



DIE WEITERBILDUNGSUNIVERSITÄT DER TU DRESDEN

in Kooperation mit der



OsteopathieSchule
Deutschland

Konsenstraining – Verbesserung der Interrater- Reliabilität bei der Testung nach Lien Mécanique Ostéopathique

**Abschlussarbeit zur Erlangung des Titels
Bachelor of Science**

Name: Marie-Annick Schmid

Klasse: Jahrgang 14

Matrikelnummer: 7006479

Geburtsdatum: 18.08.1989

1. Prüfer: Florian Dreyer

2. Prüfer: Tobias Dobler

Berlin, am 29.05.2018

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Berlin, am 29.05.2018

Danksagung

Mein Dank gilt zuallererst dem Team, das diese Studie gemeinsam mit mir organisiert und durchgeführt hat: Lisa Ebach und Anika Schulze. Weiters gilt mein Dank unseren beiden Testenden Juliette Bersou und Sebastian Spielhofer. Ebenfalls, allerdings (selbsterklärend) nicht namentlich, möchte ich mich bei den Studienteilnehmern und Studienteilnehmerinnen ganz herzlich bedanken, ohne sie wäre die Studie nicht realisierbar gewesen. Auch der Lektorin Burgi Schmid, dem Lektor Thomas Schweitzer und den Übersetzern ins Englische Peter Kaplar und Daniel Rosenfels möchte ich hier einen gesonderten Dank sagen.

Zusammenfassung

Hintergrund: Studierende der Osteopathie werden vor die Herausforderung gestellt, ihren Palpationssinn zu einem zuverlässigen Testinstrument zu entwickeln. Die Autorin nimmt für diese Studie an, dass durch mehrere Testende durchgeführte manuelle Testungen an einer Person, zu den gleichen Ergebnissen führen können. Diese Arbeit untersucht, ob ein Konsenstraining der Studierenden zu zuverlässigeren Ergebnissen im Bereich von palpatorischen Testungen, die keinen Sichtbefund ergeben, führen kann.

Material und Methoden: Anhand von ausgewählten Tests der unteren Extremität aus dem Konzept der Lien Mécanique Ostéopathique wurde ein Konsenstraining durchgeführt. Vor und nach dem Konsenstraining fanden Messungen statt, bei welchen fünf Testende jeweils 18 Probanden und Probandinnen testeten. Die Ergebnisse wurden durch die Berechnung der Kappa-Werte verglichen.

Ergebnisse: Die Ergebnisse von Prämessung ($p= 0,246-0,955$) und Postmessung ($p=0,11-0,953$) sind nicht signifikant und erreichen den Grenzwert ($\kappa=0,4$) für reliable Ergebnisse nicht (Prämessung: $\kappa= -0,11-0,04$; Postmessung: von $\kappa = -0,09-0,07$).

Schlussfolgerung: Es konnte keine ausreichende Verbesserung durch Konsenstraining erzielt werden. Die Ergebnisse sind weder in der Prä- noch in der Postmessung zuverlässig. Unter dem angewandten Studiendesign (Wahl von Tests, Probanden und Probandinnen, Statistikverfahren) ließ sich durch Konsenstraining bei dem vorliegenden Ausbildungsstand keine Ergebnisverbesserung erzielen.

Schlüsselwörter: Lien Mécanique Ostéopathique, Mechanical Link, Konsenstraining, Interrater-Reliabilität, Kappa, Osteopathie

Abstract

Background: Students of Osteopathy are given the challenge of developing their palpatory skills into a reliable test instrument. The author assumes for the purposes of this study that tests made on the same person by different testers should lead to the same outcome. This study examines if consensus training with the students can lead to more reliable results in the realm of testing in which no visual findings are evident.

Subjects and methods: On the basis of tests selected from the concept of Lien Mécanique Ostéopathique, a consensus training was conducted. Measurements were made on the lower extremities of 18 subjects by each of 5 testers before and after the consensus training. The measurements were then compared using a computation of the Kappa values.

Results: The results of the pre-measurement ($p=0.246-0.955$) and the post-measurement ($p=0.11-0.953$) are not significant, and do not reach either in the pre-measurement (from $\kappa = -0.11-0.04$) or in the post-measurement ($\kappa = -0.09-0.07$) the limiting value ($\kappa = 0.4$) for reliable results.

Conclusion: The consensus training was not able to achieve a sufficient improvement. The results of both the pre- and post-measurements were not reliable. Using the methods applied (selection of tests, selection of subjects, statistical procedure), the consensus training was not, at the current level of training, able to achieve a statistically significant improvement.

Keywords: Lien Mécanique Ostéopathique, Mechanical Link, consensus training, interrater reliability, Kappa, Osteopathy

I INHALTSVERZEICHNIS

| | |
|---|----|
| II Abkürzungsverzeichnis..... | 3 |
| III Abbildungsverzeichnis..... | 4 |
| IV Tabellenverzeichnis..... | 5 |
| 1. Einleitender Teil..... | 1 |
| 2. Hintergrundteil..... | 2 |
| Das Konzept der LMO..... | 2 |
| Studienlage LMO - Testungen..... | 4 |
| Konsenstraining..... | 5 |
| Interrater-Reliabilität..... | 8 |
| Fehlerquellen LMO-Interrater-Reliabilitäts-Studien..... | 9 |
| Fragestellung und Hypothesen..... | 11 |
| 3. Methodikteil..... | 12 |
| Variablen..... | 12 |
| Rekrutierung, Stichprobe, Ein- und Ausschlusskriterien..... | 12 |
| Voraussetzungen für die Testenden..... | 13 |
| Information und Datenschutz..... | 13 |
| Erhebungs-Ablauf..... | 14 |
| Tests der unteren Extremität..... | 15 |
| Ausführung und Bewertung der Tests..... | 15 |
| Test-Protokoll..... | 17 |
| Konsenstraining..... | 18 |
| Statistik..... | 20 |
| 4. Ergebnisse..... | 22 |
| Probanden und Probandinnen Verteilung und Dropout..... | 22 |
| Datenauswertung..... | 24 |
| Herausgearbeitete Inhalte des Konsenstrainings..... | 26 |
| 5. Diskussion..... | 28 |
| Methoden-Diskussion..... | 28 |
| 1 Rekrutierung..... | 28 |
| 2 Studiendesign..... | 28 |
| 3 Gewebe-Veränderungen durch mehrfache Testungen..... | 28 |

| | |
|---|----|
| 4 Ausführung und Bewertung der Tests..... | 30 |
| 5 Konsenstraining..... | 30 |
| Ergebnis-Diskussion..... | 31 |
| 1 Literatur im Hintergrundteil..... | 31 |
| 2 Signifikanz der Ergebnisse..... | 31 |
| 3 Kappa-Werte und ihre statistische Auswertung..... | 31 |
| Beantwortung der Fragestellung..... | 33 |
| Ausblick..... | 33 |
| Schlussbetrachtung..... | 33 |
| 6. Literaturverzeichnis..... | 34 |
| 7. Anhang..... | 37 |
| Information..... | 37 |
| Einverständniserklärung..... | 38 |
| Code..... | 39 |
| Testprotokoll..... | 40 |
| Holsti-Koeffizient..... | 41 |

II ABKÜRZUNGSVERZEICHNIS

| | |
|--|--|
| Lien Mécanique Ostéopatique (LMO)..... | |
| Mechanical Link (ML)..... | |
| siehe Seite (s.S.)..... | |
| zum Beispiel (zB)..... | |
| beidseits (bds.)..... | |
| vergleiche (vgl.)..... | |
| et cetera (etc.)..... | |
| Probanden und Probandinnen (PP)..... | |
| Wirbelkörper (WK)..... | |
| beziehungsweise (bzw.)..... | |
| Wiederholung (Wh)..... | |

III ABBILDUNGSVERZEICHNIS

| | |
|---|----|
| Abbildung 1: Flussdiagramm Versuchsaufbau..... | 22 |
| Abbildung 2: Geschlechtsverteilung der Probanden und Probandinnen bei Prä- und Posttestung..... | 23 |
| Abbildung 3: Prätestung - Alter..... | 24 |
| Abbildung 4: Posttestung - Alter..... | 24 |
| Abbildung 5: Oberschenkel. Prätest $K=-0,02$ Posttest $K=0,03$ | 26 |
| Abbildung 6: Hüfte. Prätest $K=-0,03$ Posttest $K=0,07$ | 26 |
| Abbildung 7: OSG. Prätest $K=-0,11$ Posttest $K=0,1$ | 26 |
| Abbildung 8: Patella. Prätest $K=-0,05$ Posttest $K=0,00$ | 26 |
| Abbildung 9: US. Prätest $0,04$ Posttest $0,01$ | 26 |
| Abbildung 10: Knie. Prätest $K=-0,02$ Posttest $K=-0,09$ | 26 |
| Abbildung 11: Fuß. Prätest $K=-0,02$ Posttest $K=-0,04$ | 26 |
| Abbildung 12: Gesamtergebnis aller Testungen. Prätest $K=-0,01$ Posttest $K=0,04$... | 27 |

IV TABELLENVERZEICHNIS

| | |
|---|----|
| Tabelle 1: exemplarischer Ausschnitt aus dem Test-Protokoll..... | 17 |
| Tabelle 2: Fleiss' Kappa Ergebnisse für Prä- und Post-Testung..... | 25 |
| Tabelle 3: Paarweise mittlere Übereinstimmung der Prä- und Posttestung..... | 27 |

1. EINLEITENDER TEIL

Ein Ziel der Forschung im Bereich der Osteopathie ist das Überprüfen der Vergleichbarkeit von Testergebnissen zwischen Therapeuten und Therapeutinnen. Dies ist ein Grundbaustein der Forschung, der auch zur osteopathischen Diagnostik und Weiterentwicklung der Lehre beiträgt. Andere Bereiche, die im Bereich der Praxis zu immer besserem Verständnis der Patienten und Patientinnen beitragen können, sind etwa die Verbesserung einer differenzierten Anamnese, Inspektion, Überprüfung der Aussagekraft von Tests und das Verständnis über die Zusammenhänge dieser Teilbereiche, um zu einer ersten Arbeitshypothese, einer Grundlage zur Behandlung zu finden.

Diese Studie befasst sich mit der Vergleichbarkeit von Testergebnissen bei Testung durch verschiedene Personen. Dies bezeichnet man als reliabel. Reliable Tests sind in der Osteopathie-Ausbildung eine gute Grundlage für das Erlernen, die Verbalisierung des Erlernen und das Gewinnen von Sicherheit und Vertrauen in die eigenen Fähigkeiten. Aus der Erfahrung der ersten Studienjahre zeigt sich, dass je geringer das Bewegungsausmaß bei einer Testung, desto differenzierter müssen die palpatorischen Fähigkeiten sein. Es muss sich nach und nach ein palpatorischer Erfahrungsschatz aufbauen, auf den man sich bei der Beurteilung eines durchgeführten Tests bezieht. Der Bereich der Testungen, in denen der Sichtbefund keine Auskunft gibt, stellt unsere Fähigkeiten – in unserer sehr visuell geprägten Welt – vor große Herausforderungen.

Im speziellen wird sich diese Arbeit damit auseinandersetzen, ob es bei Studierenden, die sich mindestens im vierten Ausbildungsjahr befinden, zu reliablen Ergebnissen bei einer Testform kommt, bei denen eine Ergebnissicherung mittels Sichtbefund nicht möglich ist.

Dafür wird auf Elastizitätstests aus dem osteopathischen Konzept der Lien Mécanique Ostéopatique (LMO) – Synonym: Mechanical Link (ML) zurückgegriffen. Für dieses Konzept gibt es bereits Studien, die die Reliabilität der Testungen bei erfahrenen Osteopathen und Osteopathinnen bestätigen.

Das Konsenstraining wird als zentraler Bestandteil dieser Arbeit durchgeführt, um die Verbesserungsmöglichkeiten für die Übereinstimmung von Testergebnissen zwischen Studenten und Studentinnen im Bereich sehr feiner Palpation klinisch zu erheben.

In der Einleitung wurde bewusst auf das Zitieren verzichtet, um einen flüssigen

Einstieg in das Themengebiet zu ermöglichen. Alle Punkte und Quellen werden im Hintergrundteil (siehe Seite (s.S.) 2-11) aufgegriffen und eingehend erläutert.

2. HINTERGRUNDTEIL

Im Hintergrundteil werden Das Konzept der LMO, Studienlage LMO - Testungen, Konsenstraining, Interrater-Reliabilität, Fehlerquellen LMO-Interrater-Reliabilitäts-Studien und Fragestellung und Hypothesen vorgestellt.

Das Konzept der LMO

Wie (Chauffour & Prat, 2010) schreibt, hat sich die LMO seit Anfang der 1970er Jahre innerhalb der Osteopathie, in erster Linie durch Paul Chauffour und Eric Prat, entwickelt. Die LMO hat eine Testserie über acht festgelegte Test-Einheiten, die sich nach (Hafen-Bardella, 2018) in folgende Einheiten aufteilen:

1. Okzipito-vertebro-pelvine Achse,
2. anteriorer Thorax,
3. Extremitäten: osteoartikulär,
4. Extremitäten: Intraossäre Kraftlinien und Diastasen,
5. Kraniaum,
6. Viszera,
7. Vaskuläres System und Sympathikus,
8. periphere Nerven und Hirnnerven.

Die Tests aller acht Einheiten werden zu Beginn jeder Behandlung ausgeführt. Die Summe aller gefundenen positiven Tests wird als **totale Läsion** bezeichnet. Bevor sich die LMO entwickelte, prägten nach (Fossum, 2010a) den Begriff der totalen Läsion schon Becker und Fryette. Dort bezeichnet er die Summe aller Dysfunktionen im Körper und physiologischen und psychischen Stress.

Um den Ausgangspunkt für die Behandlung festzulegen werden **Balancetests** angewendet. Das bedeutet: es werden zwei Tests zeitgleich ausgeführt. Es wird immer nur ein Test positiv gewertet, obwohl vorhin beide positiv waren – es bleibt also die stärker fixierte Läsion. Dies ist das Verfahren, wie eine Entscheidung getroffen wird, welche der Läsionen die **Primäre Läsion** ist. Sie gilt dann als der Startpunkt der Behandlung. Es gibt zwei grundlegende Arten von Tests. Der

Globaltest ist ein Test, der eine bestimmte Körperregion, zum Beispiel (zB) ein Gelenk testet. Der Globaltest besteht aus zwei oder mehr der folgenden Elemente/Testrichtungen: Kompression (Druck), Traktion (Zug), Flexion (Biegung) und Torsion (Verdrehung).

Die Testrichtungen werden nacheinander eingestellt, ohne dass die Spannung der vorangegangenen Testrichtungen aufgegeben wird. So erhöht sich die Spannung auf das Gewebe. Ist eine der Testrichtungen restriktiver als erwartet (diese Restriktionen werden auch als „fester“ oder „eingeschränkter“ als gesundes Gewebe bezeichnet) gilt der Test als positiv. Als global wird er deswegen bezeichnet, weil die einzelnen Faktoren noch spezifischer untersucht werden können. Globaltests werden häufig beidseits (bds.) zeitgleich ausgeführt, wie zB bei dem Test für das Kniegelenk, bei dem die Testenden zeitgleich eine Hand am linken Knie und die andere am rechten Knie positionieren, um den Test auszuführen. Dadurch kommt es auch zu einem Balancetest (siehe weiter unten). So kann der Test: 1) negativ. 2) positiv links oder 3) positiv rechts sein.

Die **spezifischen Tests** testen eine Struktur in einer spezifischen Elastizitätsrichtung zB eine Elastizitätsrichtung eines Knochens. Die möglichen Testrichtungen/Elemente entsprechen denen der Globaltests: Traktion, Kompression, 2 Torsionen, 4 Biegungen.

(Chauffour & Prat, 2002) beschreiben diese Grundlagen und Testserien in Lehrbüchern.

In der Zeitschrift Osteopathische Medizin erschien ein Artikel von (Hafen-Bardella, 2018) zu den Grundlagen der LMO.

Die im Rahmen dieser Studie durchgeführten Tests entstammen aus der Einheit 3 Extremitäten: osteoartikulär. Daraus werden die sieben Tests der unteren Extremität durchgeführt. Die genaue Beschreibung findet sich auf Seite 15 im Methodikteil

Auf den Ablauf von Behandlung und Retest im Konzept der LMO wird hier nicht eingegangen, da es keine Relevanz für die Studie hat. Dies ist ebenfalls bei (Chauffour & Prat, 2002) und (Hafen-Bardella, 2018) nachzulesen.

Studienlage LMO - Testungen

Die wissenschaftliche Studienlage in diesem Bereich der Osteopathie zeigt keine publizierten Artikel. Dies zeigte die Suche mit pubmed. Zusätzlich wurde im osteopathic research web gesucht. Diese Suchmaschine umfasst auch Master-Thesen.

Ausgeschlossen werden Arbeiten, die

- a) sich nicht mit dem oben genannten Konzept auseinandersetzen,
- b) sich mit der Behandlung nach dem Konzept auseinandersetzen
- c) keine Interrater-Reliabilitäts-Messung (Messung der Reliabilität zwischen mehreren Messenden) durchführen

Eingeschlossen werden Arbeiten, die sich mit der Reliabilität von Testungen nach dem LMO-Konzept auseinandersetzen.

pubmed (<https://www.ncbi.nlm.nih.gov/pubmed/>):

- mechanical link osteopath* → 6 Ergebnisse. Ausschluss a) → 0 Ergebnisse.
- lien mécanique osteopathique osteopath* → 0 Ergebnisse.

osteopathic research web (<http://www.osteopathicresearch.com/>):

- mechanical link → 1 Ergebnis. Ausschluss b) → 0 Ergebnisse.
- lmo → 2 Ergebnisse. Kein Ausschluss → 2 Ergebnisse.

Aus anderen Quellen ließen sich noch zwei weitere Master-Arbeiten zur Reliabilität im Bereich der LMO finden.

Im Folgenden stelle ich die Ergebnisse der Arbeiten vor, die für diese Studie von Bedeutung sind. Andere Teilbereiche der Studien, die mit der Interraterreliabilität nichts zu tun haben, werden außer Acht gelassen.

(Hafen, 2009) untersuchte die Reliabilität der 1. Einheit: okzipito-vertebro-pelvine Achse nach LMO. Hier wurden gute Interraterübereinstimmungen erreicht. Errechnet wurden sie mit dem Holsti-Koeffizienten. In einer Follow-up-Studie hat (Kühn, 2018) gezeigt, dass die positiven Ergebnisse wiederholbar sind. In dieser Studie wurde die Übereinstimmungen innerhalb aller Tests, innerhalb des Tests mit dem höchsten und dem niedrigsten Übereinstimmungswert berechnet. Die statistische Auswertung geschah mittels Holsti-Koeffizienten, Fleiss' Kappa, sowie Krippendorff's Alpha. Somit wird eine Vergleichbarkeit zwischen den verschiedenen Statistik-Varianten und auch

zwischen verschiedenen Arbeiten erreicht.

Ein weiteres Beispiel ist die Interrater-Reliabilitäts-Messung der craniellen Testungen im Rahmen dieses Konzeptes von (Richter-Schultz, 2010) mit weitaus schlechteren Ergebnissen. Diese Studie führte die statistischen Berechnungen mit Fleiss' Kappa durch und wurde von Therapeutinnen und Therapeuten durchgeführt, die im Praxisalltag nicht mit LMO arbeiten.

(Hafen, 2018) zeigt gute Fleiss' Kappa Werte für die Interraterreliabilität bei den Testungen der Extremitäten. Die Autorin verwendet für die Berechnung den Holsti-Koeffizienten, berechnet aber zur besseren Vergleichbarkeit ebenfalls Fleiss' Kappa und Krippendorfs Alpha. Die von dem Team untersuchten Tests haben eine große Überschneidung mit den Tests, die in vorliegender Studie untersucht werden.

Konsenstraining

Die Studienlage in dem Bereich des Konsenstrainings zeigt kaum publizierte Artikel. Dies zeigte die Suche mit pubmed. Zusätzlich wird eine Bachelorarbeit zitiert.

Ausgeschlossen werden Arbeiten, die sich nicht mit Palpation und manuellen Tests auseinandersetzen.

Eingeschlossen werden Arbeiten, die sich mit Palpation und manuellen Tests auseinandersetzen.

pubmed:

- consensus training interobserver reliability osteopath* → 2 Ergebnisse

andere Quellen:

- 1 Ergebnis

Es wird angenommen, dass durch ein Konsenstraining die Interrater-Reliabilität erhöht werden kann. Innerhalb dieses Prozesses soll durch gegenseitige Absprache, Suche nach Fehlerquellen und vergleichende Testung ein besserer Übereinstimmungsgrad erhalten werden.

(Degenhardt, Snider, Snider, & Johnson, 2005), (Degenhardt, Snider, Snider, & Johnson, 2010) und (Jakobs, 2017) können durch Konsenstraining eine Verbesserung der Reliabilität aufzeigen. Sie wählen jedoch unterschiedliche Designs bei der Ausführung des Konsenstrainings.

(Degenhardt et al., 2005) wählt vier von acht Tests der Prämessung, welche drei Therapeuten ausführen. In der ersten Messung zeigten alle acht Messungen eine schlechte Interraterreliabilität, die hier als $\kappa < 0,4$ definiert werden (vergleiche (vgl.) S.20). Die vier Tests mit den besten κ -Werten wurden für das Konsenstraining ausgewählt, da es mit größerer Wahrscheinlichkeit zur Verbesserung durch ein Konsenstraining kommen kann. Die vier durchgeführten Tests der Lumbalregion sind: *positional asymmetry* (positionelle Asymmetrie), *tissue texture abnormalities* (abnormale Gewebebeschaffenheit), *tenderness* (Druckschmerz), *anterior-posterior translational springing* (Spring-Test). Die Therapeuten und Therapeutinnen treffen sich über vier Monate ein Mal pro Woche für ein bis zwei Stunden und setzen sich mit einem Test nach dem anderen auseinander und besprechen die Unterschiede. Erst wenn es zu Verbesserungen in der Reliabilität gekommen ist, wird zum nächsten Test übergegangen. Zu guten Ergebnissen, hier festgelegt mit $\kappa \geq 0,4$ in der Postmessung kommen zwei der vier Tests (*tissue texture abnormalities* κ 0,45 & *tenderness* κ 0,68). Die zwei anderen weisen Verbesserungen auf, können aber weiterhin keine gute Reliabilität ($\kappa \geq 0,4$) aufweisen (*positional asymmetries* κ 0,34 & *anterior-posterior translational springing* ab κ 0,2). Keine signifikante Verbesserung konnte bei *anterior-posterior translational springing* festgestellt werden.

In einer Folgestudie beleuchtet (Degenhardt et al., 2010) die Aufrechterhaltung guter Interraterergebnisse über einen Zeitraum von vier Monaten. An sechs Terminen führen zwei Therapeuten vier Tests aus. Die vier Tests entsprechen den Tests der vorangegangenen Studie. In diesem Studiendesign beschäftigen sich die zwei Therapeuten nicht mehr mit Definition, Ausführung und Interpretation im Allgemeinen, sondern nur mit nicht übereinstimmenden Testergebnissen, indem sie versuchen nachzuvollziehen, wie der andere Therapeut den Test macht, wo die Unterschiede liegen und wie man diese Unterschiede angleichen kann. Die erreichte Übereinstimmung kann über die Termine weitgehend aufrecht erhalten werden, liegt jedoch nie im Bereich einer guten Reliabilität. Es wird angemerkt, dass jede Verbesserung und Aufrechterhaltung von Übereinstimmungen durch Konsenstraining erst dann Sinn macht, wenn die Ergebnisse eine gute Reliabilität aufweisen.

(Jakobs, 2017) wählt folgendes Design: 16 Testende führen den Vorlauftest aneinander aus. Dies geschieht in drei Runden. Nach Runde eins und zwei findet ein

Konsenstraining statt, in dem folgendes geschieht:

Festgelegte Parameter werden besprochen, wie zB: Positionierung der Patientin und des Patienten, Therapeuten- und Therapeutinnenposition und Blickrichtung, Palpationsort, Druckstärke, Definition des positiven Tests über Abstand in cm (Ausgangsstellung spielt keine Rolle), Bewegung des Patienten. Weiters wird Zeit gegeben, um während der den Testungen aufkommende Fragestellungen zu diskutieren und sich auf eine Ausführung zu einigen. Die Ergebnisdarstellung geschieht in Prozentangaben. Jakobs kommt zu dem Ergebnis einer 11,67%igen Steigerung der Interraterreliabilität.

Die Grundlage für die Testausführung bildet die Palpation. Zur Verbesserung der Lehre in Bezug auf die Palpation haben (Aubin, Gagnon, & Morin, 2014) ein pädagogisches Konzept ausgearbeitet. Es schlägt eine Sieben-Schritte-Anleitung für den Palpationsunterricht vor, die hier wiedergegeben werden soll.

Erstens, die gute Positionierung der palpierenden Person, da jegliche unangenehme Spannung zur Beeinträchtigung des Palpationsvermögens führen kann. Zweitens, die exakte Visualisierung der Anatomie. Es folgen die Suche nach der richtigen Tiefe des Kontaktes mit dem Gewebe und die Intention der Berührung – jede und jeder muss sich klar machen, was er mit der Berührung erreichen will. Wo liegt das Punctum fixum, auf das sich die Bewegung, die die Testende und der Testende induziert, bezieht. Es soll beobachtet werden, was mit diesem Punctum fixum geschieht. Der sechste und siebte Punkt beziehen sich auf die Feinabstimmung der fünf zuvor erworbenen Fähigkeiten: das Verbalisieren und die Anwendung der zu lernenden Tests und Behandlungen.

(Krause, 2017) verweist auf die verschiedenen Herangehensweisen an die Palpation, von denen es so viele gibt, wie es Dozenten und Dozentinnen gibt. Er hebt hervor, dass Automatisierung von Abläufen dazu beiträgt, Kapazität für Wahrnehmung zu schaffen. Diese Idee findet sich auch in verschiedenen Lernpsychologietheorien wieder, wie zum Beispiel der cognitive load theory von (Sweller, Merrienboer, & Paas, 1998).

Die Automatisierung der Tests ist im Konzept der LMO besonders ausgeprägt. Die Testabfolge in den acht Test-Einheiten wird von den Therapeuten und

Therapeutinnen immer ident ausgeführt. Dies schafft die Möglichkeit, sich mehr der Wahrnehmung von Abweichungen von der Norm, als der reinen Test-Ausführung zu widmen (vgl. S.2).

Interrater-Reliabilität

Die empirische Forschung legt für quantitative Forschung nach (Chevron, Köpl, Payrhuber A., & Reinprecht, 2012) drei Gütekriterien fest, die aufeinander aufbauen: Objektivität, Reliabilität, Validität.

Objektivität bedeutet, dass es (zu messende) Werte gibt, die unabhängig vom Messenden entstehen. Die Daten variieren nicht in Abhängigkeit vom Messenden.

Reliabilität bezieht sich auf die Zuverlässigkeit des Messinstruments.

Validität gibt Auskunft über die Gültigkeit der Messung, die es ermöglicht Daten zu interpretieren. Erst wenn Objektivität und Reliabilität gegeben sind kann eine Validität entstehen.

Die Reliabilität (Zuverlässigkeit) bildet eine wichtige Grundlage in der Vergleichbarkeit von Testergebnissen. Interrater-Reliabilität bedeutet in unserem Falle, dass Testende an einer Person einen Parameter testen und anschließend vergleichen, ob eine Übereinstimmung der Ergebnisse gegeben ist. Die Ergebnisse sind reliabel, wenn die Testenden eine hohe Übereinstimmung erreichen. Das Maß der Übereinstimmung drückt sich statistisch im κ -Wert aus (s.S.20).

Wenn die Übereinstimmung von Ratern (Bewertern; in diesem Fall: Testenden) gut ist, dann gibt es eine Möglichkeit, aber keine Garantie, dass die Variablen wirklich die Dimension widerspiegeln, die sie widerzuspiegeln vorgeben. Wenn die Übereinstimmung aber gering ist, dann ist die Sinnhaftigkeit stark eingeschränkt, da es sinnlos ist zu fragen, was die Bedeutung der Variable ist (Fleiss J., Levin B., Paik M., 2003).

In unterschiedlichen Bereichen der Palpation und manuellen Testung wurden Reliabilitätsstudien durchgeführt. Unter dem Suchbegriff: *inter rater reliability manual therapy* finden sich allerdings nur drei relevante *Clinical Trials*. Als nicht relevant ausgeschlossen wurden Studien, die technische Geräte verwenden, an Kadavern forschen und die, die Reliabilität zwischen Mensch und Maschine erforschen. Nur eines der *Clinical Trials* beschäftigt sich mit passiven Bewegungstests. Bei der

Testung des Schulterblattes durch (Deepak, Raghu, & Ramesh, 2015) wurden Menschen mit Schulterschmerzen untersucht. Zwei Untersucher mussten sich darauf festlegen, ob der Test links oder rechts positiv ist, welcher auf eine Festigkeit des Musculus pectoralis minor zurückzuführen ist. Die Ergebnisse liegen mit $\kappa=0,735$ im guten Bereich.

Eine Studie von (Walker, Koppenhaver, Stomski, & Hebert, 2015), die den Anforderungen eines Clinical Trials nicht gerecht wird, untersucht die Reliabilität für die Brustwirbelsäulen-Bewegungs-Testung mit zwei erfahrenen Chiropraktikern und der statistischen Berechnung mit Cohens Kappa. Hier blieben die Ergebnisse mit einem $\kappa < 0,4$ in einem Bereich, der mit einer schlechten Reliabilität gedeutet wird.

Hier sei auch auf die auf S.4 besprochenen Reliabilitäts-Studien verwiesen.

Fehlerquellen LMO-Interrater-Reliabilitäts-Studien

In Interrater-Reliabilitäts-Studien im Bereich der LMO wurde bereits auf einige mögliche Fehlerquellen hingewiesen. Zu unterscheiden sind solche Fehlerquellen, die die Testenden selbst beeinflussen können und andere, die aus dem Studiendesign, der Datenverarbeitung et cetera (etc.) entstehen.

Für das palpatorische Konsenstraining sind vor allem die Fehlerquellen von Bedeutung, die die Testenden selbst beeinflussen können. So nennt (Hafen, 2009) folgende:

- Unterschiedliche Druckstärken der Testenden (Fehlerquelle erhoben durch Patienten- und Patientinnen-Befragung). Hafen empfiehlt eine Schulung zur Abstimmung des Drucks bei der Testung. Laut Aussage der Probanden und Probandinnen (PP) im Vergleich zu den Ergebnissen hatte der Therapeut / die Therapeutin mit dem stärksten Druck auch die meisten positiven Tests, der / die mit dem wenigsten Druck am wenigsten positive Tests. Die Schlussfolgerung daraus ergibt, dass der ausgeübte Druck Einfluss auf die Interpretation des Tests hat.
- Unterschiedliche Ausführung der Tests. Als Ursache wird genannt: unterschiedliches Handling, unterschiedlicher Druck, zu wenig Übung. Die Maßnahme in ihrem Studiendesign war: die Testenden müssen mindestens drei Jahre praktiziert haben.
- Nicht einheitliche Benennung der Wirbelkörper (WK)

- Ursache: unterschiedliche Referenzpunkte → Maßnahme: Beschriftung der WK durch eine unabhängige Person
- Ursache: Positionsänderung der PP verschiebt Beschriftung → Maßnahme: Position des / der PP festlegen und vor Testung überprüfen
- Ursache: falsche Deutung der Beschriftung (an welcher Position wird der Dornfortsatz markiert) → Maßnahme: genaue Definition
- Beeinflussung der Tester
 - Ursache: Mithören, was der andere Tester diktiert → Maßnahme: getrennte Räumlichkeiten.
 - → Einsicht in die PP-Akten → Maßnahme: Einsicht verboten.
- Auswirkung der Testung auf den Bewegungsapparat. → Ursachen: Mobilisation der WS (nur bedingt beeinflussbar), Lange Testzeit, Vegetative Reaktion durch Ventralisation. → Maßnahme: nur eine Testung pro WK pro Durchgang
- Einhaltung von Abmachungen: Abmachung war: jeden Test nur einmal und wenn dieser nicht auswertbar war, diesen dann als negativ zu werten. Alle vier Tester haben stattdessen die Testungen wiederholt, um sicher zu gehen, dass es zu keinen falschen Testergebnissen bei ihrer Testung kommt. Als Auslöser wird die Situation der Studierenerhebung genannt, die Druck erzeugt, da die Testenden möglichst richtig und exakt arbeiten wollen. Als Reaktion auf dieses Verhalten wäre es notwendig in einer Schulung deutlich zu machen, wie mit Unsicherheiten bei Testergebnissen umgegangen wird, Sicherheit im Ablauf zu schaffen und eine entspannte Situation herzustellen.
- nicht beeinflussbare Fehlerquellen sind
 - die Individualität des / der PP und
 - das subjektive Empfinden der Tester.

In einer Studie von (Richter-Schultz, 2010) zu der craniellen Testserie der LMO werden Maßnahmen aufgelistet, die von den Testenden beeinflussbare Fehlerquellen reduzieren sollen:

- Ursache: Ergebnisse der anderen Testenden sind bekannt. → Maßnahme: Verblindung zwischen den Testern durch Testungen in getrennten Räumen.
- Ursache: zu geringe Exaktheit in der Testung → Maßnahme: schriftliche Ausarbeitung des Untersuchungsprotokolls für Tester plus Bildmaterial und

Lehrfilm als zusätzliche Erklärungsebene.

- Ursache: unterschiedliche Druckstärken → Maßnahme: angleichen mittels Briefwaage und PatientInnen – Feedback.
- Ursache: Falsche Reihenfolge der Testausführung führt zu falschem Eintrag im Protokoll → Maßnahme: um die Tests in der richtigen Reihenfolge lückenlos durchzuführen, wurden sie vom Protokollanten verbal vorgegeben und vom Untersucher mit einem Befund von positiv oder negativ beantwortet. Das Ergebnis wurde umgehend in die Datenmaske eingegeben.
- Ursache: Ermüdung durch den unnatürlichen Rhythmus (nicht wie Praxisalltag) führt zu Ermüdung bei den Testenden → keine Maßnahme ergriffen.

Fragestellung und Hypothesen

Kann die Interrater-Reliabilität, von Osteopathen und Osteopathinnen in Ausbildung, für die Tests der unteren Extremität im Konzept von Mechanical Link durch ein Konsenstraining gesteigert werden?

Nullhypothese: Das Konsenstraining von Osteopathen und Osteopathinnen in Ausbildung für die Tests der unteren Extremität nach LMO erhöht die Interrater-Reliabilität nicht.

Alternativhypothese: Das Konsenstraining von Osteopathen und Osteopathinnen in Ausbildung für die Tests der unteren Extremität nach LMO erhöht die Interrater-Reliabilität.

3. METHODIKTEIL

Im Methodikteil werden die Variablen, Rekrutierung, Stichprobe, Ein- und Ausschlusskriterien, Voraussetzungen für die Testenden, Information und Datenschutz, Erhebungs-Ablauf, Tests der unteren Extremität, Ausführung und Bewertung der Tests, Test-Protokoll, Konsenstraining und Statistik vorgestellt.

Variablen

Abhängige Variable: Übereinstimmung zwischen den Testern vor und nach einem Konsenstraining.

Unabhängige Variable: Konsenstraining.

Rekrutierung, Stichprobe, Ein- und Ausschlusskriterien

Die Studierhebung geschieht an gesunden PP. Die Rekrutierung erfolgte über E-Mail-Verteiler, die in studentischen Kreisen verschickt wurden. OSD-Studenten und Studentinnen werden in die Rekrutierung eingeschlossen. Die PP können sich bei den angebotenen Terminen eintragen. Angestrebt werden je 20 Personen bei der Prä- und Postmessung.

Als Einschlusskriterium gilt:

- dass die PP bei der Prämessung in der Lage sind 30 Minuten beziehungsweise (bzw.) bei der Postmessung 90 Minuten entspannt auf dem Rücken zu liegen.

Als Ausschlusskriterium gilt:

- die Unfähigkeit der PP bei der Prämessung 30 Minuten bzw. bei der Postmessung 90 Minuten auf dem Rücken zu liegen.

Die PP sollen in der Zeit der Messung liegen bleiben, um die Messergebnisse nicht zu beeinflussen.

Voraussetzungen für die Testenden

Einschlusskriterium für die Testenden:

- Studenten und Studentinnen an einer Osteopathieschule mit Absolvierung des Seminars *Einheit 3 Extremitäten osteoartikulär mit LMO*.

Für die Testungen herangezogen wurden fünf Personen, die den oben genannten Kriterien entsprechen. Zum Zeitpunkt der Erhebung sind vier davon im 7. Semester ihrer Ausbildung, eine Person hat bereits das letzte Semester des Bachelor-Studiums abgeschlossen.

Information und Datenschutz

Im Vorfeld bekommen die PP eine ausführliche Informations-Mail (s.S.37). Sie beinhaltet Erklärungen zu Ablauf, Kleidung, Einverständniserklärung und Anonymisierungsbogen, Sinn und Zweck der Studie, zur Möglichkeit die Studie jederzeit abzubrechen, Wegbeschreibung.

Dem Informations-Mail als Anhang beigefügt findet sich

- die Einverständniserklärung (s.S.38), mit der bestätigt wird, dass die Teilnahme freiwillig und bezüglich der Studie aufgeklärt geschieht, dass die Zurücknahme der Mitwirkung an der Studie gewährleistet ist, dass die Daten in verschlüsselter Form gespeichert werden dürfen, dass eine Kopie der Einverständniserklärung ausgehändigt wurde
- der Anonymisierungsbogen (s.S.Fehler: Referenz nicht gefunden): erstellt wird ein Code zusammengesetzt aus:
 - Erster Buchstabe des Vornamen des Vaters: **Norbert**
 - Erster Buchstabe des Vornamen der Mutter: **Alma**
 - Erste zwei Ziffern des eigenen Geburtstages: **01.01.1980**
 - Letzte zwei Ziffern des eigenen Geburtstages: **01.01.1980**
 - Geschlecht: m oder w

Erhebungs-Ablauf

Die Erhebungen für diese Studie finden im Rahmen einer größeren Erhebung, für drei Bachelorarbeiten, statt. Die Themen dieser drei Bachelorarbeiten umfassen den Vergleich der Interraterreliabilität vor und nach einem Konsenstraining, eine Interraterreliabilitätsstudie und eine Intraraterreliabilitätsstudie.

Hier wird kurz der Versuchsaufbau der Gesamterhebung geschildert, um dann näher auf den für diese Studie relevanten Teil einzugehen.

Gesamterhebung

- 1) Prä-Testung mit 20 PP. Vier Termine mit je fünf PP für 30 Minuten. Fünf Testende. Umfang: 18 Tests aus der LMO-Einheit 3: Extremitäten osteoartikulär werden von jedem Testenden ein Mal an jedem / jeder PP durchgeführt. Zeit pro PP etwa 5 Minuten.
- 2) Konsenstraining bezogen auf die sieben Tests der unteren Extremität. Fünf Testende, drei Termine zu je 1,5 Stunden.
- 3) Post-Testung bzw. Ersterhebung für eine Interrater- und einer Intraraterstudie. Vier Termine mit je 10 PP für 1,5 Stunden an zwei Tagen. Fünf Testende. 18 Tests aus der LMO-Einheit 3: Extremitäten osteoartikulär werden von jedem Testenden zwei Mal an jedem / jeder PP durchgeführt. Zeit pro PP etwa 4 Minuten.

Aus diesem Studiendesign herausgearbeitete relevante Erhebungs-Details für die vorliegende klinische Studie:

Teilerhebung Konsenstraining

Fünf Osteopathen und Osteopathinnen in Ausbildung testen bei der Prä-Testung an 20 PP den Ablauf der sieben Tests der unteren Extremität aus dem Konzept von LMO.

Der Zeitaufwand für die PP beträgt 30 Minuten. Die Testungen werden von den fünf Osteopathinnen und Osteopathen nacheinander ausgeführt. Die Testenden haben etwa 2,5 Minuten Zeit für die Ausführung der sieben Tests pro Person und das handschriftliche Eintragen der Testergebnisse in vorgefertigte Protokolle.

Die Post-Testung entspricht der Prä-Testung. Allerdings ist der Zeitaufwand für die

PP durch die weiteren in der Gesamterhebung zu erfassenden Daten deutlich höher.

Es wird davon ausgegangen, dass die Testung durch die fünf Testenden die Restriktionen des Gewebes nicht mobilisieren und somit die Testergebnisse nicht verfälschen können.

Tests der unteren Extremität

Die Testreihe, die auf ihre Reliabilität vor und nach dem Konsenstraining getestet wird, besteht aus globalen Tests für die untere Extremität aus der LMO Einheit 3: Extremitäten osteoartikulär. Die getesteten Körperregionen sind: Hüfte, Oberschenkel, Knie, Patella, Unterschenkel, Sprunggelenk und Fuß.

Es sind zwei Arten von Tests vertreten:

- **Gelenktests:**
 - Hüfte – Test zwischen Ilium und Caput femoris
 - Knie – Test zwischen Femur und Tibia
 - Patella – Test zwischen Femur und Patella
 - oberes Sprunggelenk – Test zwischen Malleolengabel (bestehend aus Tibia und Fibula) und Talus
 - Fuß – Test der Tarsometatarsalgelenke (Lisfranc-Gelenklinie gebildet durch die Gelenkflächen von Ossa cuneiformia und Os cuboideum zu der Basis der Ossa metatarsalia)
- **Intraossäre Tests:**
 - Oberschenkel – Femur intraossär
 - Unterschenkel – Tibia und Fibula intraossär

Ausführung und Bewertung der Tests

Es wird durch die verschiedenen Testrichtungen wie Druck, Traktion (Zug), Flexion (Biegung) und Torsion (Verdrehung) die Elastizität des Gewebes getestet (Hafen, 2015).

Die Ausführung der Testrichtungen werden durch lateinische Richtungsbezeichnungen nach (Schünke M., Schulte E., Schumacher U., Voll M., Wesker K., 2011) näher beschrieben, die sich an der Lage des/der PP orientieren:

- kranial = zum Kopf hin
- kaudal, inferior = zum Schwanz/unten
- lateral = seitlich
- medial = mittig
- ventral, anterior = bauchwärts/vorne
- dorsal, posterior = rückenwärts/hinten

Sechs der sieben Tests werden von je einem / einer Testenden links und rechts zeitgleich ausgeführt. Dadurch ist es möglich, dass hierbei ein Balancetest geschieht. Es kann immer nur einer der beiden Tests positiv sein. Es gibt somit drei mögliche Testergebnisse: recht positiv, links positiv, negativ.

1) Hüfte - Gelenktest

- Kontakt: Daumen lateral der A.femoralis auf der Projektionsfläche vom Caput femoris. Die Finger zeigen nach lateral und umgreifen den Trochanter major.
- Ausführung: Traktion medial-kaudal, Traktion lateral-kaudal, Posteriorisierung

2) Oberschenkel – intraossärer Test

- Kontakt: zwei Hände auf einem Oberschenkel, Finger schauen vom Testenden weg
- Ausführung: Druck, Biegung (auseinanderspreizen der Finger), Torsion
- Falls beide Tests positiv sind wird ein Balancetest ausgeführt, um festzustellen welcher der beiden Tests in das Protokoll als positiv eingetragen wird.

3) Knie – Gelenktest

- Kontakt: Daumen und Mittelfinger an Gelenkspalt
- Ausführung: Druck, Traktion

4) Patella – Gelenktest

- Kontakt: Daumen, Zeigefinger, Mittelfinger Kontakt an Patella möglichst gleichmäßig an ihren Rändern
- Ausführung: Traktion, Circumduktion

5) Tibia – intraossärer Test

- Kontakt: Daumen zeigt in Richtung Tibiaplateau, liegt auf Tibia, Finger umschließen Fibula
- Ausführung: Druck, Biegung, Torsion

6) OSG – Gelenktest

- Kontakt: inferior der Malleolen auf Talus
- Ausführung: Druck, Traktion in Achse der Tibia

7) Fuß – Gelenktest

- Kontakt: Daumen medial, Finger umfassen Metatarsale, Einhaken unter Metarsus V
- Ausführung: Druck, Traktion in Achse des Fußes

Test-Protokoll

Auf dem Test-Protokoll sind die 18 ausgeführten Tests der Gesamterhebung aufgeführt. Angegeben werden die Testergebnisse, das PP-Kürzel und Datum und Uhrzeit der Erhebung.

Die drei möglichen Testergebnisse können durch Ankreuzen von

- re – für rechts positiv
- li – für links positiv oder
- nicht ankreuzen

angegeben werden. Tabelle 1 zeigt exemplarisch einen Protokoll-Ausschnitt. Das Protokoll ist auf S.40 nachzuschlagen.

Tabelle 1: exemplarischer Ausschnitt aus dem Test-Protokoll

Pr.Kürzel:

| | | |
|--------------|----|----|
| Hüfte | re | li |
| Oberschenkel | re | li |

Konsenstraining

Das Konstraining beinhaltet:

Auseinandersetzung mit den spezifischen Tests 1) - 3)

- 1) Auseinandersetzen mit anatomisch-palpatorischen landmarks der getesteten Bereiche, mittels Anatomiebildern, Palpation und gegenseitigem Angleichen der Herangehensweise an die Palpation.
 - **Hüfte.** A.femoralis in Lacuna vasorum inferior des Lig.inguinale. Lateral davon ist die Projektionsfläche des Hüftkopfes zu finden. Lateral davon in der Lacuna musculorum befindet sich der M.iliopsoas.
Trochanter major. Seitlicher Vorsprung des Femurs. Zu beachten ist die Außenrotation, die ihn etwas dorsaler liegen lässt.
 - **Oberschenkel.** Ausrichtung von Knie zu Trochanter major. Zu beachten ist die schräge Stellung sup-lat nach inf-med.
 - **Knie.** Epicondylus medialis und lateralis. Inferior davon liegt der Gelenkspalt zwischen Femur und Tibia.
 - **Unterschenkel.** Facies medialis tibiae als mediale Fläche des Schienbeins.
 - **OSG.** Malleolus lateralis und medialis, die sich um den Talus legen.
 - **Fuß.** Tuberositas ossis metatarsi V, Basis des Os metatarsale I
(Schünke M., Schulte E., Schumacher U., Voll M., Wesker K., 2011),

- 2) angleichen des ausgeübten Drucks der Testung über Rücksprache mit dem / der PP,

- 3) versuchen Lösungen für auftauchende Probleme und Unterschiede bei den einzelnen Tests zu finden,

Auseinandersetzung mit allgemeinen Faktoren 4) - 5)

- 4) Fehlerquellen besprechen:
 - besprechen möglicher studieninterner Fehlerquellen,
 - sich mit Fehlerquellen, die andere Studienautoren und Studienautorinnen

nennen, auseinandersetzen und für das eigene Testen anzuwenden. Besprochene Details entsprechen den aufgelisteten Fehlerquellen im Hintergrundteil (s.S.9),

- 5) wichtige Lernschritte für die Palpation durchgehen (Aubin et al., 2014):
- angenehme Position für den Testenden,
 - Visualisierung der Anatomie,
 - richtiger Druck für den Kontakt mit dem Gewebe,
 - Intention der Berührung klar machen,
 - bei Testung bewegte Struktur und Punctum fixum fokussieren,
 - Feinabstimmung dieser fünf Punkte und Anwendung auf die Testausführung.

Geplanter Ablauf

Das Konsenstraining findet an drei Terminen, mit Schwerpunkten auf die spezifischen Tests und allgemeinen Faktoren, statt.

Erster Termin nur mit den Testenden, 1,5 Stunden:

- Hüfte
 - Annäherung, entspricht 1): landmarks, anatomische Bilder
 - Testteil, entspricht 2) – 3) und der Testausführung s.S.15: zwei Personen in Rückenlage, drei testen die Hüfte. Die Liegenden geben Feedback über die verschiedenen Druckqualitäten. Zwei Durchgänge, nachtesten möglich und erwünscht, gegenseitige Absprache und Angleichung der Ausführung. Dann legen sich zwei andere in Rückenlage. Gleicher Ablauf.
- Oberschenkel (Ablauf wie bei Hüfte)
- Verminderung von Stressfaktoren durch Besprechung zum Ablauf der Studie, Reflexion von Stressfaktoren, Zurückkehren zu einer ruhigen inneren Haltung. Ziel: bewusst machen von Stressfaktoren und Leistungsdruck (vergleiche: Fehlerquellen anderer Studien S.9)
- Fünf Palpationspunkte von (Aubin et al., 2014) durchgehen und damit dann
- die Testungen von Hüfte und Oberschenkel wiederholen.

2. Termin mit Testenden und zwei PP:

- Fehlerquellen anderer Studien und ihre Auswirkungen auf unsere Testungen
- Wiederholung (Wh) Hüfte, Oberschenkel
- Knie (Ablauf wie bei Hüfte)
- Patella (Ablauf wie bei Hüfte)
- Unterschenkel (Ablauf wie bei Hüfte)

3. Termin mit Testenden und zwei PP:

- Wh Hüfte, Oberschenkel, Knie, Patella, Unterschenkel (Aubin)
- OSG (Ablauf wie bei Hüfte)
- Fuß (Ablauf wie bei Hüfte)

Statistik

Der Standard für Reliabilitäts-Auswertung von mehr als zwei Ratern ist Fleiss' Kappa. Dieser entstand durch die Ergänzungen von Fleiss zu Cohens Kappa, der nur einen Vergleich zweier Rater gestattet. (Fleiss J., Levin B., Paik M., 2003) zeigt in Kapitel 18 die Berechnung der Interrater Variabilität und beschreibt ein Kappa von

1 als perfekte Übereinstimmung,

>0,75 als ausgezeichnete Übereinstimmung,

0,4-0,75 als moderate Übereinstimmung,

<0,4 als schlechte Übereinstimmung,

0 als keine Übereinstimmung,

negative Werte sind möglich: weniger als zufällig.

Die Möglichkeit einer zufälligen Übereinstimmung wird erfasst. Der Nachteil dieser Berechnungsmethode für diese Studie ist, dass viele Ergebnisse in einer Kategorie dazu führen, dass der Kappa-Wert sinkt. Nachteilig ist dies deshalb, da aus der Praxis bekannt ist, dass viele der Tests negativ sind und nur ein paar positiv. Das wird sich negativ auf die Messergebnisse auswirken. Deshalb wird weiters die mittlere paarweise Übereinstimmung aller Rater in Prozent angegeben.

Die statistische Auswertung der beiden Interrater-Reliabilitäts-Messungen wird mit den opensource Programmen:

- winpepi (<http://www.brixtonhealth.com/pepi4windows.html>) und
- recal (<http://dfreelon.org/utis/recalfront>)

gemacht.

Die zeitgleiche Testung der rechten und linken Körperhälfte (s.S. 15) führte uns zu der Entscheidung drei Antwort-Kategorien festzulegen: negativ, rechts positiv, links positiv.

(Hafen, 2018), die auch mit der zeitgleichen Messung der linken und rechten Körperhälfte arbeitet, wählt zwei mögliche Antwortkategorien: positiv und negativ jeweils für links und rechts; sie haben somit doppelt so viele Tests in ihrem Protokoll.

Die Verbesserung von Prä- zu Post-Testung wird anhand der vier von (Fleiss J., Levin B., Paik M., 2003) vorgeschlagenen Kategorien gezeigt. Das Konsenstraining hat dann eindeutig zu einer Verbesserung geführt, wenn der Kappa-Wert sich von einer Kategorie in die nächst höhere Kategorie steigern lässt.

4. ERGEBNISSE

Hier werden die Ergebnisse in den Bereichen Probanden und Probandinnen Verteilung und Dropout, Datenauswertung und Herausgearbeitete Inhalte des Konsenstrainings dargestellt.

Probanden und Probandinnen Verteilung und Dropout

Wie in Abbildung 1 ersichtlich konnten je 18 PP bei Prä- und Posttestung getestet werden. Das Alter beträgt im Mittel 28,33 ($\pm 6,20$) Jahre bzw. 26,39 ($\pm 3,03$) Jahre. Die Geschlechterverteilung ist bei beiden Erhebungen sehr ähnlich, insgesamt nahmen 25 Frauen und 11 Männer teil (vgl. Abbildung 2, Abbildung 3, Abbildung 4).

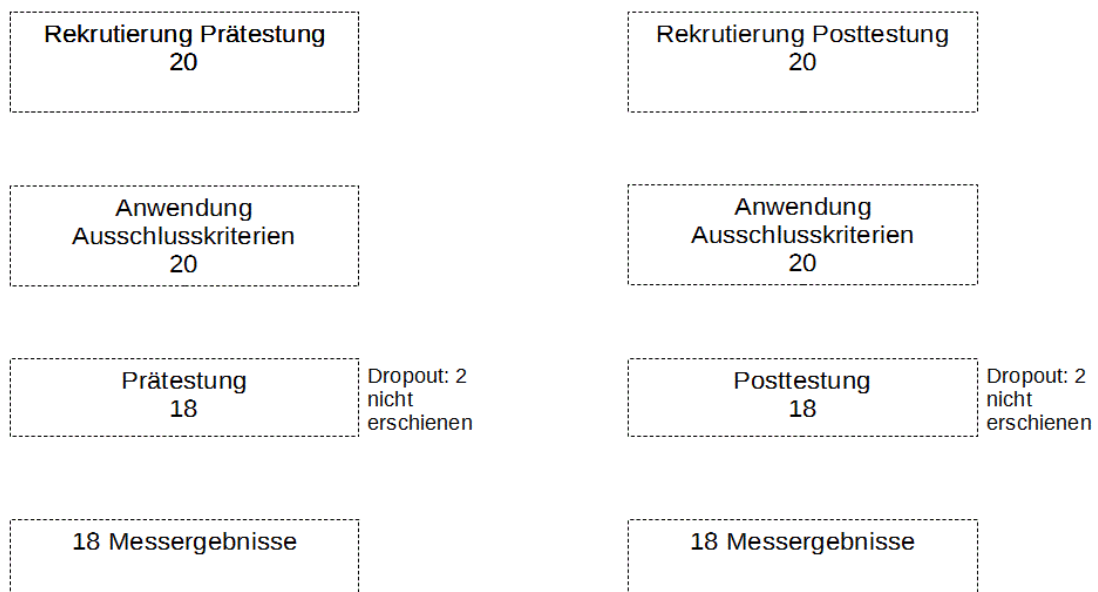


Abbildung 1: Flussdiagramm Versuchsaufbau

Anzahl der Probanden und Probandinnen

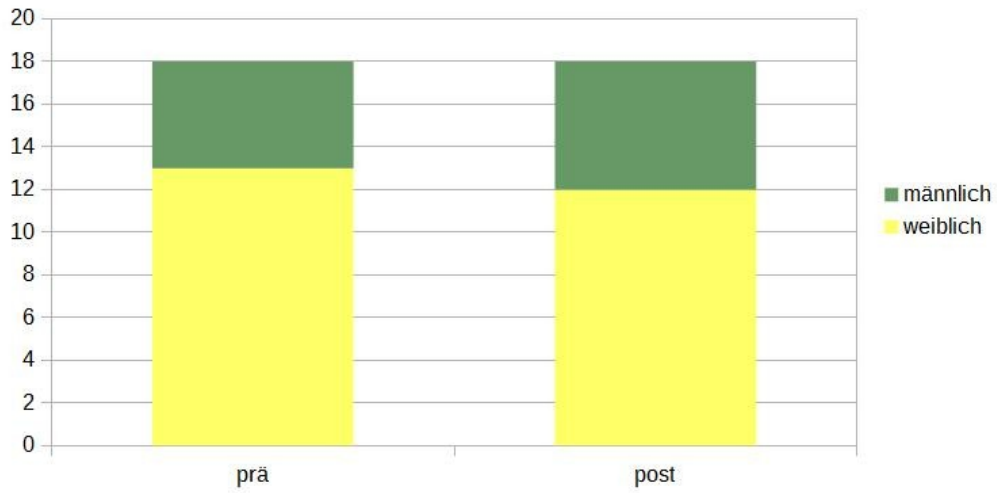


Abbildung 2: Geschlechtsverteilung der Probanden und Probandinnen bei Prä- und Posttestung



Abbildung 3: Prätestung - Alter

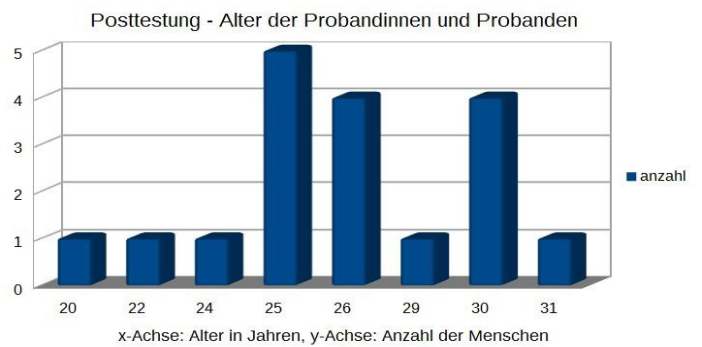


Abbildung 4: Posttestung - Alter

Datenauswertung

Fleiss' Kappa

Die kappa-Werte in der Berechnung mit Fleiss' Kappa zeigten, sowohl bei der Prä- als auch bei der Post-Testung keine signifikanten Ergebnisse, nur bei der Berechnung aller Übereinstimmungen der Postmessungen ergab sich ein signifikantes Ergebnis ($P=0,02$).

Alle sieben Tests zeigten ein κ im Bereich von -0,11 bis 0,07. Die Ergebnisse blieben somit weit unter einem $\kappa=0,4$ und es kam auch zu keiner Verbesserung durch das Konsenstraining.

Tabelle 2: Fleiss' Kappa Ergebnisse für Prä- und Post-Testung

| | Prätestung | | Posttestung | |
|---------------|------------|-------|-------------|-------|
| | Kappa | P | Kappa | P |
| Hüfte | -0,03 | 0,726 | 0,07 | 0,11 |
| Oberschenkel | -0,02 | 0,608 | 0,03 | 0,29 |
| Knie | -0,02 | 0,672 | -0,09 | 0,933 |
| Patella | -0,05 | 0,81 | 0 | 0,49 |
| Unterschenkel | 0,04 | 0,246 | 0,01 | 0,435 |
| OSG | -0,11 | 0,955 | -0,1 | 0,953 |
| Fuß | -0,02 | 0,661 | -0,04 | 0,761 |
| Gesamt | -0,01 | 0,6 | 0,04 | 0,02 |

Es ist zu sehen, dass sich vier Tests (Hüfte, Oberschenkel, Patella, OSG) minimal verbessert haben, während sich drei (Knie, Unterschenkel, Fuß) verschlechtert haben. Die Übereinstimmung aller sieben Test-Werte zusammen führen insgesamt auch zu einer leichten Verbesserung der Ergebnisse in Bezug auf den Kappa-Wert.

Die Ergebnisse im einzelnen werden in den folgenden Abbildungen dargestellt. Die vier Tests mit einer Tendenz zur Verbesserung und die drei Tests mit einer Tendenz zur Verschlechterung werden miteinander dargestellt (vgl. Abbildung 5 bis Abbildung 12).

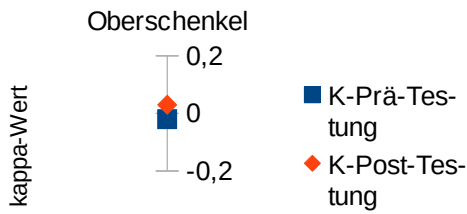


Abbildung 5: Oberschenkel. Prätest $K=-0,02$ Posttest $K=0,03$

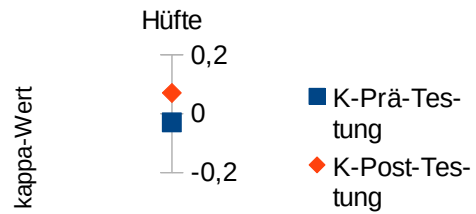


Abbildung 6: Hüfte. Prätest $K=-0,03$ Posttest $K=0,07$

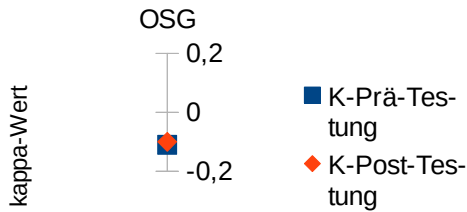


Abbildung 7: OSG. Prätest $K=-0,11$ Posttest $K=0,1$

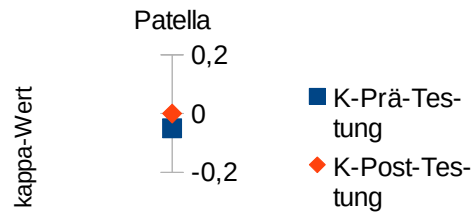


Abbildung 8: Patella. Prätest $K=-0,05$ Posttest $K=0,00$

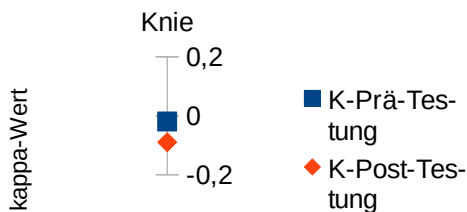


Abbildung 10: Knie. Prätest $K=-0,02$ Posttest $K=-0,09$

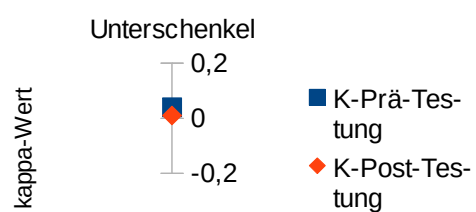


Abbildung 9: US. Prätest $0,04$ Posttest $0,01$

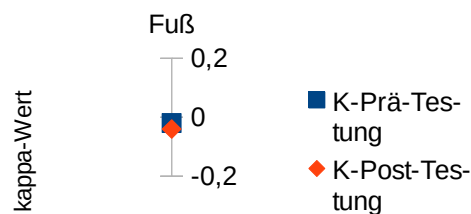


Abbildung 11: Fuß. Prätest $K=-0,02$ Posttest $K=-0,04$

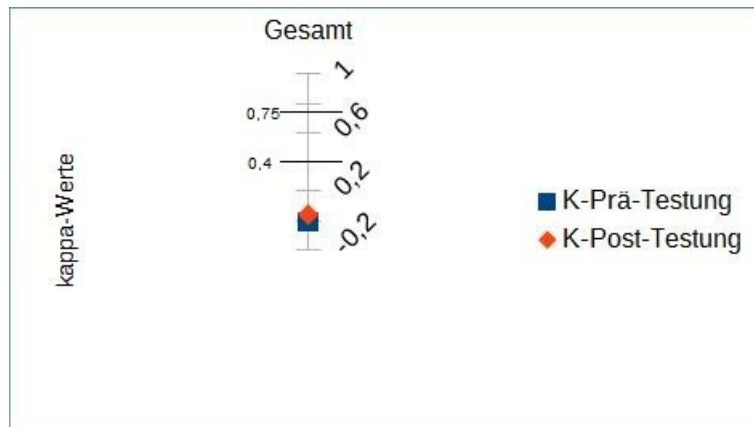


Abbildung 12: Gesamtergebnis aller Testungen. Prätest $K=-0,01$ Posttest $K=0,04$

Zur Verdeutlichung der statistischen Normwerte sind in Abbildung 12 zusätzlich die Werte 0,75 und 0,4 zu sehen, das sind die von Fleiss festgelegten Werte für schlechte Übereinstimmung ($<0,4$) und ausgezeichnete Übereinstimmung ($>0,75$).

Paarweise prozentuelle Übereinstimmung

Dargestellt als Vergleich der paarweisen prozentuellen Übereinstimmung kommt es zu den in Tabelle 3 gezeigten Ergebnissen. Der Test mit der höchsten Übereinstimmung ist der Fuß, dieser liegt als einziger über 60%.

Tabelle 3: Paarweise mittlere Übereinstimmung der Prä- und Posttestung

| | Prä | Post |
|---------------|--------|--------|
| Hüfte | 48.33% | 42.22% |
| Oberschenkel | 58.33% | 44.44% |
| Knie | 41.11% | 51.11% |
| Patella | 41.11% | 36.11% |
| Unterschenkel | 51.11% | 49.44% |
| OSG | 51.67% | 55% |
| Fuß | 56.67% | 66.67% |
| Gesamt | 49.76% | 49.29% |

Herausgearbeitete Inhalte des Konsenstrainings

Die im Methodenteil (s.S.18) genannten Punkte wurden bearbeitet. Darüber hinausgehende Schlussfolgerungen und Einigungen werden aufgelistet:

- 1) **Allgemein:** wenn eine Testrichtung positiv getestet wird, ist der ganze Test positiv.
- 2) **Hüfte, OS, Patella, US:** wenn mehrere Testrichtungen aufeinander aufgebaut werden, müssen alle ausgeführt werden. Der Test ist da positiv, wo am Ende die Elastizität am geringsten ist.
- 3) **Knie, OSG, Fuß:** die Kompression verstehen wir nur als eine Annäherung an die zu testende Struktur. Sie ist kein Teil des Tests und kann somit nicht als positiver Test gewertet werden. Aktualisierte Testrichtung: Traktion
- 4) **Oberschenkel:** die beiden Tests werden nacheinander ausgeführt und im Falle zwei positiver Tests ausbalanciert. Daher wird die Traktion anstatt des Druckes als Elastizitätsrichtung hinzugefügt. Aktualisierte Testrichtung: Traktion, Biegung (Daumen zueinander), Torsion
- 5) **Patella:** Die beiden Circumduktionen entsprechen nicht dem Übereinanderlegen mehrerer Elastizitätsrichtungen. Es gibt zwei unterschiedliche Testrichtungen. Daher werten wir den Test als positiv, wenn nur eine Circumduktion in ihrer Elastizität eingeschränkt ist, es muss die zweite Circumduktionsrichtung nicht mehr getestet werden. Aktualisierte Testrichtung: Traktion, Circumduktion in beide Richtungen
- 6) **Unterschenkel:** Ausführung der Biegung über Druck mit Daumen nach posterior und lateral. Aktualisierte Testrichtung: Biegung posterior, Biegung lateral, Torsion
- 7) **OSG:** Füße in Neutralstellung bringen
- 8) **Fuß:** Testende haben Mittelfinger auf 5. Metatarsus, Daumen auf 1. Metatarsus
- 9) **Stressfaktoren, Leistungsdruck:**
 - aus dem Blickfeld bleiben, wenn man früher fertig ist, als andere.
 - vor jedem neu zu Testenden eine kurze Pause mit zwei tiefen Atemzügen machen, um möglichst unvoreingenommen und nicht gestresst die nächste Testserie durchzuführen.
 - Ruhe bewahren, egal ob der Zeitplan eingehalten werden kann oder nicht.
 - Zu Ruhe kommen der Therapeuten und Therapeutinnen vor dem Start der Testung durch eine Meditation.
 - Wasser bereitstellen.

5. DISKUSSION

Hier werden die Methoden-Diskussion, Ergebnis-Diskussion und die Beantwortung der Fragestellung vorgestellt.

Methoden-Diskussion

In diesem Kapitel wird der Methodenteil, mit Fokus auf Rekrutierung, Studiendesign, Gewebe-Veränderungen durch mehrfache Testungen, Ausführung und Bewertung der Tests und Konsenstraining beleuchtet.

1 Rekrutierung

Die Erhebung fand mit durchschnittlich 27 Jahre alten Menschen statt. Dieser Faktor der Stichprobe hat eventuell Einfluss auf die Ergebnisse der Studie. Die Ergebnisse können nicht auf die Gesamtbevölkerung übertragen werden.

Den angenommenen Einfluss des Alters wird wie folgt begründet: Schmerzen und andere Beschwerden gehen nach (Fossum, 2010b) mit Restriktionen im Gewebe einher. Die meisten jüngeren Menschen haben weniger Restriktionen im Gewebe und weniger gesundheitliche Beschwerden. Schwach ausgeprägte Restriktionen sind palpatorisch schwieriger übereinstimmend zu bewerten.

Eine Erhebung in einem Personenkreis mit Schmerzen oder anderen Beschwerden hätte zu größerer Interraterreliabilität führen können. Die Erhebung ist somit nicht auf Extremgruppen übertragbar.

2 Studiendesign

Die Organisation, die Kommunikation mit den PP und der Ablauf der Termine zur Erhebung und Konsenstraining verliefen sehr positiv. Was im Studiendesign noch zu verbessern wäre, ist die Teilnahme der Studienleiterin an der Erhebung.

3 Gewebe-Veränderungen durch mehrfache Testungen

Es ist für alle Reliabilitätsstudien von großer Wichtigkeit, wie oft ein Test durchgeführt werden darf, bevor es zu Gewebe-Veränderungen kommt: Gibt es keine gleichbleibenden Restriktionen, kann es auch zu keinen gleichbleibenden und

reliablen Messergebnissen kommen.

Aus dem Grund, dass lebende Systeme sich ständig verändern (subject bias), bemerkt (Sommerfeld, 2006), dass die Reliabilität ein unerreichbares Ideal sei, da sich die Ausgangsbedingungen zwischen den Messungen immer unterscheiden.

In der Studie wurde vorausgesetzt, dass wiederholende Testung – in diesem Fall fünf- bzw. zehnfach – zu keinen lokalen Gewebe-Veränderungen führt. Diese Hypothese ist zu hinterfragen.

Unter Umständen verändern sich die Läsionen der PP während des Testablaufs, was dazu führt, dass die Testenden nicht zu einheitlichen Ergebnissen kommen können.

Ab wann Testungen zu Gewebe-Veränderungen führen, ist unklar. Das Ausmaß der Gewebe-Veränderung, von der wir in diesem Zusammenhang sprechen, ist sehr gering. Im schulmedizinischen Sprachgebrauch wird über Gewebe-Veränderungen meist im Zusammenhang mit kanzerogenen Tumoren gesprochen, deren Größe und Veränderung über die Zeit mittels teurer Geräte aufgezeichnet werden kann. Nach (Hilton & Jones, 2014) sind übliche technische Geräte für die Verlaufskontrolle zB CT, MRT, PET. Nach (Cooley et al., 2014) ist auch das Sonographie-Gerät ein wichtiges Gerät zur Verlaufskontrolle.

Die Erfassung der Gewebe-Veränderung aber, von der wir bei der Überprüfung der Elastizität sprechen, liegt entweder in einem viel geringeren Grad oder vielleicht auch in einer Qualitätsänderung des Gewebes, anstatt einer Vermehrung oder Verringerung. Sie lässt sich unter Umständen mit den technischen Mitteln, die heute zur Verfügung stehen, gar nicht erfassen. Und selbst wenn es diese technischen Geräte gäbe, wäre die Anwendung auf Grund der finanziellen Mittel der Forschung im manuellen Tätigkeitsfeldern sehr beschränkt. In der Grundlagenforschung bieten die Arbeiten von (Benias et al., 2018), die sich mit den Flüssigkeiten im interstitiellen Raum beschäftigen und (Schleip, 2016), der sich mit Mechanotransduktion von Zellen auseinandersetzt, vielleicht einen Ansatz, um die Vorgänge, die hier geschehen, besser zu verstehen.

Erst nach Ausarbeitung eines Modells, das beschreibt, wie es zu den Veränderungen kommen kann und Klärung der möglichen technischen Aufzeichnung, lässt sich womöglich auf wissenschaftlicher Basis ein Leitfaden erstellen, der angibt, welche Art von Berührung, in welcher Dauer, mit welcher Druckstärke und vielen anderen noch unbekanntem Faktoren eine momentane Veränderung im Gewebe hervorruft.

Die wissenschaftliche Evaluation von Testungen wird sich dann an diesen Rahmen

halten müssen, wo es noch zu keinen Gewebe-Veränderungen kommen kann.

Bis zu diesem Zeitpunkt wird mehr oder weniger willkürlich, nach bestem Wissen und Gewissen, beschlossen werden, wie das Studiendesign aufgebaut wird.

4 Ausführung und Bewertung der Tests

Die Tests wurden so ausgeführt, wie sie auch in der Praxis angewendet werden. Das führt aber dazu, dass nicht nur der Test, sondern auch der bei der gleichzeitigen Testung der linken und rechten Extremität automatisch ausgeführte Balancetest mit beurteilt werden muss. Hier werden im Praxisalltag zwei prinzipiell getrennte Arbeitsschritte zu einem zusammengefasst. Im Studienaufbau ist dies ungünstig, da eigentlich kein Rückschluss auf die einzelnen Tests, sondern nur auf Testpaare und den Balancetest gezogen werden kann. Da aber Prä- und Postmessung das gleiche Verfahren anwenden, verändert dies die Gültigkeit der Ergebnisse für das Konsenstraining nicht.

5 Konsenstraining

Der Ablauf des Konsenstrainings war strukturiert und zeitlich begrenzt.

Die Dauer des Konsenstrainings war in dieser Studie auf drei Termine beschränkt. Eine vorgegebene Anzahl an Terminen ist für die Studienergebnisse von Nachteil, allerdings erforderlich um innerhalb des Zeitrahmens die Studie zu Ende bringen zu können. Optimal wäre ein Konsenstraining, das erst dann beendet ist, wenn es zu reliablen Ergebnissen kommt, bzw. alle Teilnehmer und Teilnehmerinnen nach einer festzulegenden Zeit beschließen, dass der vorliegende Test auch mit Konsenstraining keine Aussicht auf reliable Ergebnisse bietet. Aufgrund der zeitlichen Begrenzung und der terminlichen Gegebenheiten der fünf Testenden war es nicht möglich mehr Termine für das Konsenstraining festzulegen.

Eine mögliche Ergänzung zum Konsenstraining wäre gewesen, Termine mit einer LMO-Dozentin zu vereinbaren. Dies hätte über die professionellere Herangehensweise eine weitere Annäherung an reliable Ergebnisse gebracht.

Schwierig umzusetzen war die Rückmeldung der PP über die Druckstärke. Wir fanden keinen Weg, um uns auf eine „gleiche“ Ausführung zu einigen. Hier fehlt dem Studienteam eine Handhabung zur Umsetzung dieser sehr subjektiv empfundenen und ausgeführten Druckstärken. Die Herangehensweise über eine Federwaage, wie

es (Richter-Schultz, 2010) durchgeführt haben, wäre hier eine zusätzliche Möglichkeit, die zwar die notwendigen Unterschiede der Druckstärke, aber nicht die differenzierten Testrichtungen (Torsion, Biegung, Traktion) verbessern kann.

Ergebnis-Diskussion

In diesem Kapitel wird der Ergebnisteil, mit Fokus auf Literatur im Hintergrundteil, Signifikanz der Ergebnisse und Kappa-Werte und ihre statistische Auswertung beleuchtet.

1 Literatur im Hintergrundteil

Die Literatur wurde sorgfältig recherchiert und ausgewählt. Es ist zu bemerken, dass es im Bereich der LMO wenige veröffentlichte Artikel gibt und auch nur wenige Master-Thesen, auf die in dieser Studie Bezug genommen wurde.

Der Bezug auf Lehrbücher wird als sinnvoll und notwendig betrachtet, da die Arbeitsmethoden in ihnen gut beschrieben werden.

Die Reliabilität im manuellen Bereich zeigt ein deutlich größeres Feld an veröffentlichten Studien, doch entsprechen nur wenige den wissenschaftlichen Standards eines Clinical Trials, wie die Suche auf pubmed ergab.

2 Signifikanz der Ergebnisse

Die Ergebnisse zeigen keine Signifikanz. Der einzige Wert, der signifikant war, ist die Übereinstimmung aller Testungen der Post-Messung. Es ist davon auszugehen, dass dieses Signifikanz-Niveau von 0,02 nur durch die große Anzahl der erfassten Daten entstanden ist. Somit sind die Daten in der Berechnung nach Fleiss nicht beurteilbar.

3 Kappa-Werte und ihre statistische Auswertung

Bei einem Blick auf die Kappa-Werte zeigt sich, dass alle Ergebnisse der Prä- und Posttestung weit unter dem zu übersteigenden Wert von 0,4 bleiben und somit keine Reliabilität und auch keine Verbesserung durch das Konsenstraining zu sehen ist.

Bei einem Vergleich mit anderen Studien zur Reliabilitätsmessung fällt auf, auf wie viele unterschiedliche Berechnungsmodalitäten zurückgegriffen werden kann. In

dieser Studie wurde nach der Erläuterung von (Fleiss J., Levin B., Paik M., 2003) die Berechnung mit Fleiss' Kappa durchgeführt, da Fleiss' Kappa die Standard-Berechnungs-Methode der quantitativen Forschung für mehr als zwei Rater ist. Die Nachteile wurden im Methodikteil (s.S.20) erörtert.

Die in dieser Arbeit zitierten Studien zur Reliabilität manueller Testungen berechnen ihre Ergebnisse mit Cohens Kappa, Fleiss' Kappa, Holsti Koeffizient oder Krippendorffs Alpha.

Übersichtsarbeiten zu den verschiedenen statistischen Verfahren und deren Anwendbarkeit für manuelle Techniken wurden keine gefunden. Die Ausarbeitung eines Statistikers / einer Statistikerin in Zusammenarbeit mit Personen aus den manuellen Berufen könnte eine Grundlage schaffen, auf die sich die Reliabilitätsforschung im manuellen Bereich beziehen könnte.

In anderen Forschungsbereichen, wie zB der Inhaltsanalyse von Texten nach (Rössler, 2017) wird auf den Holsti Koeffizienten als übliche Berechnungsmethode der Reliabilität verwiesen.

Unter Verwendung des Holsti-Koeffizienten fallen die Ergebnisse deutlich besser aus. Daher ist im Anhang (s.S.41) eine exemplarische Auswertung für die Testenden 1 und 2 mit dem Holsti-Koeffizienten dargestellt. Die Berechnungen führen bei drei Kategorien zu Ergebnissen, die in den Übergangsbereich von schlechter zu moderater Übereinstimmung fallen. Auch wenn die Ergebnisse dann weiterhin nicht gut sind, so sind sie doch wesentlich besser, als bei der Berechnung mit Fleiss' Kappa.

Im Vergleich mit der Studie von (Hafen, 2018) fällt bei der gleichen Art und Weise der Testung die unterschiedliche Wahl der Kategorien zu der vorliegenden Studie auf. Daher wurden in dieser Studie, ebenfalls nur exemplarisch, die Daten der Testenden 1 und 2 umgeschrieben auf das Format, das (Hafen, 2018) verwendet. Bei der erneuten Berechnung des Holsti-Koeffizienten mit nur zwei Antwortkategorien sind die Ergebnisse schon deutlich im Bereich der moderaten Übereinstimmung.

Diese Ergebnisse sind erstaunlich gut, sind aber nicht im Ergebnisteil angeführt, da wie auf Seite 15 erklärt, weder die Berechnung mit dem Holsti-Koeffizienten, noch die Berechnung anhand von zwei Antwortkategorien für diese Studie in Frage kam.

Beantwortung der Fragestellung

Die Interrater-Reliabilität von Osteopathen und Osteopathinnen in Ausbildung für die Tests der unteren Extremität im Konzept von LMO kann durch ein Konsenstraining nicht gesteigert werden.

Die Nullhypothese muss beibehalten bleiben:

Das Konsenstraining von Osteopathen und Osteopathinnen in Ausbildung für die Tests der unteren Extremität nach LMO erhöht die Interrater-Reliabilität nicht.

Ausblick

Folgende Forschungsfragen hat diese Studie aufgeworfen:

Wie oft darf eine Testung durchgeführt werden, ohne dass es bereits zu einer Gewebeveränderung kommt?

Nach welchen Kriterien und nach welcher Zeit der Auseinandersetzung darf von einem Test behauptet werden, dass es nicht möglich ist ihn, trotz Beschäftigung damit, zu einem reliablen Messinstrument zu machen?

Führen Erhebungen in einem Personenkreis mit Schmerzen oder anderen Beschwerden zu größerer Interraterreliabilität als Erhebungen an PP ohne Schmerzen oder anderen Beschwerden?

Welche Möglichkeiten gibt es, um Druckstärke und Testrichtungen anzugleichen?
Gibt es andere Wege des Konsenstrainings?

Welche statistischen Verfahren zur Berechnung der Reliabilität gibt es und welche können im Bereich der manuellen Testung eingesetzt werden?

Würde ein Konsenstraining der hier durchgeführten Tests über längere Zeit zu reliablen Ergebnissen führen?

Schlussbetrachtung

Obwohl die Ergebnisse nach Fleiss' Kappa nicht signifikant waren, lassen verschiedenen statistischen Verfahren einen breiten Interpretationsspielraum. Weitere Forschung ist erforderlich für einen sinnvollen Einsatz der statistischen Verfahren im manuellen Bereich. Das Konsenstraining kann weiter ausgebaut und erforscht werden.

6. LITERATURVERZEICHNIS

- Aubin, A., Gagnon, K., Morin, C. (2014). The seven-step palpation method: A proposal to improve palpation skills. *International Journal of Osteopathic Medicine*, 17(1), 66–72.
- Benias, P., Wells, R., Sackey-Aboagye, B., Klavan, H., Reidy, J., Buonocore, D., Theise, N. (2018). Structure and Distribution of an Unrecognized Interstitium in Human Tissues. *Scientific reports*, 8(1), 4947.
- Chauffour, P., Prat, E. (2002). *Mechanical Link: Fundamental Principles, Theory, and Practice Following an Osteopathic Approach* (1st ed.). California: North Atlantic Books.
- Chauffour, P., & Prat, E. (2010). Der Osteopathische Mechanical Link. In T., Dobler, T., Liem (Eds.), *Leitfaden Osteopathie: Parietale Techniken* (3rd ed., pp. 821–832). München: Elsevier GmbH.
- Chevron, M., Köpl, R., Payrhuber A., Reinprecht, C. (2012). Grundlagen sozialwissenschaftlicher Methodologie: Empirische Forschung in den Sozialwissenschaften: 2.6 Gütekriterien empirischer Forschung (Lernunterlage zur Unterstützung der Vorlesung). Universität Wien, Wien, URL: <http://www.univie.ac.at/sowi-online/esowi/cp/methodologiesowi/methodologiesowi-28.html>, (Zugriff am: 15.03.2018).
- Cooley, C., Nishino, M., Jagannathan, J., Ramaiya, N., Di Salvo, D., Krajewski, K. (2014). Clinical problem solving: utility of sonography in oncologic patients. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine*, 33(1), 9–22.
- Deepak, S., Raghu, C., Ramesh, M. (2015). The neck extensor endurance test: An inter-rater reliability study. *Journal of bodywork and movement therapies*, 19(2), 213–216.
- Degenhardt, B., Snider, K., Snider, E., Johnson, J. (2005). Interobserver Reliability of Osteopathic Palpatory Diagnostic Tests of the Lumbar Spine: Improvements From Consensus Training. *JAOA*, 105(10), 465–473.

- Degenhardt, B., Snider, K., Snider, E., Johnson, J. (2010). Maintenance and Improvement of Interobserver Reliability of Osteopathic Palpatory Tests Over a 4-Month Period. *JAOA*, 110(10), 579–586.
- Fleiss J., Levin B., Paik M. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, New Jersey: John Wiley & Sons.
- Fossum, C. (2010a). Ziel der Untersuchung. In T. Dobler & T. Liem (Eds.), *Leitfaden Osteopathie: Parietale Techniken* (3rd ed., p. 71). München: Elsevier GmbH.
- Fossum, C. (2010b). Die somatische Dysfunktion. In T. Dobler & T. Liem (Eds.), *Leitfaden Osteopathie: Parietale Techniken* (3rd ed., pp. 41–54). München: Elsevier GmbH.
- Hafen, C. (2009). Reliabilitätsstudie des Befunds nach LMO: Reliabilitätsstudie über die Befunderhebung der Wirbelsäule nach der Methode der Lien Mécanique Ostéopathique, Krems, Donau-Universität Krems, Master of Science.
- Hafen, C. (2015, November). Mechanical Link. Funktionseinheit 3. Berlin: 24.-25.11.2015: Einführung ML. Extremitäten global. LMO Lien Mécanique Ostéopathique nach Paul Chauffour und Eric Prat. Osteopathie Schule Deutschland, Berlin.
- Hafen, G. (2018). Reliabilitätsstudie des Befunds nach dem Mechanical Link: Reliabilitätsstudie über die Befunderhebung der Knochen und Gelenke der Extremitäten nach der Methode des Mechanical Link. Fresenius, Idstein.
- Hafen-Bardella, C. (2018). Mechanical Link – Protokoll für eine individuelle Diagnostik und Therapie. *Osteopathische Medizin*, 19(1), 19–23.
- Hilton, S., Jones, L. (2014). Recent advances in imaging cancer of the kidney and urinary tract. *Surgical oncology clinics of North America*, 23(4), 863–910.
- Jakobs, D. (2017). Verbessert sich die Interrater Reliabilität bezüglich des „stehenden Vorlauftests“, wenn die Therapeuten eine spezifische Konsensausbildung durchlaufen? Berlin, Dresden International University, Bachelor of Science.
- Krause, R. (2017). Tools zum Erlernen der Palpation. *Osteopathische Medizin*, 18(4), 21–24.

- Kühn, L. (2018). Follow-up-Studie über die Reliabilität der Befunderhebung der Wirbelsäule nach der Methode des Mechanical Link, Idstein, Fresenius, Master of Science.
- Richter-Schultz, K. (2010). Inter-Intrareliabilitätsstudie zum Test cranialer Strukturen im Konzept der "Mechanischen Vernetzung", Krems, Donau-Universität Krems, Master of Science.
- Rössler, P. (2017). Inhaltsanalyse (3rd ed.). Konstanz und München: UVK Verlagsgesellschaft mbH.
- Schleip, R. (2016). Mechanotransduktion: Von der zellulären Ebene bis zum ganzen Körper. *Osteopathische Medizin*, 17(3), 16–21.
- Schünke M., Schulte E., Schumacher U., Voll M., Wesker K. (2011). Prometheus. LernAtlas der Anatomie. Allgemeine Anatomie und Bewegungssystem. (3rd ed.). Stuttgart, New York: Georg Thieme Verlag.
- Sommerfeld, P. (2006). Methodologie, Kleine und hoffentlich kurzweilige Einführung in den Sinn und Unsinn wissenschaftlichen Arbeitens auf der Grundlage quantitativer Methoden für OsteopathInnen, Donau Universität Krems, Wien.
- Sweller, J., Merrienboer, J., Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*. (10.3), 251–296.
- Walker, B., Koppenhaver, S., Stomski, N., Hebert, J. (2015). Interrater Reliability of Motion Palpation in the Thoracic Spine. *Evidence-based complementary and alternative medicine* 815407, 1-6.

7. ANHANG

Im Anhang befinden sich Information, Einverständniserklärung, Code, Testprotokoll und Holsti-Koeffizient.

Information



Hallo liebe/r !

Danke, dass du an unserer Studie teilnimmst! Bald ist es so weit. Dein Termin ist am:

01.02. 09:00-09:30

Ablauf

Bitte komm fünf Minuten vor Beginn. Wir brauchen alle TeilnehmerInnen, damit wir starten können. Ihr werdet mit 5 anderen ProbandInnen in einem Raum liegen. Jeder auf einer Liege. 5 TherapeutInnen werden nacheinander mehrere Tests mit euch durchführen. Das sind Tests, bei denen ihr auf den Liegen liegt und die Therapeutinnen zum Beispiel euer Handgelenk anfassen, um es durch einen leichten Zug zu testen und dann das Ergebnis in eine Tabelle einzutragen.

Klamotten und Decken

Bitte zieht euch Unterwäsche an, in der ihr euch wohl fühlt. Auch eine kurze Sporthose und ein Träger-Shirt ist in Ordnung. Wir brauchen freie Schultern und freie Beine.

Unterschrift und Anonymisierungsbogen im Anhang

Im Anhang findet ihr einen Anonymisierungsbogen, wir bitten euch, diesen auszufüllen und mitzunehmen. Wir werden eure persönlichen Daten nicht speichern, sondern nur dieses Kürzel verwenden. Wir haben vor Ort dann natürlich auch noch welche, falls ihr nicht drucken könnt.

Abbruch

Ihr könnt jederzeit aus der Studie aussteigen, auch während der Erhebung oder danach. Das ist gar kein Problem. Ihr bleibt nur, wenn es für euch wirklich passt.

Anfahrtsweg



Oudenarderstraße 16, 13347 Berlin

<https://goo.gl/maps/MygDUMGFj7o>

Die Osteopathie Schule liegt in den Osram-Höfen. Eingang über die Oudenarderstraße und die Groninger Straße. Auf dem Bild haben wir euch noch eingezeichnet, wie ihr in die Osram-Höfe hinein kommt und wo ihr dann hin müsst.

2. OG und dann den Schildern folgen.

Wir freuen uns!
Anika, Lisa und Marie-Annick

Einverständniserklärung

DIU (Dresden International University) in Kooperation mit der Osteopathie
Schule Deutschland. Oudenarder Str. 16, 13347 Berlin



PatientInnen-Einverständniserklärung zur Studie über die „Befunderhebung nach dem Modell der Lien Mécanique Ostéopatique“

Name:
Vorname:
Code:

Ich erkläre mich bereit, an der Studie „Studie über die Befunderhebung nach dem Modell der Lien Mécanique Ostéopatique“ freiwillig teilzunehmen.

Ich wurde ausreichend von den Versuchsleitern über Zweck, Ablauf (Dauer, Behandlungsform, Messung), Nutzen, möglichen Risiken und Begleiterscheinungen der Studie aufgeklärt. Ich habe den Aufklärungstext gelesen und verstanden. Bei Unklarheiten meinerseits hatte ich die Möglichkeit Fragen zu stellen, die mir in verständlicher Weise beantwortet wurden. Ich hatte ausreichend Zeit mich für die Teilnahme der Studie zu entscheiden und habe derzeit keine weiteren Fragen mehr.

Ich werde den für die Durchführung der Studie erforderlichen Anordnungen Folge leisten und weiß darüber Bescheid, dass ich innerhalb von 2 Wochen meine Mitwirkung an der Studie ohne Angabe von Gründen zurücknehmen kann.

Ich bin damit einverstanden, dass meine im Rahmen dieser Studie ermittelten Daten in verschlüsselter Form gespeichert werden. Die Bestimmungen des Datenschutzgesetzes in der geltenden Fassung werden eingehalten.

Ich habe eine Kopie von der Studienteilnehmerinformation und der Einverständniserklärung erhalten. Das Original verbleibt beim Studienleiter. Ich erkläre hiermit meine freiwillige Teilnahme an dieser Studie.

Ort und Datum

Unterschrift des Studienleiters

Ort und Datum

Unterschrift des Probanden

Code

Datenspeicherung und Schutz:

Die erfassten Daten werden über einen von euch generierten Code anonymisiert gespeichert.

| | | | | |
|---|--|--|---|--------------------------|
| Erster Buchstabe des Vornamen des Vaters (Norbert) | Erster Buchstabe des Vornamen der Mutter (Alma) | Erste zwei Ziffern des eigenen Geburtstages (01.01.1980) | Letzte Zwei Ziffern des eigenen Geburtsjahres (01.01.1980) | Geschlecht (m oder w) |
| | | | | |

Testprotokoll

Pr.Kürzel:

| | | |
|---------------|----|----|
| Hand | re | li |
| Handgelenk | re | li |
| Karpaltunnel | re | li |
| Unterarm | re | li |
| Ellenbogen | re | li |
| Oberarm | re | li |
| Schulter | re | li |
| B. upslip | re | li |
| B. inflare ↓ | re | li |
| B. outflare ↓ | re | li |
| Pubis | re | li |
| Hüfte | re | li |
| Oberschenkel | re | li |
| Knie | re | li |
| Patella | re | li |
| Unterschenkel | re | li |
| Sprunggelenk | re | li |
| Fuß | re | li |

Pr.Kürzel:

| | | |
|---------------|----|----|
| Hand | re | li |
| Handgelenk | re | li |
| Karpaltunnel | re | li |
| Unterarm | re | li |
| Ellenbogen | re | li |
| Oberarm | re | li |
| Schulter | re | li |
| B. upslip | re | li |
| B. inflare ↓ | re | li |
| B. outflare ↓ | re | li |
| Pubis | re | li |
| Hüfte | re | li |
| Oberschenkel | re | li |
| Knie | re | li |
| Patella | re | li |
| Unterschenkel | re | li |
| Sprunggelenk | re | li |
| Fuß | re | li |

Pr.Kürzel:

| | | |
|---------------|----|----|
| Hand | re | li |
| Handgelenk | re | li |
| Karpaltunnel | re | li |
| Unterarm | re | li |
| Ellenbogen | re | li |
| Oberarm | re | li |
| Schulter | re | li |
| B. upslip | re | li |
| B. inflare ↓ | re | li |
| B. outflare ↓ | re | li |
| Pubis | re | li |
| Hüfte | re | li |
| Oberschenkel | re | li |
| Knie | re | li |
| Patella | re | li |
| Unterschenkel | re | li |
| Sprunggelenk | re | li |
| Fuß | re | li |

Pr.Kürzel:

| | | |
|---------------|----|----|
| Hand | re | li |
| Handgelenk | re | li |
| Karpaltunnel | re | li |
| Unterarm | re | li |
| Ellenbogen | re | li |
| Oberarm | re | li |
| Schulter | re | li |
| B. upslip | re | li |
| B. inflare ↓ | re | li |
| B. outflare ↓ | re | li |
| Pubis | re | li |
| Hüfte | re | li |
| Oberschenkel | re | li |
| Knie | re | li |
| Patella | re | li |
| Unterschenkel | re | li |
| Sprunggelenk | re | li |
| Fuß | re | li |

Pr.Kürzel:

| | | |
|---------------|----|----|
| Hand | re | li |
| Handgelenk | re | li |
| Karpaltunnel | re | li |
| Unterarm | re | li |
| Ellenbogen | re | li |
| Oberarm | re | li |
| Schulter | re | li |
| B. upslip | re | li |
| B. inflare ↓ | re | li |
| B. outflare ↓ | re | li |
| Pubis | re | li |
| Hüfte | re | li |
| Oberschenkel | re | li |
| Knie | re | li |
| Patella | re | li |
| Unterschenkel | re | li |
| Sprunggelenk | re | li |
| Fuß | re | li |

| |
|----------------|
| DATUM: |
| UHRZEIT START: |

Holsti-Koeffizient

Berechnung des Holsti-Koeffizienten nach (Rössler, 2017): es werden die Ergebnisse von zwei Testenden verglichen. Bei mehr als zwei Testenden muss jeder Testende mit jedem verglichen werden. Der Mittelwert aller Paare und die Standardabweichung werden dann dargestellt. Die Werte liegen zwischen 0 und 1. Wobei 1 hundert Prozent Übereinstimmung entspricht. Berechnung des Koeffizienten:

$(2 \cdot \text{Übereinstimmung zwischen Tester A und B}) / (\text{Testanzahl von A} + \text{B})$

Die Bewertung erfolgt analog zu den im Methodenteil angegebenen Werten des Fleiss' Kappa (s.S.20).

Exemplarische Ergebnisse mit dem Holsti-Koeffizienten mit drei und zwei Antwort-Kategorien:

Exemplarisch werden hier die Ergebnisse durch die Berechnung des Holsti-Koeffizienten für das Ergebnis von den Testenden 1 und 2 gezeigt:

Prämessung: $53 \cdot 2 / 252 = 0,42$

Postmessung: $43 \cdot 2 / 252 = 0,34$

Werden die Ergebnisse der Testenden 1 und 2 mit zwei anstatt mit drei Kategorien ausgewertet, ergibt das:

Prämessung: $169 \cdot 2 / 504 = 0,67$

Postmessung: $166 \cdot 2 / 504 = 0,66$