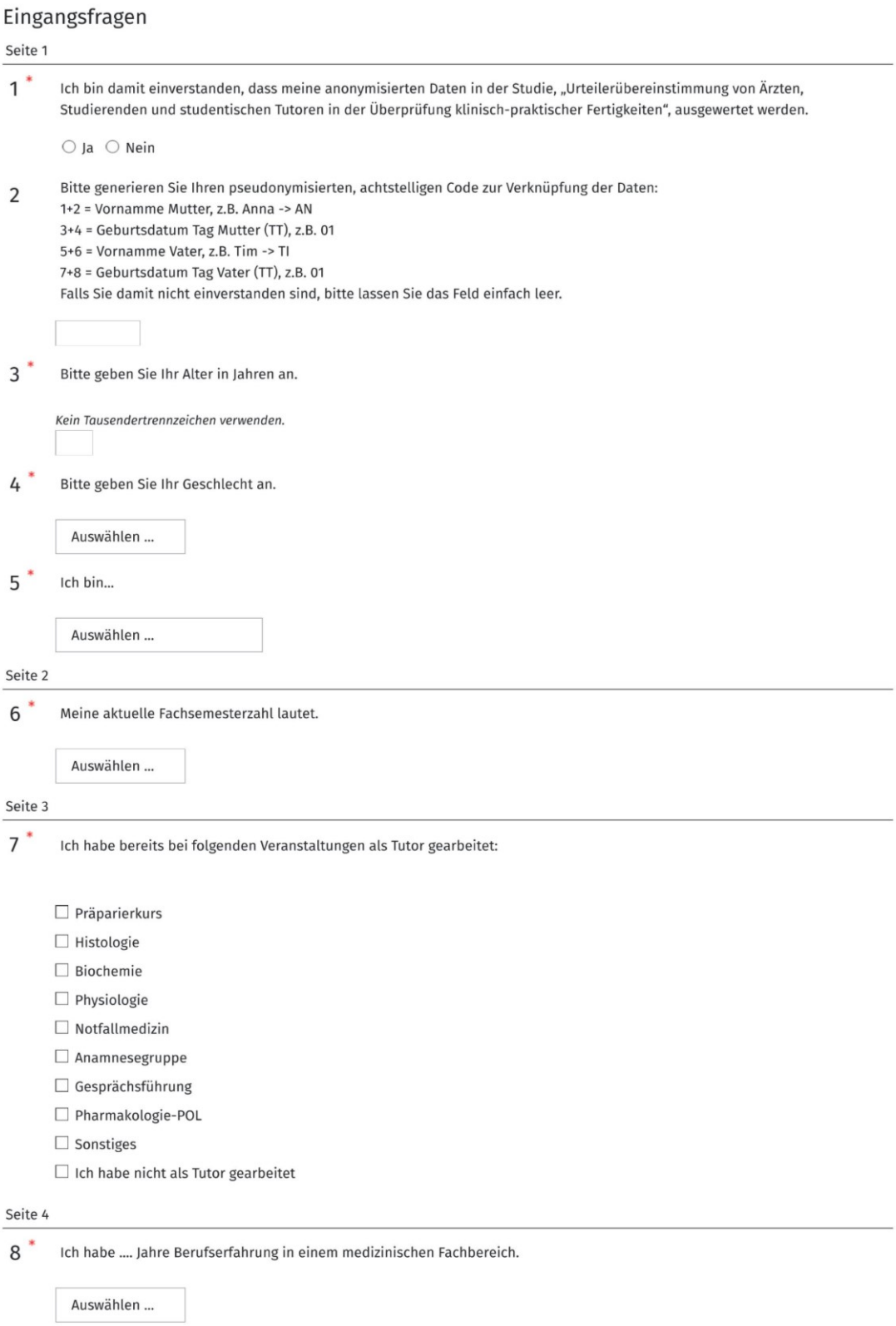
Weiteres Personal: **Proband,** passiv, kooperativ, liegt auf Untersuchungsliege

Tipps und Tricks:

Nutzen Sie die Zeit der hygienischen Händedesinfektion, um den Patienten auf die anstehenden Untersuchungen vorzubereiten und ihn anzuleiten.

Abbildung 48: Rückseite des verwendeten Praxis-Checks Untersuchung im Wintersemester 18/19, Sommersemester 19 und Wintersemester 19/20 und in der Online-Praxis-Check-Simulation vom Skills Lab der Universität Ulm im Sommersemester 20. Dieser Bogen wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

7.7 Eingangsfragebogen



Eingangsfragen

Seite 1

1 * Ich bin damit einverstanden, dass meine anonymisierten Daten in der Studie, „Urteilerübereinstimmung von Ärzten, Studierenden und studentischen Tutoren in der Überprüfung klinisch-praktischer Fertigkeiten“, ausgewertet werden.

○ Ja ○ Nein

2 Bitte generieren Sie Ihren pseudonymisierten, achtstelligen Code zur Verknüpfung der Daten:
1+2 = Vornamme Mutter, z.B. Anna -> AN
3+4 = Geburtsdatum Tag Mutter (TT), z.B. 01
5+6 = Vornamme Vater, z.B. Tim -> TI
7+8 = Geburtsdatum Tag Vater (TT), z.B. 01
Falls Sie damit nicht einverstanden sind, bitte lassen Sie das Feld einfach leer.

3 * Bitte geben Sie Ihr Alter in Jahren an.

Kein Tausendertrennzeichen verwenden.

4 * Bitte geben Sie Ihr Geschlecht an.

Auswählen ...

5 * Ich bin...

Auswählen ...

Seite 2

6 * Meine aktuelle Fachsemesterzahl lautet.

Auswählen ...

Seite 3

7 * Ich habe bereits bei folgenden Veranstaltungen als Tutor gearbeitet:

☐ Präparierkurs
☐ Histologie
☐ Biochemie
☐ Physiologie
☐ Notfallmedizin
☐ Anamnesegruppe
☐ Gesprächsführung
☐ Pharmakologie-POL
☐ Sonstiges
☐ Ich habe nicht als Tutor gearbeitet

Seite 4

8 * Ich habe .... Jahre Berufserfahrung in einem medizinischen Fachbereich.

Auswählen ...

Abbildung 49: Eingangsfragebogen der Online-Praxis-Check-Simulation Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Seite 1; Dieser Text wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.



Seite 5

9 * Bitte geben Sie Ihren Fachbereich (z.B: Innere Medizin und Kardiologie, Anatomie) an.

Seite 6

10 * Bitte geben Sie Ihre Berufserfahrung in Jahren an:

Auswählen ...

Seite 7

11 * Ich habe den Praxis-Check Abdomen bereits ... (mehrfache Antworten möglich)

- ☐ geübt
- ☐ erfolgreich abgelegt
- ☐ cogeprüft
- ☐ geprüft
- ☐ keines von den oben genannten

12 * Ich habe Erfahrung .... (mehrfache Antworten möglich)

- ☐ als Skillslabtutor
- ☐ als Prüfer
- ☐ als Staatsexamensprüfer
- ☐ in der Lehre von Studierenden
- ☐ als Prüfer in Naturwissenschaften bzw. anderen Fachgebieten
- ☐ keine Antwort zutreffend

13 * Ich fühle mich grundsätzlich dazu befähigt, eine strukturierte Abdomen Untersuchung durchzuführen.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

14 * Ich fühle mich dazu befähigt, die Qualität einer durchgeführten Abdomen Untersuchung zu beurteilen.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

15 * Ich fühle mich grundsätzlich sicher darin, studentische Leistungen in praktischen Fertigkeiten einzuschätzen.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

Seite 8

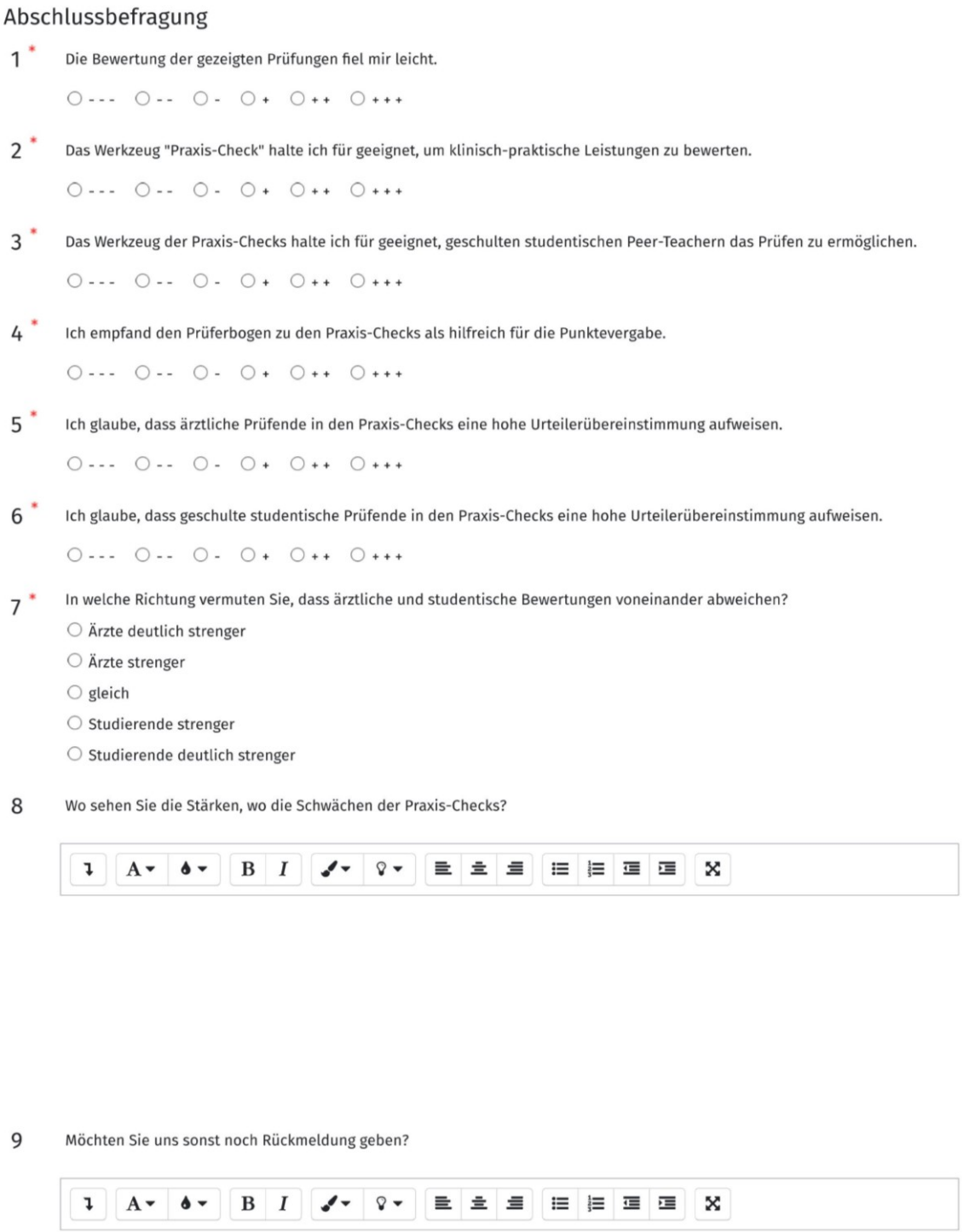
16 * Ich halte mich für einen objektiven Prüfer / eine objektive Prüferin.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

Fenster schließen

Abbildung 50: Eingangsfragebogen der Online-Praxis-Check-Simulation Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20; Seite 2; Dieser Text wurde mit dem Einverständnis von Dr. David Messerer und Astrid Horneffer verwendet.

7.8 Abschlussfragebogen


Abschlussbefragung

1 * Die Bewertung der gezeigten Prüfungen fiel mir leicht.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

2 * Das Werkzeug "Praxis-Check" halte ich für geeignet, um klinisch-praktische Leistungen zu bewerten.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

3 * Das Werkzeug der Praxis-Checks halte ich für geeignet, geschulten studentischen Peer-Teachern das Prüfen zu ermöglichen.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

4 * Ich empfand den Prüferbogen zu den Praxis-Checks als hilfreich für die Punktevergabe.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

5 * Ich glaube, dass ärztliche Prüfende in den Praxis-Checks eine hohe Urteilerübereinstimmung aufweisen.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

6 * Ich glaube, dass geschulte studentische Prüfende in den Praxis-Checks eine hohe Urteilerübereinstimmung aufweisen.

○ - - - ○ - - ○ - ○ + ○ + + ○ + + +

7 * In welche Richtung vermuten Sie, dass ärztliche und studentische Bewertungen voneinander abweichen?

○ Ärzte deutlich strenger
○ Ärzte strenger
○ gleich
○ Studierende strenger
○ Studierende deutlich strenger

8 Wo sehen Sie die Stärken, wo die Schwächen der Praxis-Checks?

9 Möchten Sie uns sonst noch Rückmeldung geben?

Abbildung 51: Abschlussfragebogen der Online-Praxis-Check-Simulation Simulation vom Skills Lab der Universität Ulm. Zeitraum: Sommersemester 20.

7.9 Email – Studienteilnehmende

7.9.1 Email: Anfrage Externe Ärzt*in

Sehr geehrte/r Herr / Frau Max Musterman, liebe/r Vorname (falls man sich kennt)

im Rahmen eines Lehrforschungsprojektes führen wir im Skills Lab der Universität Ulm eine Studie zur Lernzielkontrolle für praktische Fertigkeiten durch. Neben den im Skills Lab beschäftigen studentischen Tutoren benötigen wir hierfür insbesondere ärztliche Studienteilnehmer – daher bitten wir Sie um Ihre Teilnahme. Die Studie findet online auf der Lernplattform Moodle der Universität Ulm statt. Benötigt werden ein Computer oder Laptop (kein Tablet), stabiles Internet und ggf. Kopfhörer. Der Aufwand beträgt rund eine Stunde Bearbeitungszeit am Stück.

Im Skills Lab (Räumlichkeiten mit speziellen Simulatoren und Trainingsgeräten) können Studierende sich auf ihre ersten klinischen Einsätze vorbereiten. Unter Anleitung studentischer Tutoren werden
Untersuchungstechniken und andere praktische Fertigkeiten trainiert. Als Erfolgskontrolle dient der sog. „Praxis-Check“, eine standardisierte Prüfung.

In der Studie geht es um das standardisierte Bewerten so einer Prüfungssituation. Zu diesem Zweck werden von den Studienteilnehmern mehrere gefilmte Prüfungen bewertet. Die Bewertung erfolgt anhand eines definierten Kriterienkataloges, welcher im Rahmen der Studie vorgestellt und besprochen wird.

Bitte schicken Sie uns eine kurze E-Mail, wenn Sie an der Studie teilnehmen wollen. Sie erhalten dann einen externen Zugangscode und werden für die Studie freigeschaltet. Gerne würden wir uns, im Falle einer vollständigen Bearbeitung der Studie innerhalb von zwei Wochen, mit einem Amazon-Gutschein im Wert von 20 Euro für Ihren Zeitaufwand erkenntlich zeigen.
Wir würden uns über eine Teilnahme freuen – bitte geben Sie mir hierzu Rückmeldung.

Viele Grüße
Lukas Keßling, Astrid Horneffer und David Messerer

7.9.2 **Email: Studienteilnehmer*in externe Ärzt*in**

Sehr geehrter Max Musterman, liebe/r Vorname (falls man sich kennt)

vielen Dank, dass Sie an unserer Studie teilnehmen. Ihre Zugangsdaten werden Ihnen demnächst von dem Rechenzentrum der Universität Ulm zugeschickt. Sie finden die Studie unter: https://moodle.uni-ulm.de -> Suchen: Virtuelles Skills Lab -> Prüferschulung

Direktlink: https://moodle.uni-ulm.de/course/view.php?id=8399&section=3

Für die Studie benötigen Sie einen Computer oder Laptop (kein Tablet), eine stabile Internetverbindung und ggf. Kopfhörer sowie eine Stunde Zeit am Stück.

Für eine vollständige Bearbeitung der Studie innerhalb von zwei Wochen bieten wir Ihnen als Aufwandsentschädigung einen 20€ Amazongutschein an. Bitte lassen Sie mir nach Beendigung der Studie Ihre Postadresse zukommen.

Für technische Rückfragen stehe ich Ihnen gerne zur Verfügung.

Mit freundlichen Grüßen
Lukas Keßling, Astrid Horneffer und David Messerer

7.9.3 **Email: Anfrage Studierende**

Sehr geehrte/r Herr / Frau Max Musterman, liebe/r Vorname (falls man sich kennt)

im Rahmen eines Lehrforschungsprojektes führen wir im Skills Lab der Universität Ulm eine Studie zur Lernzielkontrolle für praktische Fertigkeiten durch. Die Studie findet online auf der Lernplattform Moodle der Universität Ulm statt. Benötigt werden ein Computer oder Laptop (kein Tablet), stabiles Internet und ggf. Kopfhörer. Der Aufwand beträgt rund eine Stunde Bearbeitungszeit am Stück.

Im Skills Lab (Räumlichkeiten mit speziellen Simulatoren und Trainingsgeräten) können Studierende sich auf ihre ersten klinischen Einsätze vorbereiten. Unter Anleitung studentischer Tutoren werden Untersuchungstechniken und andere praktische Fertigkeiten trainiert. Als Erfolgskontrolle dient der sog. „Praxis-Check“, eine standardisierte Prüfung.

In der Studie geht es um das standardisierte Bewerten so einer Prüfungssituation. Zu diesem Zweck werden von den Studienteilnehmern mehrere gefilmte Prüfungen bewertet. Die Bewertung erfolgt anhand eines definierten Kriterienkataloges, welcher im Rahmen der Studie vorgestellt und besprochen wird.

Bitte schicken Sie uns eine kurze E-Mail, wenn Sie an der Studie teilnehmen wollen. Gerne würden wir uns, im Falle einer vollständigen Bearbeitung der Studie innerhalb von zwei Wochen, mit einem Amazon-Gutschein im Wert von 10 Euro für Ihren Zeitaufwand erkenntlich zeigen.

Wir würden uns über eine Teilnahme freuen – bitte geben Sie mir hierzu Rückmeldung.

Viele Grüße
Lukas Keßling, Astrid Horneffer und David Messerer

7.9.4 **Email: Studienteilnehmende Studierende**

Sehr geehrter Max Musterman, liebe/r Vorname (falls man sich kennt)

vielen Dank, dass Sie an unserer Studie teilnehmen. Sie wurden für die Studie auf der Lernplattform freischaltet und finden diese unter: https://moodle.uni-ulm.de -> Suchen: Virtuelles Skills Lab -> Prüferschulung

Direktlink: https://moodle.uni-ulm.de/course/view.php?id=8399&section=3

Für die Studie benötigen Sie einen Computer oder Laptop (kein Tablet), eine stabile Internetverbindung und ggf. Kopfhörer sowie eine Stunde Zeit am Stück.

Für eine vollständige Bearbeitung der Studie innerhalb von zwei Wochen bieten wir Ihnen als Aufwandsentschädigung einen 10€ Amazongutschein an. Bitte lassen Sie mir nach Beendigung der Studie Ihre Postadresse zukommen.

Für technische Rückfragen stehe ich Ihnen gerne zur Verfügung.

Mit freundlichen Grüßen
Lukas Keßling, Astrid Horneffer und David Messerer

7.10 Drehbücher Film 1 - 4

A = Arzt

P = Patient

K = Kamera

X = Anweisung

- Fehler für die Filme
    - Film 2 – Fehlerfrei-Variante (Moodle 2)
    - Film 1 – Skillfehler-Variante 9 4 11 (Moodle 1)
    - Film 3 – Kontextfehler-Variante 2 15 17 (Moodle 3)
    - Film 4 – Mischfehler-Variante 1 3 8 10 12 15 16 18 20 (Moodle 4)

Material:

- 2x Kamera
- Stativ
- Armband

Film 2: Fehlerfrei-Variante

1.Szene

K: Schaut vom Fußende auf die Liege. Weitwinkel

X: Kamera (weit) auf Liege. Patient sitzt auf der Liege. Schiebewände sind offen aufgestellt. Material lieg auf kleinem Tisch links neben der Liege bereit. Timer steht rechts neben der Liege. Arzt kommt hinter den Schiebewänden vor.

(Die Tutorin ist nicht zu sehen, damit sie nicht ablenkt. Uhr hinstellen, an der sich die Prüfer*innen orientieren können.)

X: Tutorin liest Anweisung

Anweisung: „Sie sollen beim Probanden das Abdomen vollständig sowie die Niere orientierend untersuchen und Ihre Befunde dokumentieren. Der Proband hat eine unauffällige Anamnese und keine Schmerzen im zu untersuchenden Bereich. Eine weiterführende Anamnese ist nicht erforderlich. Sie haben ab jetzt 10 Minuten Zeit.“

X: Tutorin startet Timer (Tutorin ist dadurch kurz im Bild zu sehen)

A: Guten Tag. X: kein Handschlag / Wände zuziehen

P: Hallo

X: Anfang Händedesinfektion 30 s.

A: Ich bin Herr Hammerschmidt. Ich bin Medizinstudent und würde gerne Ihren Bauch untersuchen. Ist das für Sie ok?

P: nickt

A: Bitte nennen Sie Ihren vollständigen Namen und Geburtsdatum!

P: Maximilian Holz, 23.3.2003

A: Bitte legen Sie sich auf die Liege, öffnen Sie den Gürtel und ziehen Sie das T-Shirt etwas nach oben.

X: P führt Anweisung aus. Legt sich hin.

A: Ich sehe jetzt hier keine Narben. Hatten Sie schonmal eine OP im Bauchbereich?

P. Nein

X: A schaut sich den Bauch an / nimmt Stethoskop in die Hand

A: Ich werde zunächst Ihren Bauch abhören.

X: Auskultation in allen 4 Quadranten für ca. 10 Sekunden pro Quadranten. Im Uhrzeigersinn beginnend links oben. Auskultation Aorta mittig über dem Bauchnabel. Stethoskop auf Tisch legen.

A: Das hört sich alles unauffällig an. Jetzt werde ich Ihren Bauch abklopfen und anschließend abtasten

X: Perkussion in allen 4 Quadranten beginnend li oben Uhrzeigersinn.

A: Jetzt werde ich den Bauch noch abtasten. Ich taste zunächst oberflächlich und Sie sagen mir bitte Bescheid, wenn Sie dabei Schmerzen haben.

X: Palpation aller Quadranten beginnend links oben im Uhrzeigersinn, dabei Blickkontakt, zuerst komplett oberflächlich

A: Jetzt taste ich noch etwas mehr in die Tiefe. Sie sagen mir wieder Bescheid wenn Sie schmerzen haben.

X folgend komplett tief Palpieren

A: können Sie mir sagen, was Sie heute zum Mittag gegessen haben?

X: Prüfung Loslass-Schmerz. Druck linker Unterbauch. Schnelles Loslassen ohne nachdrücken. Tief drücken ca. 5-10s

A: Dann untersuche ich noch Ihre Leber. Bitte atmen Sie vollständig aus. Holen Sie dann tief Luft und halten Sie den Atem an.

X: Palpation der Leber. Kaudal beginnen, um bei Inspiration kaudalen Rand zu tasten. Dabei Anleitung des Patienten zur Ex-/Inspiration

A: Bitte atmen Sie normal weiter. Ich werde jetzt noch die Größe ihrer Leber ermitteln.

X: Kratzauskultation, Stethoskop unter Proc. Xiphoideus. Kratzen von kaudal nach kranial in MCL, Stethoskop auf Tisch legen.

A: Das sind jetzt 11cm, also ein ganz normales Maß. Nun taste ich noch nach Ihrer Milz. Bitte drehen Sie sich dafür etwas auf die rechte Seite.

X: Milz Palpation. Linke Hand umgreift flanke rechte Hand tastet.

A: Bis jetzt ist alles unauffällig. Bitte setzen Sie sich zur Untersuchung der Nieren hin. Haben Sie Schmerzen, wenn ich hier klopfe?

X: Nierenklopfschmerz. Mit Handballen links / rechts neben Wirbelsäule klopfen. Proband sitzt mit Rücken zur Kamera. Beine zeigen zur rechten Seite der Liege. Arzt steht auf linken Seite der Liege am Fußende.

A: Die Untersuchung des Bauches ist nun beendet, ich habe nichts Beunruhigendes gefunden. Bitte ziehen Sie sich wieder an. Ich bespreche die Befunde mit dem Stationsarzt und komme dann auf sie zurück. Vielen Dank.

X: Desinfektion Stethoskop, Händedesinfektion 30 s., (Proband verlässt Zimmer)

X: Tutorin Anweisung: „Welche beiden Untersuchungen gehören darüber hinaus zu einer vollständigen Abdomen Untersuchung?“

A: Ich würde noch eine digital-rektale Untersuchung sowie eine Prüfung möglicher Bruchpforten durchführen.

X: Tutorin „Bitte dokumentieren Sie den Befund“. Arzt verlässt Zimmer

2.Szene:

X: A setzt sich mit Bogen an Tisch

3.Szene

X: fertiger Dokumentationsbogen wird eingeblendet. Handschriftlich sauber ausgefüllt und dann abfotografiert. Stammdaten werden weggelassen.

Film 1: Skillfehler-Variante

Fehler:

9 Palpation Perkussion Auskultation

4 Auskultation nur im oberen Bauchbereich

11 Milzpalpation fehlerhaft

1.Szene

K: Schaut vom Fußende auf die Liege. Weitwinkel

X: Kamera (weit) auf Liege. Patient sitzt auf der Liege. Schiebewände sind offen aufgestellt. Material lieg auf kleinem Tisch links neben der Liege bereit. Timer steht rechts neben der Liege. Arzt kommt hinter den Schiebewänden vor.

(Die Tutorin ist nicht zu sehen, damit sie nicht ablenkt. Uhr hinstellen, an der sich die Prüfer*innen orientieren können.)

X: Tutorin liest Anweisung

Anweisung: „Sie sollen beim Probanden das Abdomen vollständig sowie die Niere orientierend untersuchen und Ihre Befunde dokumentieren. Der Proband hat eine unauffällige Anamnese und keine Schmerzen im zu untersuchenden Bereich. Eine weiterführende Anamnese ist nicht erforderlich. Sie haben ab jetzt 10 Minuten Zeit.“

X: Tutorin startet Timer (ist dadurch kurz im Bild zu sehen)

A: Guten Tag.

X: kein Handschlag / Wände zuziehen

P: Hallo

X: Anfang Händedesinfektion 30s

A: Ich bin Herr Schmitt. Ich bin Medizinstudent und würde gerne Ihren Bauch untersuchen. Ist das für Sie ok?

P: nickt

A: Bitte nennen Sie Ihren vollständigen Namen und Geburtsdatum!

P: Maximilian Holz, 23.3.2003

A: Bitte legen Sie sich auf die Liege, öffnen Sie den Gürtel und ziehen Sie das T-Shirt etwas nach oben.

X: P führt Anweisung aus. Legt sich hin

A Ich sehe jetzt hier keine Narben. Hatten Sie schonmal eine OP im Bauchbereich?

P. Nein

X: A schaut sich den Bauch an

A: Ich werde zunächst Ihren Bauch abklopfen und anschließend abtasten.

X: Perkussion in allen 4 Quadranten beginnend links oben Uhrzeigersinn.

A: Jetzt werde ich den Bauch noch abtasten. Ich taste zunächst oberflächlich und Sie sagen mir bitte Bescheid, wenn Sie dabei Schmerzen haben.

X: Palpation alle Quadranten beginnend links oben im Uhrzeigersinn, dabei Blickkontakt, zuerst komplett oberflächlich

A: Jetzt taste ich noch etwas mehr in die Tiefe. Sie sagen mir wieder Bescheid, wenn Sie Schmerzen haben.

X folgend komplett tief palpieren

A: Können Sie mir sagen was Sie heute zum Mittag gegessen haben? (Ablenkung für Loslass-Schmerz)

X: Prüfung Loslass-Schmerz. Druck linker Unterbauch. Schnelles Loslassen ohne nachdrücken. Tief drücken ca. 5-10s.

A: Jetzt werde ich nur Ihren Bauch abhören.

X: Arzt nimmt sich Stethoskop. Auskultation nur oberen 2 Quadranten für ca. 10 Sekunden im Uhrzeigersinn beginnend links oben. Auskultation Aorta mittig über Bauchnabel. Stethoskop auf Tisch legen.

A: Das hört sich alles unauffällig an. Dann untersuche ich noch Ihre Leber. Bitte atmen Sie vollständig aus. Holen Sie dann tief Luft und halten Sie den Atem an.

X: Palpation der Leber. Kaudal beginnen, um bei Inspiration kaudalen Rand zu tasten. Dabei Anleitung des Patienten zur Ex-/Inspiration

A: Bitte atmen Sie normal weiter. Ich werde jetzt noch die Größe ihrer Leber ermitteln.

X: Kratzauskultation, Stethoskop unter Proc. Xiphoideus. Kratzen von kaudal nach kranial in MCL Stethoskop auf Tisch legen.

A: Das sind jetzt 11cm, also ein ganz normales Maß. Nun taste ich noch nach Ihrer Milz. Bitte drehen Sie sich dafür etwas auf die rechte Seite.

X: Milz Palpation sehr oberflächlich, auf dem Rippenbogen, ohne Widerlager

A: Bis jetzt ist alles unauffällig. Bitte setzen Sie sich zur Untersuchung der Nieren hin. Haben Sie Schmerzen, wenn ich hier klopfe?

X: Nierenklopfschmerz. Mit Handballen links / rechts neben Wirbelsäule klopfen. Proband sitzt mit Rücken zur Kamera. Beine zeigen zur rechten Seite der Liege. Arzt steht auf linken Seite der Liege am Fußende.

A: Die Untersuchung des Bauches ist nun beendet, ich habe nichts Beunruhigendes gefunden. Bitte ziehen Sie sich wieder an. Ich bespreche die Befunde mit dem Stationsarzt und komme dann auf sie zurück. Vielen Dank.

X: Desinfektion Stethoskop, Händedesinfektion 30 s., Patient verlässt Zimmer

X: Prüfer Anweisung: „Welche beiden Untersuchungen gehören darüber hinaus zu einer vollständigen Abdomen Untersuchung?“

A: Ich würde noch eine digital-rektale Untersuchung sowie eine Prüfung möglicher Bruchpforten durchführen.

2.Szene:

X: A setzt sich mit Bogen an Tisch

3.Szene

X: fertiger Dokumentationsbogen wird eingeblendet. Handschriftlich sauber ausgefüllt und dann abfotografiert. Dokumentation: lebhafte Darmgeräusche in den oberen beiden Quadranten, untere Quadranten nicht untersucht.

Film 3: Kontextfehler-Variante

Fehler:

2 keine Ankündigung der Untersuchung

15 hygienische Händedesinfektion 15s

17 Weglassen der Stammdaten

1.Szene

K: Schaut vom Fußende auf die Liege. Weitwinkel

X: Kamera (weit) auf Liege. Patient sitzt auf der Liege. Schiebewände sind offen aufgestellt. Material lieg auf kleinem Tisch links neben der Liege bereit. Timer steht rechts neben der Liege. Arzt kommt hinter den Schiebewänden vor.

(Die Tutorin ist nicht zu sehen, damit sie nicht ablenkt. Uhr hinstellen, an der sich die Prüfer*innen orientieren können.)

X: Tutorin liest Anweisung

Anweisung: „Sie sollen beim Probanden das Abdomen vollständig sowie die Niere orientierend untersuchen und Ihre Befunde dokumentieren. Der Proband hat eine unauffällige Anamnese und keine Schmerzen im zu untersuchenden Bereich. Eine weiterführende Anamnese ist nicht erforderlich. Sie haben ab jetzt 10 Minuten Zeit.“

X: Tutorin startet Timer (ist dadurch kurz im Bild zu sehen)

A: Guten Tag. Ich bin Herr Schmitt und studiere Medizin

X: kein Handschlag / Wände zuziehen / Anfang Händedesinfektion 30 s.

P: Hallo

A: Bitte nennen Sie Ihren vollständigen Namen und Geburtsdatum!

P: Maximilian Holz, 23.3.2003

A: Bitte legen Sie sich auf die Liege, öffnen Sie den Gürtel und ziehen Sie das T-Shirt etwas nach oben.

X: P führt Anweisung aus. Legt sich hin.

A: Ich sehe jetzt hier keine Narben. Hatten Sie schonmal eine OP im Bauchbereich?

P. Nein

X: A schaut sich den Bauch an / nimmt Stethoskop in die Hand

A: Ich werde zunächst Ihren Bauch abhören.

X: Auskultation in allen 4 Quadranten für ca. 10 Sekunden pro Quadranten im Uhrzeigersinn beginnend links oben. Auskultation Aorta mittig über dem Bauchnabel. Stethoskop auf Tisch legen.

A: Das hört sich alles unauffällig an. Jetzt werde ich Ihren Bauch abklopfen und anschließend abtasten

X: Perkussion in allen 4 Quadranten beginnend links oben Uhrzeigersinn.

A: Jetzt werde ich den Bauch noch abtasten. Ich taste zunächst oberflächlich und Sie sagen mir bitte Bescheid, wenn Sie dabei Schmerzen haben.

X: Palpation alle Quadranten beginnend links oben im Uhrzeigersinn, dabei Blickkontakt,zuerst komplett oberflächlich

A: Jetzt taste ich noch etwas mehr in die Tiefe. Sie sagen mir wieder Bescheid, wenn Sie Schmerzen haben.

X folgend komplett tief palpieren

A: können Sie mir sagen was Sie heute zum Mittag gegessen haben?

X: Prüfung Loslass-Schmerz. Druck linker Unterbauch. Schnelles loslassen ohne nachdrücken. Tief drücken ca. 5-10s

A: Dann untersuche ich noch Ihre Leber. Bitte atmen Sie vollständig aus. Holen Sie dann tief Luft und halten Sie den Atem an.

X: Palpation der Leber. Kaudal beginnen, um bei Inspiration kaudalen Rand zu tasten. dabei Anleitung des Patienten zur Ex-/Inspiration

A: Bitte atmen Sie normal weiter. Ich werde jetzt noch die Größe ihrer Leber ermitteln.

X: Kratzauskultation, Stethoskop unter Proc. Xiphoideus. Kratzen von kaudal nach kranial in MCL. Stethoskop auf Tisch legen.

A: Das sind jetzt 11cm, also ein ganz normales Maß. Nun taste ich noch nach Ihrer Milz. Bitte drehen Sie sich dafür etwas auf die rechte Seite.

X: Milz Palpation. Linke Hand umgreift flanke rechte Hand tastet.

A: Bis jetzt ist alles unauffällig. Bitte setzen Sie sich zur Untersuchung der Nieren hin. Haben Sie Schmerzen, wenn ich hier klopfe?

X: Nierenklopfschmerz. Mit Handballen links / rechts neben Wirbelsäule klopfen. Proband sitzt mit Rücken zur Kamera. Beine zeigen zur rechten Seite der Liege. Arzt steht auf linken Seite der Liege am Fußende.

A: Die Untersuchung des Bauches ist nun beendet, ich habe nichts Beunruhigendes gefunden. Bitte ziehen Sie sich wieder an. Ich bespreche die Befunde mit dem Stationsarzt und komme dann auf sie zurück. Vielen Dank.

X: Desinfektion Stethoskop, Händedesinfektion 15 s. Patient verlässt das Zimmer.

X: Prüfer Anweisung: „Welche beiden Untersuchungen gehören darüber hinaus zu einer vollständigen Abdomen Untersuchung?“

A: Ich würde noch eine digital-rektale Untersuchung sowie eine Prüfung möglicher Bruchpforten durchführen.

X: Prüfer „Bitte dokumentieren Sie den Befund“. Arzt verlässt Zimmer.

2.Szene:

X: A setzt sich mit Bogen an Tisch

3.Szene

X: Fertiger Dokumentationsbogen wird eingeblendet. Handschriftlich sauber ausgefüllt und dann abfotografiert. Dokumentation der Perkussion wird weggelassen

Film 4: Mischfehler-Variante

Fehler:

1 Ansprechen der Person mit einer rhetorischen Frage

3 Armband an der linken Hand

8 Prüfung Loslass-Schmerz: zu kurzes Eindrücken

10 Kratzauskultation nur bis untere Lebergrenze

12 Nierenklopfschmerz auf Beckenschaufel

15 Händehygiene weglassen

16 Bauch Sono und DRU

18 Perkussionsbefund weglassen

20 Dokumentation Leber u Milz weglassen (Zeitmangel)

1.Szene

K: Schaut vom Fußende auf die Liege. Weitwinkel

X: Kamera (weit) auf Liege. Patient sitzt auf der Liege. Schiebewände sind offen aufgestellt. Material lieg auf kleinem Tisch links neben der Liege bereit. Timer steht rechts neben der Liege. Arzt kommt hinter den Schiebewänden vor.

(Die Tutorin ist nicht zu sehen, damit sie nicht ablenkt. Uhr hinstellen, an der sich die Prüfer*innen orientieren können.)

X: Tutorin liest Anweisung

Anweisung: „Sie sollen beim Probanden das Abdomen vollständig sowie die Niere orientierend untersuchen und Ihre Befunde dokumentieren. Der Proband hat eine unauffällige Anamnese und keine Schmerzen im zu untersuchenden Bereich. Eine weiterführende Anamnese ist nicht erforderlich. Sie haben ab jetzt 10 Minuten Zeit.“

X: Tutorin startet Timer (ist dadurch kurz im Bild zu sehen)

A: Guten Tag. Sie sind Herr Holz, richtig?

X: Wände zuziehen / kein Handschlag

P: Guten Tag.

X: nickt deutlich

A: Ich bin Medizinstudent und werde jetzt Ihren Bauch untersuchen. Ist das für Sie ok?

P: nickt

X: Anfang Händedesinfektion 15 s

A: Bitte sagen Sie mir ihr Geburtsdatum!

P: 23.3.2003

A: Bitte legen Sie sich auf die Liege, öffnen Sie den Gürtel und ziehen Sie das T-Shirt etwas nach oben.

X: P führt Anweisung aus. Legt sich hin

A: Hatten Sie bereits eine Operation im Bauchbereich?

P. Nein

X: A schaut sich den Bauch an / nimmt Stethoskop in die Hand

A: Ich werde zunächst Ihren Bauch abhören.

X: Auskultation in allen 4 Quadranten für ca. 10 Sekunden pro Quadranten im Uhrzeigersinn beginnend links oben Auskultation Aorta mittig über dem Bauchnabel. Stethoskop auf Tisch legen.

A: Das hört sich alles unauffällig an. Jetzt werde ich Ihren Bauch abklopfen und anschließend abtasten

X: Perkussion in allen 4 Quadranten beginnend links oben im Uhrzeigersinn.

A: Jetzt werde ich den Bauch noch abtasten. Ich taste zunächst oberflächlich und Sie sagen mir bitte Bescheid, wenn Sie dabei Schmerzen haben.

X: Palpation alle Quadranten beginnend links oben im Uhrzeigersinn, dabei Blickkontakt zuerst komplett oberflächlich

A: Jetzt taste ich noch etwas mehr in die Tiefe. Sie sagen mir wieder Bescheid, wenn Sie Schmerzen haben.

X folgend komplett tief palpieren

A: können Sie mir sagen was Sie heute zum Mittag gegessen haben?

X: Prüfung Loslass-Schmerz. Druck in rechten Unterbauch. Schnelles Loslassen ohne nachdrücken.

A: Dann untersuche ich noch Ihre Leber. Bitte atmen Sie vollständig aus. Holen Sie dann tief Luft und halten Sie den Atem an.

X: Palpation der Leber. Kaudal beginnen, um bei Inspiration kaudalen Rand zu tasten. Dabei Anleitung des Patienten zur Ex-/Inspiration

A: Bitte atmen Sie normal weiter.

X: Kratzauskultation von distal kommend nur bis zur unteren Leberkante. Obere Leberkante wird geschätzt. Stethoskop unter Proc. Xiphoideus. Kratzen von kaudal nach kranial in MCL Stethoskop auf Tisch legen

A: Nun taste ich noch nach Ihrer Milz. Bitte drehen Sie sich dafür etwas auf die rechte Seite.

X: Milz Palpation, Linke Hand umgreift flanke rechte Hand tastet.

A: Bis jetzt ist alles unauffällig. Bitte setzen Sie sich zur Untersuchung der Nieren hin. Haben Sie Schmerzen, wenn ich hier klopfe?

X: Nierenklopfschmerz. Mit Handballen links / rechts neben Wirbelsäule klopfen Höhe der Beckenschaufel. Proband sitzt mit Rücken zur Kamera. Beine zeigen zur rechten Seite der Liege. Arzt steht auf linken Seite der Liege am Fußende.

A: Die Untersuchung des Bauches ist nun beendet, ich habe nichts Beunruhigendes gefunden. Bitte ziehen Sie sich wieder an. Ich bespreche die Befunde mit dem Stationsarzt und komme dann auf sie zurück. Vielen Dank.

X: Desinfektion Stethoskop, Händedesinfektion 30 s. Patient verlässt das Zimmer.

X: Prüfer Anweisung: „Welche beiden Untersuchungen gehören darüber hinaus zu einer vollständigen Abdomen Untersuchung?“

A: Ich würde noch eine digital-rektale Untersuchung sowie ein Bauchsonographie durchführen.

X: Prüfer „Bitte dokumentieren Sie den Befund“, Arzt verlässt Zimmer

2.Szene:

X: A setzt sich mit Bogen an Tisch

3.Szene

X: A sitzt am Tisch, schreibt. Nach 10s klingelt Timer.

X: Tutorin „Die Zeit ist abgelaufen, bitte geben Sie ihr Blatt ab.“ A gibt Blatt ab.

X: Fertiger Dokumentationsbogen wird eingeblendet. Handschriftlich sauber ausgefüllt und dann abfotografiert. Dokumentation des Perkussionsbefundes / der Leber / der Milz wird weggelassen.

8 Danksagung

Ich danke Herrn Prof. Dr. Markus Huber-Lang für die freundliche Überlassung des Themas der Dissertation.

Besonders möchte ich mich bei Dr. David Messerer für die exzellente Betreuung und Unterstützung meiner Arbeit bedanken.

Des Weiteren möchte ich mich bei Astrid Horneffer und den an den Filmen beteiligten Schauspieler*innen für ihre tatkräftige Unterstützung danken.

Auch möchte ich mich bei meiner Familie und bei meinen Freunden bedanken die mich immer unterstützt und motiviert haben.

9 Lebenslauf

Name:	Lukas Keßling
Geburtsjahr:	1996
Geburtsort:	München
Staatsangehörigkeit:	deutsch

Ausbildung:

07.2002 - 07.2006	Volksschule Haimhausen, Bayern
09.2006 - 06.2014	Gisela-Gymnasium, München: Allgemeine Hochschulreife
10.2015 - 12.2022	Universität Ulm Medizinische Fakultät (2015-2021), Technische Universität München (2021-2022)

München, Mai 2021

Lukas Keßling