

Die Inter- Rater- Reliabilitätstestung des Vorlauftests im Stand und im Sitz

MASTER - THESIS

zur Erlangung des akademischen Grades

Master of Science

im Universitätslehrgang Osteopathie MSc

vorgelegt von

Tanja Flicker- Niederl, BSc.

11850402

Department für Gesundheitswissenschaften, Medizin und Forschung

an der Donau-Universität Krems

Betreuer 1: Andreas Brandl, MSc.

Betreuer 2: Raimund Engel, MSc.

Datum, 14.06.2024



EIDESSTATTLICHE ERKLÄRUNG

Ich, Tanja Flicker- Niederl, BSc., geboren am 03.03.1988 in Graz erkläre,

1. dass ich meine Master Thesis selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfen bedient habe,
2. dass ich meine Master Thesis bisher weder im In- noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe,
3. dass ich, falls die Master Thesis mein Unternehmen oder einen externen Kooperationspartner betrifft, meinen Arbeitgeber über Titel, Form und Inhalt der Master Thesis unterrichtet und sein Einverständnis eingeholt habe.

DANKSAGUNGEN / WIDMUNGEN

An dieser Stelle möchte ich mich bei all jenen Bedanken, die mich während meiner Ausbildung zur Osteopathin bis hin zur Entstehung dieser Master- Thesis unterstützt haben. Meine Familie und möchte ich dabei hervorheben. Sie haben mich stets motiviert weiter zu machen und durchzuhalten.

Ich danke meinen StudienkollegInnen der Gruppe „Oddi“ für die tolle Zeit, speziell in Klagenfurt.

Das wohl größte Geschenk in meinem Leben, die Geburt meines Sohnes, ließ das Verfassen der Arbeit erstmal hintenanstehen. Mein Mann gab und gibt mir aber zu jeder Zeit das Gefühl für mich und unseren Sohn da zu sein. Meine Arbeit in der Praxis und auch das Verfassen der These wäre also bereits zu einem früheren Zeitpunkt möglich gewesen. Vielen Dank Didi! Ergänzend möchte ich mich auch bei allen Babysittern, die während dem Verfassen der These, Gabriel ihre Zeit gewidmet haben, bedanken. Danke an die Omas und Opas, die Urlis und Katrin.

Ich möchte mich außerdem bei Alina Kampl, meiner Kollegin und Freundin bedanken, dass ich mich an ihre Zeitvorgaben hängen durfte. Alina ist eine der zielstrebigsten Menschen, ohne dabei egoistisch oder vernachlässigend zu sein, die ich kenne. Vielen Dank!

Vielen Dank an die Praxisgemeinschaft XUNDOS, im Besonderen Simon Kampl, für das Bereitstellen ihrer Räumlichkeiten.

Ohne euch, liebe Elsa und liebe Simone, wäre die Durchführung der Studie gar nicht erst möglich gewesen. Vielen Dank für die Zeit, die ihr mir als Testerinnen zur Verfügung gestellt habt.

Seit 20 Jahren meine Rechtschreib- und Grammatikunterstützung ist Martina. Bereits in der Oberstufe wäre ich ohne dich verloren gewesen. Vielen Dank für das Korrekturlesen sämtlicher Texte und Arbeiten von mir. Nach dieser Master- Thesis kommt jetzt mal länger nichts, versprochen! ☺

Liebe Mary, vielen Dank für dein Engagement.

Ein großes Dankeschön gilt auch Thomas und Carina, die mein absolutes Nicht-Können am Computer und sämtlichen dazugehörigen Programmen kompensiert haben.

Auch meinem Betreuer Andreas Brandl, möchte ich auf diesem Wege danken.

ABSTRACT

Inter- Rater- Reliabilitätstestung des Vorlaufftests im Stand und im Sitz bei Erwachsenen

Ziel: Der Vorlaufftest im Stand und im Sitz stellt eine häufig angewandte Methode zur Eingrenzung des Untersuchungsbereiches im Rahmen der globalen Untersuchung, von OsteopathInnen dar. Bisher wurde der Test im Zusammenhang mit Dysfunktionen des Iliosakralgelenkes untersucht und als wenig valide und reliabel bestätigt. Ziel dieser Arbeit ist es, Werte für die Inter- Rater- Reliabilität des Vorlaufftests im Stand und im Sitz zu erheben, um herauszufinden, wie hoch die Übereinstimmung zweier unabhängiger OsteopathInnen ist.

Methode: Bei der vorliegenden Studie handelt es sich um eine Inter- Rater- Reliabilitätsstudie. Der Vorlaufftest im Stand und im Sitz wurde an 47 symptomatischen und asymptomatischen Erwachsenen von zwei unabhängigen OsteopathInnen durchgeführt. Zusätzlich wird die subjektive Angabe der Testerinnen von sicher und unsicher im Rahmen der Confidence- Rating- Scale gefordert.

Ergebnisse: Aus den erhobenen Daten wurden die Cohens- Kappa- Koeffizienten für den Vorlaufftest im Stand und den Vorlaufftest im Sitz berechnet. Für den Vorlaufftest im Stand beträgt der Wert $\kappa = 0,36$. Für den Vorlaufftest im Sitz wurde ein Wert von $\kappa = 0,28$ berechnet. Es liegt bei beiden Tests eine mäßige Übereinstimmung vor. Es konnte sowohl beim Vorlaufftest im Stand als auch im Sitz nur eine geringe Korrelation (0,29 und 0,21) mit der CRS errechnet werden.

Konklusio: Anlässlich der berechneten Koeffizienten für den Vorlaufftest im Stand und im Sitz kann derzeit keine Empfehlung für den Einsatz der Testungen als ein Teil der globalen Untersuchung von OsteopathInnen gegeben werden. Die Studie muss hinsichtlich des Kappa- Paradoxon optimiert werden. Die Übereinstimmung der Ergebnisse des Vorlaufftests im Stand und im Sitz sind nicht von der Angabe sicher und unsicher abhängig.

Schlüsselwörter: Inter- Rater- Reliabilität, Vorlaufftest im Stand, Vorlaufftest im Sitz, Kappa- Paradoxon, Osteopathie

ABSTRACT

Inter- rater reliability of the standing and sitting forward flexion test in symptomatic and asymptomatic adults

Objective: The standing and sitting forward flexion test is a common method in the global examination of osteopathic patients. Currently, the validity and reliability of the forward flexion test for evaluating dysfunction of the lliosacral joint are not well- established. This study aims to assess the inter- rater reliability of the standing and sitting forward flexion test as a tool in the global examination, to determine the level of agreement between two independent osteopathic practitioners.

Methods: This study employed an inter- rater reliability design, where two independent osteopathic practitioners examined 47 symptomatic and asymptomatic adults using the standing and sitting forward flexion test. Each examiner also had to indicate their confidence in the test results using a rating scale of “confident” or “not confident”.

Results: Cohen’s Kappa coefficients were calculated for the standing and sitting forward flexion tests, resulting in $\kappa = 0.36$ for the standing forward flexion test and $\kappa = 0.28$ for the sitting flexion test, indicating fair agreement. There was a low correlation (0.29 and 0.21) between the test results and the confidence- rating- scale.

Conclusion: Based on the fair agreement observed in the standing and sitting forward flexion tests, neither can be recommended as a standalone test in the global examination of osteopathic practitioners. Further optimization of the study is needed to address the Kappa paradox. The results of the standing and sitting forward flexion test were not influenced by the confidence level of the raters.

Keywords: inter- rater reliability, standing forward flexion test, sitting forward flexion test, global examination, osteopathy

INHALTSVERZEICHNIS

1	Einleitung	1
2	Theorie	4
2.1	Gütekriterien von Messmethoden	4
2.1.1	Objektivität	4
2.1.2	Validität	4
2.1.3	Reliabilität	4
2.1.4	Praktikabilität	4
2.2	Messmethoden in der Reliabilität	5
2.2.1	Kappa- Koeffizient	5
2.3	Optimale Ausgangsbedingungen einer Reliabilitätsstudie	6
2.3.1	Die vier Perioden nach Patijn (2019)	6
2.3.2	„Golden Rules“ von Patijn (2019)	7
2.4	Confidence- Rating- Scale (CRS)	7
2.5	Grundlagen Palpation, Anatomie, Biomechanik	7
2.5.1	Palpation und Interpretation	7
2.5.2	Anatomie und Biomechanik	8
2.5.3	Tensegrity- Modell	9
2.5.4	Funktionelle Anatomie	9
2.6	Evidenced based practice und aktueller Forschungsstand	10
3	Forschungsfrage und Hypothesen	12
3.1	Forschungsfragen	12
3.2	Nullhypothesen	12
3.3	Alternativhypothesen	12
4	Methodik	13
4.1	Forschungsdesign	13
4.2	Der Untersuchungstest	13
4.3	Beurteilung	13
4.4	Die Untersucherinnen	14
4.5	Der Assistent	14
4.6	Stichprobenbeschreibung	14
4.6.1	Einschlusskriterien	15
4.6.2	Ausschlusskriterien	15
4.6.3	Stichprobengröße	15
4.6.4	Auswahl der ProbandInnen	16
4.6.5	Die ProbandInnen	16
4.7	Zielparameter	17
4.7.1	Primäre Zielparameter	17
4.7.2	Sekundäre Zielparameter	17

4.8	Der Probedurchlauf	17
4.9	Studienablauf	19
4.9.1	Verblindung	19
4.9.2	Randomisierung	20
4.9.3	Durchführung	20
4.10	Datenverarbeitung und Auswertung	21
5	Ergebnisse	22
5.1	Darstellung und Berechnung der Daten	22
5.2	Sozialdemographische Daten zu den ProbandInnen	22
5.3	Ergebnisse der Raterinnen	24
5.3.1	Ergebnisse Raterinnen in absoluten Zahlen Vorlaufetest im Stand	24
5.3.2	Ergebnisse Raterinnen in absoluten Zahlen Vorlaufetest im Sitz	25
5.4	Kappa Koeffizienten der Vorlaufetestungen	25
5.4.1	Kappa- Koeffizient Vorlaufetest im Stand gesamt	25
5.4.2	Explorative Analyse Vorlaufetest im Stand	26
5.4.3	Kappa- Koeffizient Vorlaufetest im Sitz gesamt	27
5.4.4	Explorative Analyse Vorlaufetest im Sitz	28
5.5	P_{index} Vorlaufetest Stand und Vorlaufetest Sitz	28
5.6	Sicherheit der Raterinnen	28
5.6.1	Sicherheit der Raterinnen Vorlaufetest im Stand	29
5.6.2	Sicherheit der Raterinnen Vorlaufetest im Sitz	30
5.7	Korrelation von Inter- Rater- Reliabilität und CRS	30
5.8	Überprüfung der Hypothesen	31
5.8.1	Überprüfung der Nullhypothesen	31
5.8.2	Überprüfung der Alternativhypothesen	31
6	Diskussion	33
6.1	Vergleich der Ergebnisse mit bestehender Literatur	33
6.2	Interpretation der Ergebnisse	34
6.3	Bedeutung der Ergebnisse im Kontext der Osteopathie	36
6.4	Kritische Reflexion des Studienpersonals	37
6.5	Kritische Reflexion der Stichprobe	38
6.6	Kritische Reflexion des Studienablaufes	39
7	Konklusio	40
7.1	Zusammenfassung	40
7.2	Schlussfolgerung	40
7.3	Ausblick	40
	Literaturverzeichnis	42
	Tabellenverzeichnis	46
	ABBILDUNGSVERZEICHNIS	47

ABKÜRZUNGSVERZEICHNIS.....	48
ANHANG A: LITERATURRECHERCHE.....	49
ANHANG B: PROBANDINNENAUFTRUF.....	50
ANHANG C: ZEITEINTEILUNG	51
ANHANG D: GESUNDHEITSFRAGEBOGEN PROBANDIN	52
ANHANG E: EINVERSTÄNDNISERKLÄRUNG PROBANDIN.....	53
ANHANG F: ERGEBNISBOGEN GRUPPE 1 BIS ZUR PAUSE	54
ANHANG G: ERGEBNISBOGEN GRUPPE 2 AB DER PAUSE	56
ANHANG H: VARIABLEN UND KODIERUNG FÜR MICROSOFT EXCEL	58
ANHANG I: KREUZTABELLENVORLAGE	60
ANHANG J: WEITERE BERECHNUNGEN	61
ANHANG K: KORRELATIONEN.....	62
ANHANG L: INFORMATIONSBLETT.....	63
ANHANG M: ENGLISCHE KURZFASSUNG	64

1 Einleitung

Stills Aussage "Find it, fix it and leave it alone" hat in der Osteopathie einen hohen Stellenwert. Laut Fossum (2010, S. 70) ist es zunächst erforderlich, die mechanischen und palpatorischen Fähigkeiten zu entwickeln, um das Problem zu identifizieren („find it“), was dem diagnostischen Prozess entspricht. Anschließend muss man den richtigen Ansatz, also modus operandi finden, um das Problem zu beheben („fix it“) und schließlich das Prinzip der geringsten Intervention („leave it alone“) sowie das Vertrauen in die Selbstregulation des Körpers anwenden.

Für die osteopathische Diagnose ist es wichtig, eine ausführliche Anamnese und eine aussagekräftige klinische Untersuchung durchzuführen. Fossum (2010, S. 70) geht sogar davon aus, dass 75% der Behandlungszeit für die Untersuchung beansprucht werden. Beschrieben wird die Anamnese, eine globale und lokale osteopathische Untersuchung und eine medizinische Untersuchung (Fossum, 2010, S. 71- 72). Croibier (2006, S. 14- 16) beschreibt einen ähnlichen Aufbau. Die allgemeine Anamnese, eine allgemeine klinische Untersuchung, die manuelle Übersichtsdiagnose, eine Differentialdiagnose, die spezifische und die ätiologische Diagnose bilden den Weg zur osteopathischen Diagnose. Ein systematischer, strukturierter und vollständiger Aufbau sind dabei unumgänglich (Croibier, 2006, S. 282).

Die klinische Untersuchung und somit die Auswahl an Tests bleibt den einzelnen OsteopathInnen vorbehalten (Fossum, 2010). Sie ist individuell, flexibel und veränderbar (Croibier, 2006, S. 62). Die globale osteopathische Untersuchung wird in mehreren Positionen durchgeführt, um ein Bild der Läsionsmuster in seiner Totalität zu erhalten (Fossum, 2010, S. 71). Bei der lokalen osteopathischen Untersuchung werden laut Fossum (2010, S. 72) die Läsionsparameter qualitativ und quantitativ bewertet. Der Zustand des Gewebes wird beurteilt und Parameter für die Korrektur werden aufgestellt (Fossum, 2010, S. 72).

In einem Beispiel für eine Untersuchungsroutine im Leitfaden Osteopathie wird eine Testpalette vorgestellt, welche neben anderen Testungen den Vorlauftest im Stand und im Sitz enthält, um ein Gesamtbild zu erhalten und das Untersuchungsgebiet einzugrenzen (Fossum, 2010, S. 80- 88).

An der Wiener Schule für Osteopathie werden diese Testverfahren für die globale und lokale Untersuchung von Beginn an gelehrt. Biberschick (2010) hat im Rahmen einer Arbeit zur Begriffsdefinition „Routineuntersuchung“ festgestellt, dass es eine Vielzahl von Tests im Untersuchungsablauf gibt. Neben aktiven Funktionstests, General Listening Tests, der Ganganalyse, Hip Drop, Gillet Test usw. gehören auch der Vorlauftest im Sitz und im Stand zur klinischen Untersuchung (Biberschick, 2010). Es zeigt sich in der Häufigkeitsverteilung von Biberschick (2010), dass zwei Tests, nämlich die Überprüfung der Atmung und der Vorlauftest

im Stand von 50% der Autoren (n=10) immer ausgeführt werden. Alle anderen Testungen werden laut Biberschick (2010) nur zwei bis vier Mal in seinen Interviews erwähnt.

In der Literatur wird der Vorlaufstest im Stand und im Sitz auch als „standing“ und „sitting forward flexion test“ oder „bent over test“ beschrieben. Der Test wird in der Osteopathie im Screeningverfahren bzw. in der globalen Untersuchung eingesetzt (Biberschick 2010; Croibier 2006, S. 220; Fossum, 2010, S. 87, S. 89). Über das Einsatzgebiet des Vorlaufstests waren sich alle befragten OsteopathInnen von Biberschick (2010) einig. Er konnte es unter dem Punkt „Ziele der Routineuntersuchung“ gut veranschaulichen (Biberschick, 2010).

Als Anreiz für die Durchführung und Verfassung der Masterthese wurde die fehlende Evidenz des Vorlaufstests im Sitz und im Stand gewählt. Es konnte in der Literatur nichts gefunden werden, was den Einsatz des Vorlaufstests berechtigt. Die Evidenz dafür soll mit dieser Arbeit geschaffen werden.

Nicht nur im Diagnoseleitfaden für Osteopathie von Magga Corts (2020, S. 26, S. 37, S. 38) wird der Vorlaufstest im Rahmen der globalen Untersuchung genannt und beschrieben. In der Grundausbildung an der Wiener Schule für Osteopathie kommt der Vorlaufstest bereits sehr früh zur Anwendung. Er wird zu Beginn gelehrt, um den Test in der globalen Untersuchung zu integrieren.

Der Test zeigt Asymmetrien bzw. Einschränkungen der mechanischen Körperfunktionen, welche wahrscheinlich durch strukturelle und neuromuskuläre Anpassungen verursacht werden (Fossum, 2010, S. 87). So wird laut Beschreibung im Leitfaden für Osteopathie zum Beispiel ein positiver Vorlauf, welcher nur im Stand aber nicht im Sitz auftritt, als aufsteigendes Problem bezeichnet (Fossum, 2010, S. 89). Im Vergleich dazu kann bei einem positiven Vorlauf, welcher nur im Sitz festgestellt wird, die untere Extremität ausgeschlossen werden (Fossum, 2010, S. 90). Ein positiver Vorlauf im Sitz und im Stand deutet auf ein absteigendes Problem hin (Fossum, 2010, S. 89). Laut Fossum (2010, S. 87) ist es außerdem interessant zu beobachten, ob sich das Testergebnis vor und nach der Behandlung verändert.

Es wird darauf hingewiesen, dass es sich in folgender Arbeit nicht um eine spezifische Testung für das Iliosakralgelenk handelt. Es wurde bereits 1996 und 1999 festgestellt sowie 2021 durch Ribeiro et al. bestätigt, dass ein positiver Vorlaufstest kein ausreichender Hinweis auf eine Funktionsstörung im ISG ist (Egan et al., 1996; Ribeiro et al., 2021; Vincent-Smith & Gibbons, 1999).

Der Nachweis von guten und aussagekräftigen Tests und deren Ergebnisse ist in der Osteopathie von großer Bedeutung. Die Osteopathie muss evidenzbasiert arbeiten um an Seriosität zu gewinnen. Nach wie vor wird die Osteopathie im österreichischen Gesundheitssystem nicht richtig anerkannt. Die Rückverrechnung mit den einzelnen Kassen ist schwierig bis unmöglich.

In der Durchführung folgender Inter- Rater- Reliabilitätsstudie hat sich die Autorin und Studienleiterin an die Vorgaben der Guideline von Patjin (2019) gehalten. Der Einsatz des Vorlauftests im Stand und im Sitz in der globalen Untersuchung sollen durch eine gute Reliabilität seine Berechtigung, speziell in Fachkreisen, bekommen. Die Zuverlässigkeit zwischen RaterInnen ist für den Informationsaustausch unter Fachleuten erforderlich.

2 Theorie

2.1 Gütekriterien von Messmethoden

Für die Evidenz von Tests und Assessments gibt es mehrere Gütekriterien (Scherfer & Bossmann, 2011, S. 242). Scherfer und Bossmann (2011, S. 242) definieren zwei Hauptgütekriterien, die Reliabilität und die Validität. Weiß (2013) beschreibt noch ein drittes Hauptgütekriterium, die Objektivität.

2.1.1 Objektivität

Die Objektivität beschreibt die Unabhängigkeit eines Messergebnisses von der Person, welche die Daten erhebt. Die Objektivität eines Tests gilt als Voraussetzung für die anderen Gütekriterien (Hussy et al., 2013, S. 23- 24).

2.1.2 Validität

Die Validität stellt fest, ob der Test misst was er messen soll. Es bezieht sich auf die Richtigkeit des Messinstrumentes (Scherfer & Bossmann, 2011, S. 242- 243).

2.1.3 Reliabilität

Die Zuverlässigkeit eines Tests, die sogenannte Reliabilität, beschreibt inwieweit Messwiederholungen, die unter denselben Bedingungen stattfinden, tatsächlich dieselben Ergebnisse liefern (Scherfer & Bossmann, 2011, S. 247; Weiß, 2013). Unterschieden werden die Intra- Rater- Reliabilität und die Inter- Rater- Reliabilität. Als „Rater“ werden die TesterInnen bezeichnet. Bei der Intra-Rater- Reliabilität wird untersucht, ob ein und der/dieselbe TesterIn bei mehrmaliger Durchführung des Tests, unter identen Bedingungen, zu den gleichen Ergebnissen kommt (Scherfer & Bossman 2011, S. 250).

Für die vorliegende Masterarbeit ist die Inter-Rater- Reliabilität von Relevanz. Es wird untersucht, ob zwei oder mehr TesterInnen, unter den gleichen Bedingungen, bei einer Gruppe von ProbandInnen zu einem identen Ergebnis kommen. Eine Messung kann reliabel und trotzdem nicht valide sein (Scherfer & Bossmann, 2011, S. 244).

2.1.4 Praktikabilität

Für die Vollständigkeit der Gütekriterien wird noch das Kriterium der Praktikabilität erwähnt. Scherfer und Bossmann (2011, S. 246) beschreiben unter diesem Punkt die ethische und ökonomische Vertretbarkeit eines Tests. Messinstrumente, Assessments und Tests sollen mit geringem zeitlichen und apparativen Aufwand sowie möglichst wenig gesundheitlichen Belastungen für TherapeutInnen und PatientInnen durchgeführt werden können (Scherfer & Bossmann, 2011, S. 246).

2.2 Messmethoden in der Reliabilität

Es gibt unterschiedliche Messmethoden welche in der Reliabilitätsforschung verwendet werden, um die Konsistenz und Zuverlässigkeit von Messungen und Tests zu überprüfen und sicherzustellen (Scherfer & Bossmann, 2011, S. 249- 252). Bei der Inter- Rater- Reliabilität führen mehrere Beurteiler die gleiche Messung durch. Die Ergebnisse und die Konsistenz der Messung werden verglichen und überprüft. Es werden statistische Methoden, wie die Berechnung von Korrelationen und die Verwendung von Koeffizienten eingesetzt. Zum Beispiel wird der Cronbach`s Alpha oder der Kappa- Koeffizient verwendet, um die Reliabilität der durchgeführten Messung zu bewerten. Für die Bewertung einer Reliabilitätsstudie mit nominalem Messniveau wird die Reliabilität mit dem Kappa- Koeffizienten angegeben (Patijn, 2019; Scherfer & Bossmann, 2011, S. 251- 252).

2.2.1 Kappa- Koeffizient

Der Kappa- Koeffizient nach Cohen ist ein statistisches Maß für die Übereinstimmung zwischen zwei oder mehreren TesterInnen. Der Koeffizient berücksichtigt dabei den Anteil an Übereinstimmungen, welche über den Zufall hinausgehen (Scherfer & Bossmann, 2011, S. 251- 252). Vereinfacht bedeutet das, dass eine 100%ige Übereinstimmung zweier oder mehrerer UntersucherInnen einen Wert von 1 ergeben würde. Bei einem Wert von 0 entspricht die Anzahl der Übereinstimmung der Zufallserwartung (Scherfer & Bossmann, 2011, S. 251- 252). Ein Wert von -1 gibt eine vollständig negative, also gegensätzliche Übereinstimmung an (Landis & Koch, 1977). Für die Berechnung des Kappa- Koeffizienten wird die genaue Anzahl der Übereinstimmungen, die Anzahl der Nicht- Übereinstimmungen und die Anzahl der zufälligen Übereinstimmungen benötigt. Abhängig von der Art der Daten und der Anzahl der TesterInnen ergibt sich die Formel zur Berechnung des Kappa- Koeffizienten. Ein hoher Kappa- Koeffizient deutet auf eine gute Inter- Rater- Reliabilität hin. Eine gute Orientierung und Veranschaulichung der Interpretation des Kappa- Wertes bietet die Tabelle von Landis und Koch (1977) (siehe Tabelle 1). Fjellner et al. (1999) und McHugh (2012) nennen Werte von mindestens 0,61 als Indikator einer guten Inter- Rater- Reliabilität. Ein $\kappa = 0,41$ wäre zumindest akzeptabel.

TABELLE 1 KLASSIFIKATION DER COHENS- KAPPA- KOEFFIZIENTEN ADAPTIERT NACH LANDIS UND KOCH (1977) ÜBERSETZT VON DER AUTORIN

Cohens Kappa Koeffizient	Grad der Übereinstimmung
<0.00	Poor (mangelhaft)
0.00-0.20	Slight (gering)
0.21-0.40	Fair (mäßig, ausreichend)
0.41-0.60	Moderate (mittelmäßig)
0.61-0.80	Substantial (beachtlich)
0.81-1.00	Almost perfect (fast perfekt)

2.3 Optimale Ausgangsbedingungen einer Reliabilitätsstudie

Für die Durchführung einer erfolgreichen Reliabilitätsstudie hat Patijn (2019) eine Guideline aufgestellt. Die Einhaltung aller Punkte ist sehr wichtig, um optimale Voraussetzungen für die Studie zu schaffen, die Reproduzierbarkeit zu garantieren, Vergleichbarkeit zu ermöglichen und vor allem eine gute Evidenz zu erreichen. Patijn (2019) unterteilt in vier Perioden, welche untenstehend genauer beschrieben werden.

2.3.1 Die vier Perioden nach Patijn (2019)

1. Logistische Periode: Alle an der Studie teilnehmenden Personen werden festgelegt. Die Rahmenbedingungen werden besprochen. Im Optimalfall werden zwei TesterInnen aus derselben Einrichtung gefunden, da dies die Trainings- und Studienphase vereinfacht. Trainingsphasen mit zehn ProbandInnen werden vereinbart um die Abläufe zu vereinheitlichen.

2. Trainingsperiode: Für eine erfolgreiche Reliabilitätsstudie ist die Trainingsphase essentiell. Es muss eine mündliche Vereinbarung zwischen den TesterInnen über alle Details bezüglich Durchführung und Bewertung des Tests geben. Im Optimalfall gibt es eine Zusammenkunft, bei der alles an einer Person getestet wird. Es wird verglichen und eine einheitliche Ausführung bestimmt. Für den Probelauf werden zehn Personen benötigt. Letztendlich stehen die Reproduzierbarkeit und das Ergebnis im Fokus. Der Einsatz in der täglichen Praxisroutine wird empfohlen, um die Skills zu schulen. Es sollte immer nur ein Test zu einem Zeitpunkt überprüft werden. Die TesterInnen sollen sich über die Hypothese zum Test der aktuellen Studie im Klaren sein und vergangene bisher unbestätigte Hypothesen ignorieren. Bei der Interpretation des Ergebnisses sollen die TesterInnen ebenfalls etwaige Veränderungen aus deren täglichen Routine vernachlässigen und einheitliche Interpretationen einsetzen. Für die Verblindung ist es essentiell, dass weder die TesterInnen untereinander noch der/die TesterIn mit dem/der ProbandIn während des Studienablaufes kommunizieren. Natürlich dürfen TeilnehmerInnen über den Ablauf informiert werden. Bereits beim Probelauf soll alles möglichst genau, wie bei der eigentlichen Studie, durchgeführt werden. Sollte bei der Auswertung der Ergebnisse in der Trainingsphase ein P_0 kleiner als 0.80 rauskommen, wird eine weitere Trainingsphase eingeleitet. Zuvor sollen mögliche Fehlerquellen diskutiert und behoben werden.

3. Periode der Studie: Die Studie wird von den gleichen TesterInnen, die auch den Probelauf durchgeführt haben, ausgeführt. Die Größe der Stichprobe enthält 40 Personen. Wie in der Trainingsperiode beschrieben, ist die Verblindung unbedingt einzuhalten. Die Evaluationsbögen mit den Ergebnissen sollen direkt nach Durchführung des Tests von einer unabhängigen Person eingesammelt werden.

Statistische Periode: Die abgesammelten Ergebnisse werden in die vorgesehene Kreuztabelle eingetragen. Berechnet werden müssen P_0 , P_{index} und der Kappa- Koeffizient.

4. Publikationsperiode: Für die Publikation ist es wichtig, den/die LeserIn mit adäquaten Informationen darüber zu versorgen, auf welchen Fakten ihre Konklusio beruht.

2.3.2 „Golden Rules“ von Patijn (2019)

1. Eine Person ist für die Aufsicht und die Führung des Logbooks der Studie zuständig.
2. Führe ein Logbook.
3. Es muss eine ausgiebige Trainingsphase geben. Die Rater müssen sich in allen Belangen des durchzuführenden Tests und der Interpretation des Ergebnisses einig sein.
4. Eine standardisierte Pilotstudie muss durchgeführt werden.
5. Es muss ein P_0 von 0,8 erreicht werden, sonst wird die Trainingsphase verlängert.
6. Es wird nur ein Test zur gleichen Zeit in der Reliabilitätsstudie überprüft.
7. Rohdaten werden in Kreuztabellen veranschaulicht. P_0 , P_{index} und der Kappa- Koeffizient werden gemeinsam dargestellt.

2.4 Confidence- Rating- Scale (CRS)

Cooperstein et al. (2010) haben in ihrer Inter- Rater- Reliabilitätsstudie festgestellt, dass ein hoher Überzeugungsscore der Rater eine bessere Übereinstimmung bei der Testung ergab. In der vorliegenden Studie soll mittels Confidence- Rating- Scale (CRS) ebenfalls ermittelt werden, ob sich die TesterInnen bei der Angabe ihrer Ergebnisse sicher oder unsicher sind. Es ist eine reine Selbsteinschätzung die die TesterInnen, ergänzend zum eigentlichen Ergebnis ihrer Testdurchführung, in den Bewertungsbogen eintragen.

2.5 Grundlagen Palpation, Anatomie, Biomechanik,

Im nachfolgenden Abschnitt wird die Bedeutung einer präzisen Palpation sowie die korrekte Interpretation palpiert Strukturen erläutert. Zudem werden die theoretischen Grundlagen hinsichtlich der Anatomie und Biomechanik erläutert, die für den Vorlauftest sowohl im Stand als auch im Sitz ausschlaggebend sind.

2.5.1 Palpation und Interpretation

Der Vorlauftest im Stand und im Sitz sind abhängig von der richtigen Palpation der Spinae iliacae posteriores superiores rechts und links. Diese Landmarken sind ausschlaggebend für die Interpretation der Testergebnisse. Für manuell arbeitende TherapeutInnen ist die Palpation essenziell und Voraussetzung für eine adäquate Behandlung (Och, 2010, S. 72). Ein geschulter Tastsinn und sehr gute Anatomiekenntnisse sind dafür erforderlich (Och, 2010, S. 72). Wahrgenommen werden können bei der Palpation Temperaturunterschiede, Feuchtigkeitsverhältnisse der Oberfläche, die Oberflächenbeschaffenheit, die Beschaffenheit

von tiefer liegendem Gewebe, die Gewebedichte und Gewebeelastizität sowie feinste Bewegungen, Beweglichkeit und Energien in den Geweben. Och (2010, S. 72) beschreibt außerdem die Übermittlung von Informationen über Funktionsstörungen und Strukturveränderungen. Bei der Palpation von Knochen bzw. knöchernen Landmarken palpiert man „vom Tal zum Berg“. Man tastet sich von den Weichteilen zum Knochen hin (Och, 2010, S. 78). Bei länglichen Strukturen wie Muskeln, Sehnen, Ligamenten, Röhrenknochen und Knochenkanten wird sowohl quer zur als auch entlang der Struktur palpiert (Och, 2010, S. 78). Laut Och (2010, S. 79) unterliegt die Palpation der Subjektivität der behandelnden Personen und der PatientInnen. Liem (2014) geht sogar soweit dass er sagt, die osteopathische Zuverlässigkeit kann mit der Wahrnehmung und Interpretation der Palpation zusammenhängen, die von früheren Erfahrungen, der Art der zu sammelnden Informationen gewohnheitsmäßigen und kontextbezogenen Einflüssen sowie kulturellen und sozialen Prägungen beeinflusst wird.

Von Relevanz für diese Arbeit ist die Palpation der Crista iliaca rechts und links sowie die Spina iliaca anterior superior und posterior beidseits. Die Hüftextensoren, der M. quadratus lumborum und der M. piriformis, müssen ebenfalls palpiert werden. Erschwerend kommt hinzu, dass die Palpation während der Durchführung einer Bewegung aufrechterhalten werden muss.

2.5.2 Anatomie und Biomechanik

Die Beschreibung der Anatomie folgt aus dem Anatomieatlas Sobotta von Paulsen und Waschke (2010). Das Os Ilium ist dorsal durch das Articulatio sacroiliaca mit dem Os Sacrum verbunden (Paulsen & Waschke, 2010, S. 250). Die verdickte kraniale Kante wird Crista iliaca genannt und verbindet zwei markante Knochenvorsprünge, die ventral gelegene Spina iliaca anterior superior und die dorsal gelegene Spina iliaca posterior superior (SIPS) (Paulsen & Waschke, 2010, S. 252). Die Crista iliaca und die SIPS können gut von außen palpiert werden. Kaudal von den SIPS befinden sich die Spinae iliaca posterior inferior (SIPI) (Paulsen & Waschke, 2010, S. 252).

Generell ist das Os Ilium an zwei Gelenken beteiligt. Zum einen durch das Acetabulum am Hüftgelenk, zum anderen über die dorsale Gelenkfläche der Facies auricularis am Iliosakralgelenk (Paulsen & Waschke, 2010, S. 250, S. 253). Die Gelenkflächen von Os Ilium und Os Sacrum sind miteinander verzahnt und über kräftige Ligamente verbunden, den Ligamenti sacroiliaca anteriora, interossea und posteriora (Paulsen & Waschke, 2010, S. 266 & S. 268). Das Ligamentum iliolumbale verbindet die Crista iliaca mit dem fünften Lendenwirbel (Paulsen & Waschke, 2010, S. 266). Die Bewegungsmöglichkeit des Os Iliums gegenüber dem Os Sacrum besteht in einer Rotation nach anterior oder nach posterior (Vincent, 2010, S. 397). Bei einer Flexion in der Wirbelsäule sollen beide SIPS gleichzeitig beginnen und aufhören sich nach superior zu bewegen (Vincent, 2010, S. 398). Da laut Vincent (2010, S. 398) die Hüftextensoren, der Musculus quadratus lumborum und der Musculus

piriformis, Einfluss auf den Vorlauftest im Sitz und im Stand haben, werden diese nachfolgend genauer beschrieben. Die Musculi gluteus maximus, medius und minimus haben ihren Ursprung an der Facies glutea und arbeiten als Extensoren in der Hüfte (Paulsen & Waschke, 2010, S. 307). Der Musculus quadratus lumborum entspringt an der Crista iliaca und hat seinen Ansatz an der 12. Rippe. Als Funktion ergibt sich die Streckung in der Lendenwirbelsäule und die ipsilaterale Lateralflexion (Paulsen & Waschke, 2010, S. 93). Der Ursprung des Musculus piriformis ist an der Facies pelvica des Kreuzbeins und der Ansatz an der Spitze des Trochanter major femoris (Paulsen & Waschke, 2010, S. 306). Der Musculus piriformis arbeitet hauptsächlich als Außenrotator im Hüftgelenk. Bei Flexion im Hüftgelenk verändert der Muskel seine Funktion und macht eine Innenrotation. Der Musculus piriformis unterstützt außerdem die Abduktion und Retroversion des Hüftgelenks und stabilisiert dieses.

2.5.3 Tensegrity- Modell

Fossum (2010, S. 87) geht davon aus, dass Asymmetrien bei einem auffälligen Vorlauftest auf strukturelle (ligamentäre, fasziale und muskulöse) sowie neuromuskuläre Adaptationen zurückzuführen sind. Die Einheit des Körpers und die Wechselwirkung zwischen Struktur und Funktion bilden in der Osteopathie die Grundbausteine der therapeutischen Philosophie (Ciranna- Raab, 2010, S. 67). Das Tensegrity- Modell wird häufig als Erklärungsmodell in der Osteopathie herangezogen. Die Worte „Tension“ (Spannung) und „Integrity“ (Ganzheit) beschreiben das Grundprinzip des Modells ganz gut (Ciranna- Raab, 2010, S. 65). Das Tensegrity- Modell besteht als Konstrukt aus starren diskontinuierlichen Anteilen, welche durch kontinuierliche Spannungselemente verbunden sind (Ciranna- Raab, 2010, S. 65). Viele Konzepte der Biomechanik wurden durch das Tensegrity- Modell verändert. Es stellt die Spannungsverhältnisse im Körper in den Vordergrund (Ciranna- Raab, 2010, S. 67). Schon vor mehr als 100 Jahren erkannte Still die Wichtigkeit von Faszien, die laut seiner Aussage als Kommunikationsnetz von Krankheit und Gesundheit arbeiten und durch das Tensegrity- Modell bestätigt wird (Ciranna- Raab, 2010, S. 68). Solche komplexen Wechselwirkungen machen die Wichtigkeit einer guten globale Untersuchung mit verlässlichen Testungen zur Eingrenzung deutlich.

2.5.4 Funktionelle Anatomie

Die funktionelle Anatomie beschäftigt sich mit der Struktur des menschlichen Körpers und dessen Verhalten in Bewegung. Jeder Muskel hat eine individuelle Aufgabe, arbeitet aber auch zum Beispiel innerhalb des Fasziennetzes über funktionell integrierte, den gesamten Körper umspannende Einheiten hinweg (Myers, 2004, S. 1). Myers (2004, S. 1) spricht vom Verständnis von Kompensations- und Spannungsmustern, welche den ganzen Körper umspannen. Klinisch bedeutet das, dass ein schmerzhaftes Problem in einer Körperregion von einer symptomfreien Region in einiger Entfernung ausgehen kann (Myers, 2004, S. 1). Myers

(2004, S. 98) hat zum Beispiel beschrieben, dass es bei einer Dysbalance zwischen den sogenannten Zuglinien, der Frontalen Linie und der Rückenlinie, zu einem anterioren sowie posterioren Tilt im Becken kommen kann.

Im Buch von Meert (2017, S. 5) wird der Körper als eine funktionierende Einheit beschrieben, bei der nicht die Summe der Einzelteile, sondern das Zusammenspiel und die Wechselbeziehungen entscheidend für die Gesundheit sind. Eine Störung des Gleichgewichts kann daher zu vielfältigen und komplexen Läsionen führen. Der Körper strebt kontinuierlich nach einem Gleichgewicht, was zur Entstehung von sekundären Dysfunktionen führen kann (Meert, 2017, S. 8). Meert (2017, S. 17) betont, dass das Becken als ein großer Stoßdämpfer zwischen dem Rumpf und den unteren Extremitäten fungiert, wodurch Belastungsstress durch auf- und absteigende Kräfte entsteht.

Die Notwendigkeit einer aussagekräftigen und zielführenden globalen Untersuchung wird von Meert (2017, S. 6) unterstrichen.

2.6 Evidenced based practice und aktueller Forschungsstand

„Evidence- based- practice“ leitet sich aus der „evidence- based- medicine“ ab und hat folgendes Ziel zu Grunde: sie möchte osteopathisch klinische Entscheidungen für die Testauswahl und auch die Auswahl der Intervention optimieren (Bordoni, 2019). Wissenslücken zu identifizieren, sie als Grundlage für ein praxisrelevantes und problemorientiertes Lernen zu sehen, ist in der Evidence- based- practice (EBP) wichtig (Scherfer & Bossmann, 2011, S. 34). Ein weiterer wesentlicher Bestandteil der EBP ist die kritische Wahrnehmung von Forschungsergebnissen (Scherfer & Bossmann, 2011, S. 14). Drei wichtige Auswahlkriterien, um den Evidence- based- practice Ansatz zu erlangen, werden von Scherfer und Bossmann (2011, S. 13- 14) beschrieben. Zum einen die externe und interne Evidenz, zum anderen die individuelle Patientensituation (Scherfer & Bossmann, 2011, S. 15). Zusammenfassend kann gesagt werden, dass die Anwendung der EBP einen hohen ethischen und professionellen Standard ausdrückt, dass es eine Bescheidenheit hinsichtlich des eigenen Wissensstandes gibt und auch die Bereitschaft, diesen im Dienste der PatientInnen zu verbessern (Scherfer & Bossmann, 2011, S. 29).

Bei der Suche nach bereits vorhandener Evidenz zu folgenden Forschungsfragen, konnte eine Studie von Ribeiro et al. (2021) gefunden werden. In dieser Studie wird die Validität und Reliabilität des Vorlaufstests, allerdings in Bezug auf eine Dysfunktion des Iliosakralgelenks, untersucht. Die Ergebnisse sind unzureichend.

Nolet et al. (2021) haben in ihrem systematischen Review von 2023 Artikeln, in der der Nutzen von Palpation bei PatientInnen mit unterem Rückenschmerzen untersucht wurde, alle palpierenden Maßnahmen in Frage gestellt. Die Gelenkspalpation während einer Bewegung wird als nicht reliabel bei PatientInnen mit „low back pain“ bewertet.

Fryer et al. (2005) untersuchten in ihrer Studie wie sich ein „consensus“- Training einer Palpation auf die Inter- und Intra- Reliabilität auswirkt. Der Vorlaufstest im Sitz wurde zur Überprüfung herangezogen. Die Ergebnisse zeigen sich unzureichend. Die Gruppe mit dem Training hat einen Kappa- Wert von 0,14, die Gruppe ohne vorangegangenes Training einen Kappa- Wert von 0,07 (Fryer et al. 2005).

Vincent- Smith und Gibbons (1999) haben eine Studie mit neun asymptomatischen ProbandInnen durchgeführt, die keine guten Ergebnisse geliefert hat. Es stand auch hier die Dysfunktion des Iliosakralgelenks im Fokus.

Boddenberg (2013) untersuchte in einer kontrollierten Studie, ob ein positiver Vorlaufstest durch das Einlegen von Zahnwatterollen zwischen den Zähnen im Ober- und Unterkiefer, bei Patienten mit einer Kiefergelenkspathologie verändert werden kann. Dies ist im Hinblick auf vorliegende Fragestellung interessant, da es auf die Auswirkung von entfernten Veränderungen auf das Becken und deren mögliche Erfassung mittels Vorlaufstest hinweist. In der Konklusio beschreibt der Autor den modifizierten Vorlaufstest als geeignete Methode.

Es gibt keine Literatur in Bezug auf die Reliabilität des Vorlaufstests als globales Untersuchungsinstrument. Mit einer Inter-Rater-Reliabilitätsstudie soll der Vorlaufstest auf Zuverlässigkeit überprüft werden, diese Wissenslücke geschlossen und für weitere Studie angeregt werden.

3 Forschungsfrage und Hypothesen

In diesem Teil der Master- Thesis werden die Forschungsfragen, die Null- und die Alternativhypothesen vorgestellt.

3.1 Forschungsfragen

Wie hoch ist die Übereinstimmung zweier OsteopathInnen gleichen Ausbildungsstandards bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand?

Wie hoch ist die Übereinstimmung zweier OsteopathInnen gleichen Ausbildungsstandards bei der Beurteilung der Auffälligkeit des Vorlauftests im Sitz?

3.2 Nullhypothesen

Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand ist kleiner als der Cohens Kappa Koeffizient von 0,61 ($\kappa < 0,61$) und liegt unter dem beachtlichen Bereich (Landis & Koch, 1977).

Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand ist kleiner als der Cohens Kappa Koeffizient von 0,61 ($\kappa < 0,61$) und liegt unter dem beachtlichen Bereich (Landis & Koch, 1977).

3.3 Alternativhypothesen

Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand ist gleich oder größer als der Cohens Kappa Koeffizient von 0,61 ($\kappa \geq 0,61$) und liegt im oder über dem beachtlichen Bereich (Landis & Koch, 1977).

Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Sitz ist gleich oder größer als der Cohens Kappa Koeffizient von 0,61 ($\kappa \geq 0,61$) und liegt im oder über dem beachtlichen Bereich (Landis & Koch, 1977).

4 Methodik

Im folgenden Kapitel beschreibt die Autorin das gewählte Forschungsdesign, die zwei Untersuchungstests und deren Beurteilung, das Personal welches für die Studie benötigt wurde, die für diese Inter- Rater- Reliabilitätsstudie gewählte Stichprobe mittels sozialdemographischer Angaben der ProbandInnen, den Probedurchlauf mit den dazugehörigen Ergebnissen, den gesamten Studienablauf inklusive des Verblindungs- und Randomisierungsvorganges und die Form der Datenverarbeitung und Auswertung. Die Art der Literaturrecherche wird in Anhang A erläutert.

4.1 Forschungsdesign

Bei der hier beschriebenen Studie handelt es sich um eine Inter- Rater- Reliabilitätsstudie. Die Übereinstimmung zweier TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand und des Vorlauftests im Sitz wird überprüft.

4.2 Der Untersuchungstest

Bei dem in dieser Inter- Rater- Reliabilitätsstudie untersuchten Test, handelt es sich um einen Palpationstest, welcher im Verlauf einer Bewegung ausgeführt wird. Für die korrekte Ausführung ist die Palpation der Spina iliaca posterior superior rechts und links Voraussetzung. Die Ausführung des Vorlauftests im Stand und im Sitz hält sich ohne Abweichungen an die Vorgaben, die im Leitfaden für Osteopathie beschrieben werden (Liem & Dobler, 2010, S. 397- 400).

Beim Vorlauftest aus dem Stand steht der/die ProbandIn hüftbreit vor dem/der TesterIn. Die Augen des/der TesterIn befinden sich auf Höhe der Spinae iliaca posterior superior. Mit beiden Daumen werden die Spinae iliaca posterior superior von dem/der TesterIn palpirt. Der/Die ProbandIn wird aufgefordert den Oberkörper langsam nach vorne Richtung Boden zu beugen („Hände in Richtung der Füße bewegen“).

Beim Vorlauftest im Sitz sitzt der/die ProbandIn vor dem/der TesterIn. Der/Die ProbandIn hat die Füße hüftbreit abgestellt und wird aufgefordert, den Oberkörper langsam nach vorne zu beugen („Hände zwischen den Knien in Richtung der Füße bewegen“). Wie auch bei der Testung im Stand, befinden sich die Augen des/der TesterIn auf Höhe der Spinae iliaca posterior superior. Die Daumen befinden sich wieder auf den genannten Knochenvorsprüngen, den Spinae iliaca posterior superior.

4.3 Beurteilung

Der/Die TesterIn achtet auf die Bewegung der Spinae Iliaca posterior superior. Das Ergebnis ist unauffällig, wenn sich beide Spinae iliaca posterior superior gleichzeitig

während der Flexionsbewegung beginnen und aufhören, sich nach superior zu bewegen. Das Ergebnis ist auffällig, wenn sich eine Seite deutlich weiter nach superior mitbewegt. Die Beurteilung gilt für den Vorlauftest im Stand und für den Vorlauftest im Sitz.

4.4 Die Untersucherinnen

Patijn (2019) gibt in seinen Richtlinien an, dass es keinen Grund für mehr als zwei TesterInnen bei Inter- Rater- Reliabilitätsstudien gibt. Es wurden daher zwei Osteopathinnen für die Durchführung der Testungen rekrutiert. Voraussetzung war, dass beide TesterInnen ihre Ausbildung zur OsteopathIn an der Wiener Schule für Osteopathie positiv abgeschlossen haben. Der Grundberuf als PhysiotherapeutIn war ebenfalls ein Einschlusskriterium. Beide Osteopathinnen haben den Masterlehrgang „Osteopathie MSc.“ in Kooperation mit der Donau-Universität Krems belegt. Ihre Teilnahme an der Studie war freiwillig.

Testerin A ist 33 Jahre alt und arbeitet seit 2014 als Physiotherapeutin. 2017 begann sie ihre Ausbildung an der Wiener Schule für Osteopathie. Die Klinikprüfung wurde im September 2022 abgelegt. Der Vorlauftest wird im Stand und im Sitz in der „Routineuntersuchung“ eingesetzt und kommt somit im Praxisalltag regelmäßig zum Einsatz.

Testerin B ist 50 Jahre alt und arbeitet seit 1998 als Physiotherapeutin. Die Klinikprüfung wurde im Februar 2021 abgelegt. Wie auch Testerin A ist Testerin B freiberuflich in eigener Praxis tätig und setzt den Vorlauftest im Stand und im Sitz regelmäßig bei der „Routineuntersuchung“ oder auch als Wiederbefundungsparameter in einer Therapieeinheit ein.

4.5 Der Assistent

Die Studienleiterin wurde am Tag der Studiendurchführung von einem Assistenten unterstützt. Die Aufgabe des Assistenten war die Bereitstellung der vorbereiteten Fragebögen an die ProbandInnen und die folgende Überprüfung, ob alle Felder im Daten- und Fragebogen ausgefüllt und das Blatt unterschrieben sind. Weiter sorgte er für die Bereitstellung von Getränken und die Beantwortung aller Studien irrelevanten Fragen (z. Bsp.: Wo müssen die Schuhe ausgezogen werden? Wo befindet sich die Toilette? Wo kann man parken? usw.). Bevor die ProbandInnen an die Studienleiterin weitergeleitet wurden, entschied sich durch einen Münzwurf des Assistenten die Zuteilung zu Raum blau oder Raum rot. Der Assistent konnte einen reibungslosen Ablauf sicherstellen und die Studienleiterin vor unnötiger Ablenkung bewahren.

4.6 Stichprobenbeschreibung

Auf die Stichprobe der Studie, mit ihren Einschluss- und Ausschlusskriterien, der Stichprobengröße und die Akquirierung der ProbandInnen wird unter diesem Kapitel genauer eingegangen.

4.6.1 Einschlusskriterien

Da es sich beim Vorlauftest im osteopathischen Zusammenhang um eine globale und nicht um eine spezifische Untersuchung handelt, wurden alle Personen, jeglichen Geschlechts, ab einem Alter von 18 Jahren in die Studie eingeschlossen. Als Voraussetzung wurde die selbstständige Ausführung der vorgegebenen Ausgangsstellungen angegeben. Der Stand und der Sitz mussten ohne Hilfsmittel und Hilfspersonen eingenommen werden können. Des Weiteren mussten die Testpersonen aus freiem Willen an der Studie teilnehmen. Es wurden symptomatische und asymptomatische Personen zur Studie zugelassen. Die Einschlusskriterien wurden telefonisch abgefragt und am Tag der Durchführung der Studie Vorort überprüft. Durch eine Unterschrift haben die ProbandInnen der freiwilligen Teilnahme zugestimmt und bestätigt, dass die Angaben wahrheitsgemäß sind und sämtliche Daten zur Weiterverarbeitung genutzt werden dürfen.

4.6.2 Ausschlusskriterien

Personen bei denen auf Grund von Verletzungen (z. Bsp. offenen Wunden) keine Palpation der Spine iliace posterior superior möglich gewesen wäre, wären aus der Studie ausgeschlossen worden. Telefonisch wurde bereits im Vorhinein geklärt, ob der Ausschluss durch eine frische Hüftprothese (bis sechs Wochen postoperativ) vorliegt. Die Ein- und Ausschlusskriterien wurden erstmalig bei der telefonischen Rekrutierung der ProbandInnen und zur definitiven Absicherung am Tag der Durchführung der Studie vom Assistenten bei der Anmeldung persönlich und mittels Fragebogen abgefragt. Die Zustimmung und Einwilligung in Form einer Unterschrift des/der ProbandIn wurde zusätzlich von der Studienleiterin kontrolliert.

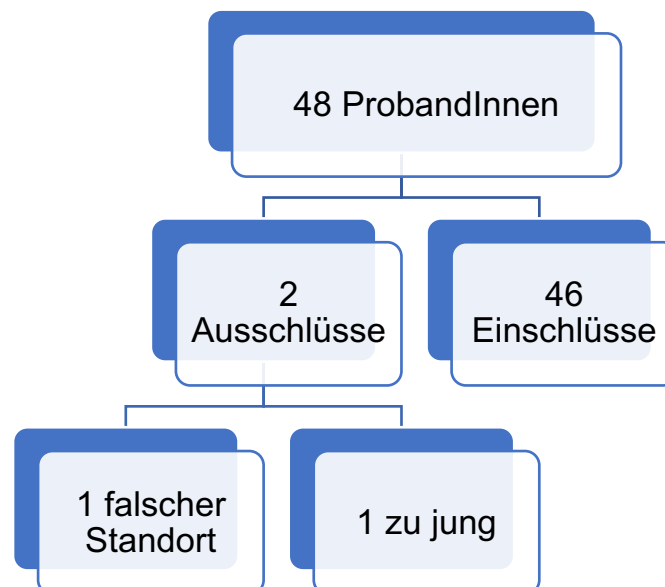
4.6.3 Stichprobengröße

Die Anzahl von 40 ProbandInnen wird im Protokoll für Reliabilitäts- und Validitätsstudien von Patijn (2019) angegeben. Um die Mindestanzahl von 40 Personen nicht zu unterschreiten wurden mehr Personen zur Studienteilnahme eingeladen. Akquiriert wurden 48 Personen. Eine Probandin war am falschen Standort. Sie wurde aus der Studie ausgeschlossen, damit das zeitliche Management nicht beeinflusst wurde. Die Reliabilitätsstudie wurde anhand von 47 Testpersonen durchgeführt. Eine weitere Person wurde auf Grund des zu geringen Alters (17 Jahre) im Nachhinein aus der Studie ausgeschlossen (siehe Abbildung 1). Es kam auf Grund einer Drittperson, der Mutter des Probanden, zu diesem Missverständnis. Die Mutter hat das telefonische Erstgespräch geführt und wollte ihren älteren Sohn mitbringen, der dann am Tag der Studie verhindert war. Der Assistent und auch die Studienleiterin haben das Geburtsdatum im Rahmen der Datenerhebung nicht kontrolliert. Erst bei der sozialdemographischen Datenerhebung wurde dieser Fehler ersichtlich. Proband F wurde im Nachhinein aus der Studie ausgeschlossen. Sämtliche Daten wurden in der vorliegenden Studie nicht berücksichtigt. Ausgewertet wurden die Daten von 46 Personen.

4.6.4 Auswahl der ProbandInnen

Die Akquisition der ProbandInnen erfolgte ab Dezember 2023. Es wurde ein ProbandInnenaufruf gestartet (siehe Anhang B). Aus der kurzen Information konnte man entnehmen, dass es sich um Testungen am Becken handelt, wann und wo die Studie durchgeführt wird und dass man bei Interesse gerne telefonisch genauer informiert wird. Dieses Blatt zur Akquisition von ProbandInnen wurde in der Praxis natürlich- bewegt in 8184 Anger, in der Praxis Xundos in 8200 Gleisdorf sowie in zwei Apotheken ausgelegt. Auf der Homepage der Praxis natürlich- bewegt wurde zusätzlich ein Aufruf gestartet. Außerdem wurde mittels Statusbeitrag auf Whats APP ebenfalls das Informationsblatt geteilt. Etwa eine Woche vor Durchführung der Studie am 14. Jänner konnten 48 Personen zur Studie eingeteilt werden. Die InteressentInnen wurden telefonisch über die Studie, deren Ablauf und Ziel aufgeklärt sowie über die Ein- und Ausschlusskriterien informiert. Sie wurden darauf hingewiesen, dass es sich ausschließlich um Testungen und keine Behandlung handelt. Nach telefonischer Zustimmung zur Teilnahme wurde ihnen ein genauer Termin zugeteilt (siehe Anhang C). Soweit möglich wurden dabei individuelle Wünsche berücksichtigt. Drei Tage vor Durchführung der Studie wurden alle ProbandInnen per Nachricht an den Termin erinnert. Dabei wurden nochmals der Standort und die genaue Uhrzeit mitgeteilt sowie um pünktliches Erscheinen gebeten.

ABBILDUNG 1 AKQUISITION PROBANDINNEN



4.6.5 Die ProbandInnen

Wie in den Einschlusskriterien beschrieben, wurden symptomatische und asymptomatische erwachsene Personen jeglichen Geschlechts zur Studie zugelassen. Dies garantiert eine praxisnahe Situation. Evaluiert wurden etwaige Symptome durch einen Fragebogen (siehe Anhang D). Gleichzeitig wurde mit dem Fragebogen schriftlich das einzige Ausschlusskriterium, das Bestehen einer frischen Hüftprothese bis sechs Wochen

postoperativ, erfragt. Vervollständigt wurde der Fragebogen durch eine Inspektion des unteren Rückens durch die Studienleiterin, um offene Wunden auszuschließen. Mittels Unterschrift wurden alle Angaben sowie die Freiwilligkeit zur Teilnahme und auch die Zustimmung zur Verwertung der Daten seitens des/der ProbandIn gegeben (siehe Anhang E).

4.7 Zielparameter

Während der Recherche zur vorliegenden Studie haben sich außer den primären Zielparametern noch sekundäre Zielparameter ergeben.

4.7.1 Primäre Zielparameter

In der vorliegenden Studie ist die abhängige Variable das Ergebnis des Vorlauftests im Stand und im Sitz mit den zwei Antwortmöglichkeiten, auffällig oder unauffällig. Die unabhängige Variable ist der/die testende OsteopathIn.

4.7.2 Sekundäre Zielparameter

Ergänzend zum primären Zielparameter soll herausgefunden werden, ob eine Korrelation zwischen der Selbsteinschätzung und der Inter- Tester- Reliabilität besteht. Dazu wird zusätzlich die Selbsteinschätzung der RaterInnen zum getesteten Ergebnis, mit sicher und unsicher, abgefragt.

4.8 Der Probedurchlauf

Laut Patijn (2019) ist ein Training maßgeblich für die Durchführung einer guten Reliabilitätsstudie. Reliabilitätsstudien weisen in der Osteopathie oft keine guten Ergebnisse auf. Speziell dieser umstrittene Test, welchen die Studienleiterin für ihre Studie ausgewählt hat, könnte sich durch optimale Voraussetzungen von bisherigen Studien abheben.

Anfang Dezember traf sich die Studienleiterin das erste Mal mit den beiden Kolleginnen, die sich bereiterklärt haben als Testerinnen zu fungieren. Bereits bei der Rekrutierung wurden die zwei Osteopathinnen telefonisch über die Studie aufgeklärt. Im Anschluss wurden ihnen per Mail alle relevanten Informationen zugeschickt. Sie konnten sich also bereits vor dem Treffen mit dem studienrelevanten Ablauf des Vorlauftests im Sitz und im Stand auseinandersetzen. Sie wurden auch über die notwendigen Anatomiekenntnisse informiert und gebeten, sich bereits im Vorhinein damit auseinanderzusetzen. Für Rückfragen wäre die Studienleiterin jederzeit zur Verfügung gestanden. Dies wurde jedoch nicht in Anspruch genommen. Beide Osteopathinnen haben die Ausführung des Vorlauftests in der täglichen Praxis an die Vorgaben der Studienleiterin angepasst und bis zum ersten Treffen regelmäßig im Praxisalltag eingesetzt und geübt. Am vierten Dezember wurde im Zuge des ersten Treffens der Studienleiterin mit den zwei Testerinnen die Palpation aneinander und an der Studienleiterin sowie deren Schwester geübt. Um Fehler bei der Palpation der Orientierungspunkte zu

vermeiden, wurde hierauf besonderes Augenmerk gelegt. O’Haire und Gibbons (2000) nennen diesen Punkt als klare Fehlerquelle in ihrer durchgeführten Pilotstudie. Ziel dabei war es, eine möglichst idente Palpation und Ausführung zu erreichen. Der Vorlaufstest im Stand und im Sitz wurden durchgeführt sowie sämtliche Details wie zeitliches Management, Verblindung und Randomisierung durchgesprochen.

Der Probelauf fand am 14. Dezember in der Praxis von Testerin B statt. Es wurden zehn ProbandInnen eingeladen. Bereits beim Probelauf wurden die ProbandInnen im Zehnminutenrhythmus eingeteilt. Ein Fragebogen und auch eine Einwilligungserklärung waren seitens der ProbandInnen auszufüllen. Eine möglichst ähnliche Durchführung zur eigentlichen Studie wurde forciert. Unter den ProbandInnen befanden sich vier PhysiotherapeutInnen, die im Anschluss gutes Feedback zur Palpation und den verbalen Anweisungen der Testerinnen geben konnten. Außerdem waren Personen zum Probelauf geladen, welche keine medizinischen Kenntnisse hatten. Eine Anpassung der Formulierungen des kurzen Fragebogens hat daraufhin stattgefunden. So wurden Abkürzungen, wie zum Beispiel „DM“ für Diabetes Mellitus, entfernt und durch das ausgeschriebene Wort ersetzt und auch das Wort „Fraktur“ durch „Knochenbruch“ zum besseren Verständnis ausgebessert. Es hat sich herausgestellt, dass es bezüglich Palpation und Testung keinerlei Unklarheiten gab. Seitens der Testerinnen wurde lediglich eine Veränderung des Ergebnisbogens gefordert. Dieser wurde in erster Instanz als zu wenig übersichtlich beschrieben. Durch die räumliche Aufteilung der Praxis kam es beim Probelauf teilweise zu relativ viel Unruhe. Der Wechselbereich der Testerinnen befand sich im Aufenthaltsbereich der wartenden und auch verabschiedenden Personen.

Die Ergebnisse des erfolgten Probelaufs waren sowohl organisatorisch als auch rechnerisch zufriedenstellend. Die Testerinnen haben den Probelauf, vor allem hinsichtlich Nervosität, als positiv empfunden. Beide gaben im Feedback an, dass sie mit jeder zu testenden Person sicherer im Ablauf, den Anweisungen und der Dokumentation der Ergebnisse wurden. Der berechnete P_0 , die relative Übereinstimmung, war mit 0,8 für den Vorlaufstest im Stand und 0,8 für den Vorlaufstest im Sitz laut Angaben von Patijn (2019) in Ordnung (siehe Tabelle 2). Von einer weiteren Trainingsphase wurde daher abgesehen.

Aus Interesse der Studienleiterin, welches im Rahmen der Literaturrecherche entstand, wurde zusätzlich die Confidence- Rating- Scale zur Bewertung der eigenen Sicherheit der Testerinnen herangezogen. Dies war in der ursprünglichen Planung nicht vorgesehen, wurde aber vor allem in der Studie von Cooperstein et al. (2010) als sehr interessant wahrgenommen und kann mit sehr geringem Aufwand eine gute zusätzliche Information liefern. Beim Probelauf der vorliegenden Studie war die 100% Übereinstimmung der Testerinnen bei der Angabe der Sicherheit ihres Ergebnisses auffällig (siehe Tabelle 2). Beide Testerinnen waren beim Probelauf der Studie neun Mal sicher und einmal unsicher. Beide

Testerinnen gaben bei Proband D an, im Ergebnis unsicher zu sein. Laut Kurzfragebogen von Proband D gab es weder aktuelle Beschwerden noch eine Grunderkrankung während der Testung. Interessant ist allerdings, dass bei Proband D 14 Tage später akute Rückenschmerzen mit Verdacht auf Prolaps in der Lendenwirbelsäule aufgetreten sind, welcher sich später auch bestätigt hat.

TABELLE 2 ERGEBNISSE PROBELAUF

VORLAUFTEST	STAND	SITZ
Anzahl der Kategorien	2	2
ProbandInnenanzahl [n]	10	10
TesterInnenanzahl [n]	2	2
Relative Übereinstimmung P_0 [%]	0,8	0,8
Zufällige Übereinstimmung P_e [%]	0,68	0,82
Cohens- Kappa κ [-1 bis +1]	0,375	-0,011
CRS Übereinstimmung	100%	100%

4.9 Studienablauf

Nachfolgend wird der Ablauf der Studie genau beschrieben. Die Verblindung, Randomisierung und Durchführung der Studie stehen dabei im Fokus.

4.9.1 Verblindung

Die Verblindung zwischen ProbandInnen und Testerinnen wurde bereits im Vorhinein durch örtliche Trennung erreicht. Beide Testerinnen sind aus Graz. Die meisten ProbandInnen aus dem Bezirk Weiz. Da die Studienleiterin den Fragebogen mit aktuellen Beschwerden und Grunderkrankungen in Abwesenheit der Testerinnen ausfüllen ließ, waren die Testerinnen auch bezüglich symptomatisch und asymptomatisch sowie in Bezug auf Grunderkrankungen geblindet. Die ProbandInnen und Testerinnen wurden mündlich und schriftlich darauf hingewiesen, nicht mehr als notwendig miteinander zu kommunizieren. Natürlich durften sie sich begrüßen und die Testerinnen die notwendigen Anweisungen aussprechen (bitte ziehen Sie sich die Hose aus, hüftbreit hinstellen, zum Boden bücken, hinsetzen, usw.). Die Blindierung zwischen den beiden Testerinnen während des Raumwechsels am Gang wurde durch die Aufsicht der Studienleiterin gewährleistet. Beide Testerinnen wurden im Vorhinein angewiesen, am Gang nicht miteinander zu sprechen oder auf sonstige Art und Weise zu kommunizieren.

4.9.2 Randomisierung

Sobald der Fragebogen ausgefüllt, das Informationsblatt sorgfältig gelesen, Fragen beantwortet und das Datenblatt mit der freiwilligen Teilnahme unterschrieben waren, kam es zur Zulosung des Therapieraumes. Ein Münzwurf des Assistenten entschied über die Zuteilung der ProbandInnen zum roten oder blauen Therapieraum. Die Benennung mittels Farben wurde gewählt, um eine Wertung der Räumlichkeiten oder der Testerinnen durch die ProbandInnen zu vermeiden.

4.9.3 Durchführung

Als Tag für die Studiendurchführung wurde aus organisatorischen Gründen ein Sonntag gewählt. Als Standort wurde die Praxis eines Kollegen in 8200 Gleisdorf ausgewählt. Es waren jeweils zwei ProbandInnen im Zehnminuten Rhythmus zeitgleich eingeplant. Obwohl Patijn (2019) davon abrät mehrere Testungen innerhalb einer Studie durchzuführen, hat sich die Studienleiterin für eine praxisnahe Ausführung entschieden. Der Vorlauftest wird von OsteopathInnen zuerst im Stand und im Anschluss im Sitz ausgeführt.

Die Testerinnen haben die an der Studie teilnehmenden ProbandInnen gebeten, sich Schuhe und Hose auszuziehen. Zuerst wurde der Vorlauftest im Stand ausgeführt. Eine einmalige Durchführung des Vorlauftests wurde mit den Testerinnen im Vorhinein festgelegt. Danach wurde der/die ProbandIn gebeten sich auf die Liege zu setzen, die Füße sollten dabei hüftbreit abgestellt und der Bodenkontakt gehalten werden.

Der Vorlauftest im Sitz wurde direkt im Anschluss an den Test im Stand durchgeführt. Wie im Stand war auch der Vorlauftest im Sitz nur ein einziges Mal pro ProbandIn auszuführen und zu bewerten.

Zur Identifikation durften die ProbandInnen das Informationsblatt der Studie mit in den Raum nehmen. Das Blatt war mit einem Groß- oder Kleinbuchstaben versehen, welcher den Testerinnen zur anonymisierten Zuordnung ihres Ergebnisses diente. War das Blatt also mit einem großen D versehen, musste das Ergebnis des Tests in der Zeile des Großbuchstabens D eingetragen werden (siehe Ergebnisliste Anhang F).

Nach Durchführung der beiden Testungen wurde der/die ProbandIn gebeten aufzustehen und ein paar Schritte durch den Raum zu gehen. In der Zwischenzeit wechselten die Raterinnen unter Aufsicht der Studienleiterin die zwei Therapieräume. Jede Raterin führte ihre Ergebnisliste beim Wechseln der Räume mit sich. Nach dem Wechsel führte die jeweils andere Testerin die Testungen durch und vermerkte wiederum das Ergebnis in der eigenen Ergebnisliste mit dem zugehörigen Identifikationsbuchstaben. Der/Die ProbandIn durfte sich wieder anziehen und wurde durch die Studienleiterin mit einem kleinen Dankeschön für die Teilnahme verabschiedet.

Die Ergebnislisten der beiden Testerinnen wurden vor der eingeplanten Pause von der Studienleiterin eingesammelt und direkt in einer Mappe eingeordnet. Nach der Pause wurden neue Ergebnislisten ausgeteilt (siehe Anhang G).

Die Identifikation der zweiten Untersuchungsgruppe war alphabetisch mit Kleinbuchstaben gegeben. Nach der letzten Testung wurden die Ergebnislisten wiederum von der Studienleiterin eingesammelt und ebenfalls sofort in eine Mappe eingeordnet. Die Testerinnen erhielten von der Studienleiterin auch ein kleines Dankeschön (Geschenkskörbe) für die in Anspruch genommene Zeit und die Bemühungen, alle Richtlinien und Vorgaben eingehalten zu haben.

4.10 Datenverarbeitung und Auswertung

Alle Rohdaten wurden ins Microsoft Excel eingegeben und verarbeitet. Um Fehler zu vermeiden, wurde ein Informatiker hinzugezogen. Mit den zugehörigen Formeln wurden genaue Berechnungen durchgeführt. Die Auswertung wurde von der Autorin in Zusammenarbeit mit dem Informatiker durchgeführt. Erhoben wurden Alter, Geschlecht, aktuelle Beschwerden, Grunderkrankungen, die Testergebnisse der beiden Testungen, dem Vorlauftest im Stand und dem Vorlauftest im Sitz sowie die jeweilige Sicherheit in der Angabe des jeweiligen Testergebnisses.

Das nominale Messniveau der durchgeführten Reliabilitätsstudie verlangt zur Berechnung den Kappa- Koeffizienten nach Cohen (Patijn, 2019). Da Cooperstein et al. (2010) angeben, dass die Inter- Rater- Reliabilität höher ist, wenn sich die Rater bei deren Angaben bezogen auf das Testergebnis sicher sind, wurde die Confidence- Rating- Scale als zusätzlicher Parameter herangezogen. Dieser sekundäre Zielparameter der Selbstbeurteilung wurde mittels zusätzlicher Angabe von sicher oder unsicher auf der Ergebnisliste der Testerinnen abgefragt (siehe Anhang F und G).

5 Ergebnisse

Zuerst wird im folgenden Kapitel die deskriptive Statistik beschrieben. Danach werden die Ergebnisse des Vorlauftests im Stand und die Ergebnisse des Vorlauftests im Sitz präsentiert und zuletzt wird auf die subjektive Angabe der eigenen Sicherheit Bezug genommen.

5.1 Darstellung und Berechnung der Daten

Microsoft Excel ist ein mit Zahlen und Formeln arbeitendes Programm. Sämtliche Variablen der vorliegenden Studie wurden mit Zahlen kodiert. Um die zugeordneten Zahlen zu den Variablen nachzuvollziehen zu können, werden die einzelnen Variablen mit ihren Kodierungen in den dazugehörigen Unterkapiteln kurz erläutert und im Anhang H nochmals übersichtlich dargestellt.

5.2 Sozialdemographische Daten zu den ProbandInnen

Es wurden 48 ProbandInnen zur Teilnahme an der Studie eingeladen. Am Ende haben 47 Personen an der Testung teilgenommen. Auf Grund eines Ausschlusses durch den falschen Standort einer Probandin und eines Ausschlusses im Nachhinein auf Grund des Alters wurden bei der Angabe der sozialdemographischen Daten nur 46 ProbandInnen berücksichtigt (siehe Abbildung 1).

Die Erhebungen der Daten von Proband F wurden sowohl sozialdemographisch als auch bei den Ergebnissen nicht berücksichtigt.

Am Datenbogen konnten die ProbandInnen die Frage nach dem Geschlecht frei beantworten. Es wurde ausschließlich mit weiblich oder männlich geantwortet.

Für die Berechnungen wurde die Variable Geschlecht mit der Zahl eins für weiblich und der Zahl zwei für männlich kodiert.

In der unten angeführten Tabelle 3 werden die absoluten Zahlen und die prozentuellen Angaben dargestellt. Die Übersicht der Altersverteilung inklusive der Standardabweichung (SD) kann Tabelle 4 entnommen werden.

TABELLE 3 GESCHLECHTERVERTEILUNG

Geschlecht (n= 46)	n	%
weiblich	36	78,26
männlich	10	21,74

TABELLE 4 ALTERSVERTEILUNG

Alter [Jahre]	Minimum	Maximum	Median	Mittelwert	SD
(n = 46)	22	77	42	44,33	14,3

Zusätzlich zum Geburtsdatum und dem Geschlecht wurden auch die Fragen aus dem Gesundheitsfragebogen, welchen die ProbandInnen am Tag der Studie ausfüllen mussten, ausgewertet. Der Fragebogen umfasste vier Entscheidungsfragen mit der Antwortmöglichkeit Ja oder Nein sowie einer möglichen individuellen Ergänzung bei den Fragen nach aktuellen Symptomen und der Grunderkrankungen durch die Person selbst.

Erstens wurde das einzige Ausschlusskriterium, mit der Frage nach einer frischen Hüftprothese, sichergestellt.

Zweitens und drittens wurde nach etwaigen Grunderkrankungen und derzeitigen Beschwerden gefragt.

Viertens wurde durch die Inspektion der Studienleitung ein Ausschluss aus der Studie auf Grund einer offenen Wunde im Bereich des unteren Rückens sichergestellt.

Aus dem kurzen Fragebogen ging hervor, dass elf ProbandInnen im Moment der Datenerhebung Beschwerden hatten. Diese Personen wurden in der vorliegenden Studie als symptomatisch eingestuft. Die Kodierung für symptomatische Patienten im Microsoft Excel lautet zwei. Vier Personen gaben Probleme im Bereich des Nackens und Schultergürtels an. Zwei Probandinnen nannten Probleme im unteren Rücken. Außerdem wurden ein Mal Schmerzen im Gesäß, ein Mal Hüftprobleme und ein Mal eine Knieprothese links mit bestehenden Knieschmerzen genannt. Eine Person äußerte einen verringerten Allgemeinzustand bzw. Gliederschmerzen durch einen zur Zeit der Studiendurchführung bestehenden Infekt. Eine weitere Person beantwortete die Frage nach derzeitigen Beschwerden mit Ja, machte aber keine weiteren Angaben dazu.

Asymptomatische ProbandInnen wurden in den Berechnungen mit der Zahl eins kodiert. Eine Übersicht zur Verteilung symptomatischer ProbandInnen und asymptomatischen ProbandInnen zeigt Tabelle 5.

TABELLE 5 VERTEILUNG SYMPTOMATISCH/ ASYMPTOMATISCH

(n=46)	n	[%]
symptomatisch	11	23,91%
asymptomatisch	35	76,09%

Es wurden im Rahmen des Fragebogens auch Grunderkrankungen erhoben. Grunderkrankungen wurden nur unter den Frauen genannt (siehe Tabelle 6).

Sieben Probandinnen beantworteten die Frage mit Ja. Ein Mal wurde eine Rheumaerkrankung genannt. Zweimal eine Fehlfunktion der Schilddrüse und am häufigsten Osteoporose. Drei Probandinnen mit einer Grunderkrankung hatten zusätzlich aktuelle Beschwerden (Gesäßschmerz, unterer Rücken, Hüftbeschwerden).

Die Kodierung für die Berechnungen im Microsoft Excel lautet eins für keine Grunderkrankung und die zwei für Grunderkrankung. Dreimal wurden Probandinnen mit einer Grunderkrankung auch als symptomatisch eingeordnet.

TABELLE 6 VERTEILUNG GRUNDERKRANKUNGEN

Grunderkrankung	(n=46)	weiblich	männlich
Ja	7	7	0
Nein	39	29	10

5.3 Ergebnisse der Raterinnen

Wie von Patijn (2019) empfohlen werden die Daten der in der vorliegenden Studie durchgeführten Testungen mittels Kreuztabelle dargestellt. Die verwendete Vorlage befindet sich im Anhang I.

Um den Kappa- Koeffizienten zu erhalten, müssen einige Variablen zuvor berechnet werden. Die prozentuelle Übereinstimmung (P_0) und die prozentuelle Zufallsübereinstimmung (P_e) sind dabei von Relevanz.

Die Ergebniskategorien für die durchgeführten Testungen lauten unauffällig oder auffällig. Die Kodierung für die Analyse ist eins für unauffällig und zwei für auffällig.

Die Interpretation des Kappa- Koeffizienten folgt nach der Klassifikation von Landis und Koch (1977) (siehe Tabelle 7). Die Tabelle wurde für die Übersichtlichkeit an dieser Stelle nochmals eingefügt. Laut Patijn (2019) können Werte zwischen -1 und 1 angenommen werden. Die Übersetzung wurde von der Autorin vorgenommen.

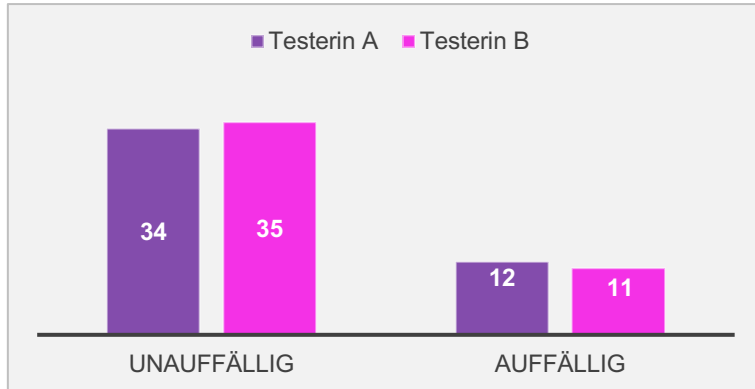
TABELLE 7 KLASSIFIKATION DER COHENS KAPPA KOEFFIZIENTEN- QUELLE: ADAPTIERT NACH LANDIS UND KOCH (1977)

Cohens Kappa Koeffizient	Grad der Übereinstimmung
<0.00	Poor (mangelhaft)
0.00-0.20	Slight (gering)
0.21-0.40	Fair (mäßig, ausreichend)
0.41-0.60	Moderate (mittelmäßig)
0.61-0.80	Substantial (beachtlich)
0.81-1.00	Almost perfect (fast perfekt)

5.3.1 Ergebnisse Raterinnen in absoluten Zahlen Vorlauftest im Stand

In Abbildung 2 werden die Ergebnisse in absoluten Zahlen von Testerin A und Testerin B gegenübergestellt.

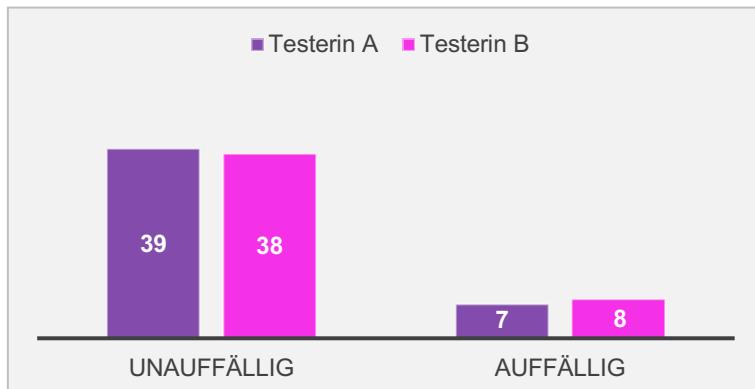
ABBILDUNG 2 ERGEBNIS IN ABSOLUTEN ZAHLEN VORLAUFTEST STAND



5.3.2 Ergebnisse Raterinnen in absoluten Zahlen Vorlauftest im Sitz

Die Ergebnisse von Testerin A und Testerin B in absoluten Zahlen für den Vorlauftest im Sitz können Abbildung 3 entnommen werden.

ABBILDUNG 3 ERGEBNIS IN ABSOLUTEN ZAHLEN VORLAUFTEST SITZ



5.4 Kappa Koeffizienten der Vorlauftestungen

5.4.1 Kappa- Koeffizient Vorlauftest im Stand gesamt

In der untenstehenden Kreuztabelle sind die beobachteten Häufigkeiten der Übereinstimmung aller ProbandInnen des Vorlauftests im Stand ersichtlich. Die prozentuelle Zufallsübereinstimmung liegt bei 62,5%. Berechnet aus P_0 und P_e ergibt sich ein Kappa-Koeffizient von $\kappa = 0,36$. Das Konfidenzintervall mit 95% wird von 0.057 bis 0.668 angegeben. Zusätzlich wurde der p- Wert mit einem Signifikanzniveau von 0,05 berechnet. Alle Werte in absoluten Zahlen und in Prozenten angegeben können Tabelle 8 und 9 entnommen werden.

TABELLE 8 KREUZTABELLE GESAMTERGEBNIS VORLAUFTTEST STAND

Kreuztabelle Testerin A / Testerin B – Gesamtergebnis Vorlauftest Stand					
		Testerin B		Gesamt	
		auffällig	unauffällig		
Testerin A	auffällig	Anzahl	6	6	12
		% d. Gesamt-A	13,04%	13,04%	26,08%
	unauffällig	Anzahl	5	29	34
		% d. Gesamt-A	10,87%	63,04%	73,91%
Gesamt	Anzahl	11	35	46	
	% d. Gesamt-A	23,91%	76,09%	100%	
		Übereinstimmung		76,09%	
		Keine Übereinstimmung		23,91%	

TABELLE 9 DARSTELLUNG ERGEBNISSE VORLAUFTTEST STAND

Kappa- Koeffizient	0,36
P ₀	76,09%
P _{Index}	0,25
p- Wert	0,8097

5.4.2 Explorative Analyse Vorlauftest im Stand

Die höchste Übereinstimmung gibt es nach Cohens- Kappa bei der explorativen Datenanalyse in der Gruppe der ProbandInnen vor der Pause mit $\kappa = 0,51$, also den ProbandInnen mit Identifikation Großbuchstaben A bis X. Der größte Unterschied bei der Erhebung der Kappa-Werte liegt zwischen den Testungen beim Vorlauftest im Stand vor und nach der Pause (siehe Tabelle 10). Nicht berücksichtigt wurde in der explorativen Analyse die Übereinstimmung, bei der sich beide Raterinnen bei der gleichen Probandin unsicher waren. Da die Anzahl mit $n=1$ keine sinnvolle Berechnung zulässt.

Aus den vorhandenen Daten (Geschlecht, symptomatisch und asymptomatisch, Grunderkrankung und keine Grunderkrankung) wurden interessehalber weitere Ergebnisse berechnet (siehe Anhang J).

TABELLE 10 ERGEBNISSE EXPLORATIVE ANALYSE VORLAUFTTEST STAND

	Cohens-Kappa Koeffizient [-1 bis +1]	Relative Übereinstimmung P_0 [%]	Zufällige Übereinstimmung P_e [%]	CRS Übereinstimmung [%]	p- Wert
vor der Pause n= 23 (A bis X)	$\kappa= 0,51$	78,26%	55,95%	69,6%	p= 0,7542
nach der Pause n=23 (a-w)	$\kappa= 0,09$	73,91%	71,27%	69,6%	nicht ermittelbar
Beide Raterinnen sicher n=31	$\kappa= 0,35$	83,87%	75,13%	100%	p= 0,7184
Eine Raterin unsicher n=14	$\kappa= 0,16$	57,14%	48,98%	0%	p= 0,2557

5.4.3 Kappa- Koeffizient Vorlaufstest im Sitz gesamt

Wie in Tabelle 8 gibt es in Tabelle 11 zwei Kategorien. Die Ergebnisse des Vorlaufstests im Sitz können Tabelle 11 und Tabelle 12 entnommen werden. Das Konfidenzintervall mit 95% liegt zwischen -0.069 und 0,637.

TABELLE 11 KREUZTABELLE GESAMTERGEBNIS VORLAUFTTEST SITZ

Kreuztabelle Testerin A / Testerin B – Gesamtergebnis Vorlaufstest Sitz					
		Testerin B		Gesamt	
		auffällig	unauffällig		
Testerin A	auffällig	Anzahl	3	4	7
		% d. Gesamt-A	6,52%	8,7%	15,28%
	unauffällig	Anzahl	5	34	39
		% d. Gesamt-A	10,87%	73,91%	84,78%
Gesamt	Anzahl	8	38	46	
	% d. Gesamt-A	17,39%	82,61%	100%	
		Übereinstimmung		80,43%	
		Keine Übereinstimmung		19,57%	

TABELLE 12 DARSTELLUNG ERGEBNISSE VORLAUFTTEST SITZ

Kappa- Koeffizient	0,28
P_0	80,43%
P_{Index}	0,16
p- Wert	0,7800

5.4.4 Explorative Analyse Vorlaufstest im Sitz

In Tabelle 13 werden alle Daten aus der explorativen Analyse des Vorlaufstests im Sitz dargestellt.

TABELLE 13 ERGEBNISSE EXPLORATIVE ANALYSE VORLAUFTEST IM SITZ

	Cohens-Kappa Koeffizient [-1 bis +1]	Relative Übereinstimmung P_0 [%]	Zufällige Übereinstimmung P_e [%]	CRS Übereinstimmung [%]	p- Wert
vor der Pause n= 23	$\kappa = 0,31$	78,26%	68,43%	69,6%	$p=0,7083$
nach der Pause n=23	$\kappa = 0,23$	82,61%	77,32%	95,7%	nicht ermittelbar
Beide Raterinnen sicher n= 38	$\kappa = 0,16$	84,21%	81,16%	100%	$p=0,4551$
Eine Raterin unsicher n= 8	$\kappa = 0,25$	62,5%	50%	0%	$p=0,3019$

5.5 P_{index} Vorlaufstest Stand und Vorlaufstest Sitz

Der nur im Nachhinein zu ermittelnde P_{index} sollte zwischen 0,2 und 0,9 liegen (Patijn, 2019) um einen Kappa- Wert über $\kappa = 0,61$ zu erreichen. Erstrebenswert wäre ein P_{index} von 0,5, dann würde sich der Kappa- Wert am höchsten Punkt der Kappa/ P_{index} Kurve befinden (Patijn, 2019). Berechnet anhand der Formel für P_{index} aus den Unterlagen von Patijn (2019), wurde aus den Daten des Vorlaufstests im Stand ein Wert von 0,25 ermittelt (siehe Tabelle 10). Für den Vorlaufstest im Sitz wurde ein P_{index} von 0,16 berechnet (siehe Tabelle 13).

5.6 Sicherheit der Raterinnen

Die Kodierung für die Angabe der eigenen Sicherheit im Microsoft Excel ist eins für sicher und zwei für unsicher.

In Abbildung 4 ist die Angabe von „sicher“ und „unsicher“ nach der Confidence- Rating- Scale (CRS) in absoluten Zahlen dargestellt.

In Tabelle 11 und 14 ist die Übereinstimmung von Sicher und Unsicher zwischen Testerin A und Testerin B in Prozenten auf die einzelnen Gruppen aufgeschlüsselt angegeben. Es konnte weder bei der Testung im Stand noch bei der Testung im Sitz eine höhere Übereinstimmung als 95,7% in der Sicherheit zwischen den Raterinnen errechnet werden. Durchschnittlich haben die beiden Testerinnen eine höhere Übereinstimmung ihrer Sicherheit bei der Bewertung ihres Ergebnisses im Sitz. Es wurde ein Mittelwert von 67,73% im Stand und ein Mittelwert von 81,74% im Sitz errechnet.

In absoluten Zahlen zeigt sich, dass sich insgesamt, also beim Vorlaufstest im Stand und im Sitz (n=92), Testerin A fast doppelt so häufig unsicher ist wie Testerin B (siehe Abbildung 5).

Insgesamt sind sich entweder Testerin A oder Testerin B 26 Mal unsicher. Bei einer einzigen Person sind sich beide Raterinnen unsicher.

ABBILDUNG 4 ERGEBNISSE CRS IN ABSOLUTEN ZAHLEN

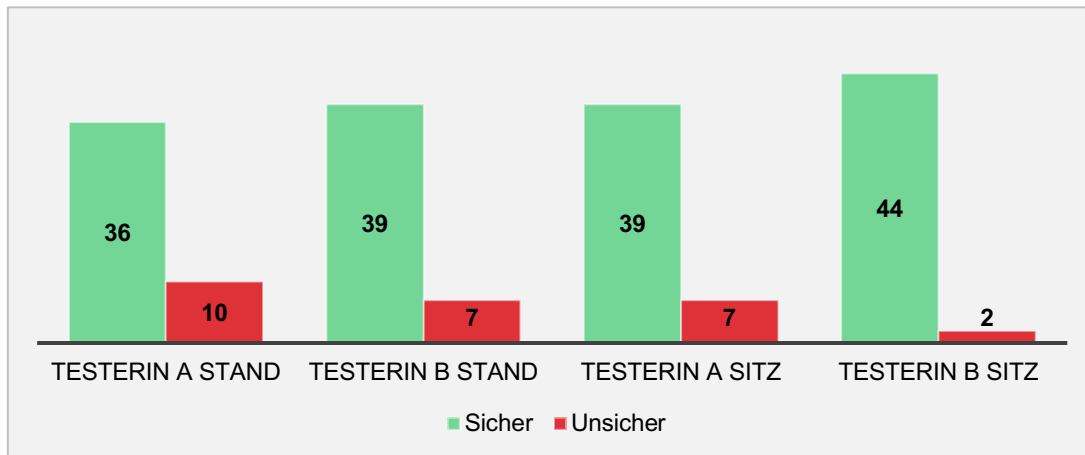
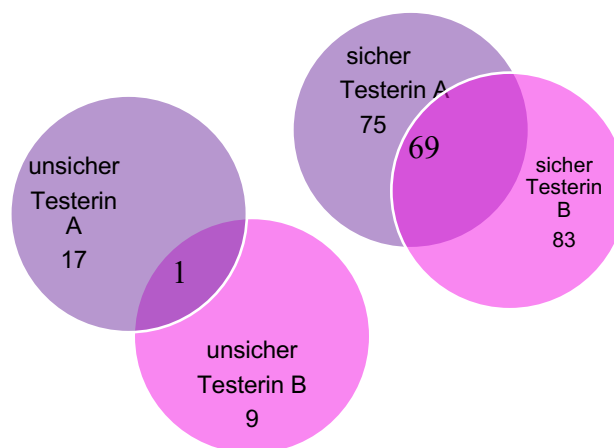


ABBILDUNG 5 ÜBERSCHNEIDUNG SICHERHEIT GESAMT (N=92) TESTERIN A/ TESTERIN B

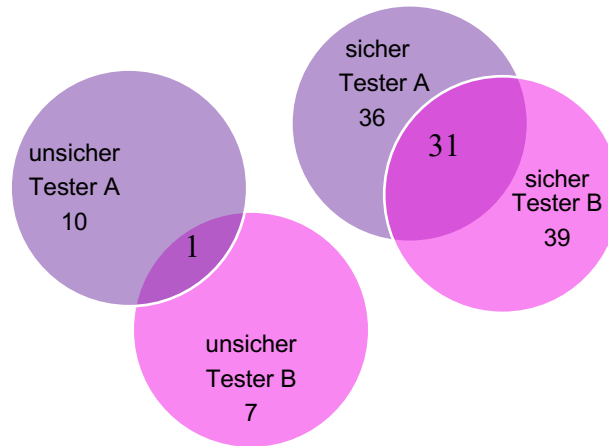


5.6.1 Sicherheit der Raterinnen Vorlaufstest im Stand

In Abbildung 4 ist die subjektive Angabe der Sicherheiten beim Ergebnis des Vorlaufstests im Stand von Testerin A und Testerin B in absoluten Zahlen ersichtlich.

In Prozenten ausgedrückt bedeutet das, dass sich Raterin A in 78,3% ihrer Angaben sicher ist. Raterin B ist sich bei der Angabe der Ergebnisse in 84,8% sicher. Abbildung 6 zeigt wie häufig sich Testerin A und Testerin B bei den gleichen ProbandInnen sicher oder unsicher sind. 31- mal sind sich Testerin A und Testerin B bei den gleichen ProbandInnen in der Angabe ihrer Ergebnisse sicher. Bei einer Person (Proband c) waren sich beide Testerinnen unsicher.

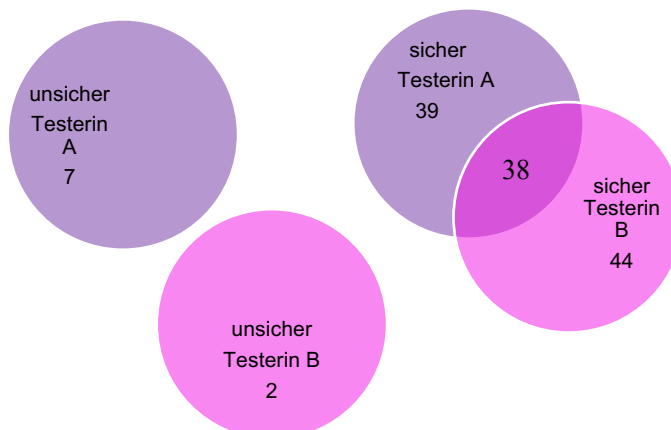
ABBILDUNG 6 ÜBERSCHNEIDUNG CRS VORLAUFTEST STAND (N=46)



5.6.2 Sicherheit der Raterinnen Vorlauftest im Sitz

Die Darstellung der absoluten Zahlen, in Bezug auf die subjektive Angabe der eigenen Sicherheit beim Ergebnis des Vorlauftests im Sitz, kann Abbildung 4 entnommen werden. Siebenmal bewertet sich Testerin A bei der Angabe ihres Ergebnisses selber mit Unsicher. Im Sitz ist sich Testerin A zu 87% in ihrer Angabe des Ergebnisses sicher. Prozentuell gesehen ist sich Raterin B in 95,7% sicher in der Angabe des Ergebnisses der Vorlaufprüfung im Sitz. 38-mal überschneiden sich die Angaben von Sicher bei Testerin A und Testerin B (siehe Abbildung 7). Wie aus der Abbildung 7 zu entnehmen, hat sich die subjektive Angabe unsicher beim Vorlauftest im Sitz bei keiner zu testenden Person überschritten.

ABBILDUNG 7 ÜBERSCHNEIDUNG CRS VORLAUFTEST SITZ (N=46)



5.7 Korrelation von Inter- Rater- Reliabilität und CRS

Die Korrelationsanalyse wurde auf Basis von Wertepaaren aus null und eins anhand einer Tabelle durchgeführt. Das Ergebnis ist aus Tabelle 14 zu entnehmen. Die detaillierte Auswertung der Korrelationen in Bezug auf Untergruppen ist im Anhang K ersichtlich.

TABELLE 14 KORRELATION PRIMÄRER UND SEKUNDÄRER ZIELPARAMETER

Korrelationsvariablen	Korrelationsfaktor (Vorlauftest im Stand)	Korrelationsfaktor (Vorlauftest im Sitz)
Übereinstimmung auffällig und unauffällig Tester A und B/ Übereinstimmung sicher und unsicher Tester A und B gesamt	0,29	0,21

5.8 Überprüfung der Hypothesen

Die unter Kapitel 3.2 und 3.3 beschriebenen Hypothesen werden unter diesem Punkt verifiziert beziehungsweise falsifiziert werden. Zur Überprüfung wird der Cohens Kappa- Koeffizient herangezogen.

5.8.1 Überprüfung der Nullhypothesen

Nullhypothese Vorlauftest Stand: Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand ist kleiner als der Cohens Kappa Koeffizient von 0,61 ($\kappa < 0,61$) und liegt unter dem beachtlichen Bereich (Landis & Koch, 1977).

→ Der Kappa- Koeffizient liegt bei $\kappa = 0,36$ und ist somit kleiner als 0,61. Das dazugehörige 95% Konfidenzintervall liegt zwischen 0.057 bis 0.668. Es wird die Nullhypothese angenommen die besagt, dass die Ergebnisse von den einzelnen Untersuchungspersonen abhängig sind.

Nullhypothese Vorlauftest Sitz: Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Sitz ist kleiner als der Cohens Kappa Koeffizient von 0,61 ($\kappa < 0,61$) und liegt unter dem beachtlichen Bereich (Landis & Koch, 1977).

→ Der Kappa- Koeffizient liegt bei $\kappa = 0,28$ und ist somit kleiner als 0,61. Das dazugehörige 95% Konfidenzintervall liegt zwischen -0.069 bis 0,637. Daher wird die Nullhypothese angenommen die besagt, dass die Ergebnisse von den einzelnen Untersuchungspersonen abhängig sind.

5.8.2 Überprüfung der Alternativhypothesen

Alternativhypothese Vorlauftest im Stand: Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlauftests im Stand ist gleich oder größer als der Cohens Kappa Koeffizient von 0,61 ($\kappa \geq 0,61$) und liegt im oder über dem beachtlichen Bereich (Landis & Koch, 1977).

→ Der Kappa- Koeffizient liegt bei $\kappa = 0,36$ und ist somit kleiner als 0,61. Das dazugehörige 95% Konfidenzintervall liegt zwischen 0.057 bis 0.668. Auf Grund der Ergebnisse wird die Alternativhypothese abgelehnt.

Alternativhypothese Vorlaufstest im Sitz: Die Übereinstimmung der TesterInnen bei der Beurteilung der Auffälligkeit des Vorlaufstests im Sitz ist gleich oder größer als der Cohens Kappa Koeffizient von 0,61 ($\kappa \geq 0,61$) und liegt im oder über dem beachtlichen Bereich (Landis & Koch, 1977).

→ Der Kappa- Koeffizient liegt bei $\kappa = 0,28$ und ist somit kleiner als 0,61. Das dazugehörige 95% Konfidenzintervall liegt zwischen -0.069 bis 0,637. Die Alternativhypothese wird daher abgelehnt.

6 Diskussion

Für die bessere Übersichtlichkeit wird das Kapitel in mehrere Subkapitel unterteilt. Zu Beginn werden die Ergebnisse mit der bestehenden Literatur verglichen, danach die Ergebnisse interpretiert und der Kontext zur Osteopathie hergestellt. Limitationen durch das Studienpersonal und die Stichprobe werden aufgezeigt und kritisch reflektiert sowie der Studienablauf diskutiert.

6.1 Vergleich der Ergebnisse mit bestehender Literatur

Wie bei der Studie von Ribeiro et al. (2021), die sich bei der Untersuchung des Vorlauftests im Stand und im Sitz auf eine Dysfunktion des Iliosakralgelenkes bezieht, konnten in der vorliegenden Arbeit keine ausreichenden Ergebnisse erreicht werden. Ribeiro et al. (2021) berücksichtigen in der Stichprobe nur asymptomatische ProbandInnen. PatientInnen mit Rückenschmerzen wurden aus der Studie ausgeschlossen. Die Inter- Rater-Reliabilität wurde für den Vorlauftest im Stand mit 10% Übereinstimmung, einem Kappa- Koeffizienten von $\kappa = -0,02$ und einem p- Wert von 0,825 und beim Vorlauftest im Sitz mit 13,3 % Übereinstimmung, $\kappa = 0,01$ und einem p- Wert mit 0,836 bestimmt.

Auch bei der Studie von Fryer et al. (2005), bei der der Vorlauftest im Sitz zur Überprüfung von Iliosakraldysfunktionen herangezogen wurde, wurden Werte im Bereich geringer Übereinstimmung zwischen den Testern erreicht ($\kappa = 0,14$ und $\kappa = 0,07$). Die Stichprobengröße von zehn ProbandInnen ist allerdings gering. Berücksichtigt wurden nur asymptomatische Frauen.

Vincent- Smith und Gibbons (1999) überprüften mit neun RaterInnen an 14 freiwilligen Osteopathiestudierenden in zwei Durchgängen die Übereinstimmung. Es geht nicht klar hervor, ob auch diese asymptomatisch waren. Die Übereinstimmung zwischen den Prüfern wurde mit 34% der Prüfungen im ersten Versuch und mit 50% im zweiten Versuch festgestellt. Dies ergab einen Inter- Prüfer- Reliabilitätskoeffizienten (κ) von 0,013 für Versuch eins und 0,09 für Versuch zwei. Die RaterInnen gaben an, in zweiter Instanz sicherer bei der Durchführung gewesen zu sein.

Wie auch in der Studie von Ribeiro et al. (2021), Fryer et al. (2005) als auch bei Vincent- Smith und Gibbons (1999) zeigt sich auch die Inter- Rater- Reliabilität des Vorlauftests als Teil der globalen Untersuchung in vorliegender Arbeit nur als mäßig bis ausreichend. Die Inter- Tester- Reliabilität dieses Tests als Teil der globalen Untersuchung zeigte sich in vorliegender Arbeit als mäßig bis ausreichend. Wie McHugh (2012) beschreibt, ist jede Übereinstimmung unter $\kappa = 0,6$ als unzureichend einzustufen.

In der Studie von Cooperstein et al. (2010) wurde nicht der Vorlauftest sondern eine Bewegungspalpation der Brustwirbelsäule durchgeführt. Die Ergebnisse werden als gut

beschrieben, wenn sich die RaterInnen in der Angabe ihrer Ergebnisse sicher waren. Insgesamt ist die Übereinstimmung zwischen den RaterInnen aber nach Definition eher moderat (ICC- Wert 0,3110). Ein direkter Vergleich ist nicht gut möglich, da sie zur Berechnung und Darstellung der Ergebnisse den Intraclass- Koeffizienten verwenden.

Da die Motivation zur vorliegenden Studie der Vorlaufstest als Teil der globalen Untersuchung ist, ist auch der wissenschaftliche Vergleich der Ergebnisse mit anderen globalen Untersuchungsmethoden interessant.

In der Recherche der Autorin konnte im Nachhinein eine Studie hinsichtlich der Reliabilität von globalen Testungen, dem Ecoute- Test gefunden, werden. Rittler (2010) hat anhand der Inter-Rater- und Intra- Rater- Reliabilitätsstudie festgestellt, dass es nur zu sehr geringen zufälligen Übereinstimmungen in der Testung des „global listenings“ kommt.

Boddenberg zeigte 2013 in seiner Studie, dass sich Veränderung im Kieferbereich durch Aufbeißen auf Watteröllchen auf das Ergebnis des Vorlauftestes auswirkt, was auf die Qualifizierung des Vorlauftestes als globales Testmittel hinweist. Laut Boddenberg dient der Vorlaufstest als guter Parameter in der Befundung.

Generell scheinen osteopathische Reliabilitätsstudien schlechte Übereinstimmungen vorzuweisen. Auch die Ergebnisse vorliegender Studie sind als nicht zufriedenstellend zu bewerten. Somit stellt sich, wie auch im Review von Nolet et al. (2021), der Nutzen der untersuchten Tests in Frage.

Fehlende Literatur in Bezug auf Reliabilitätsstudien von globalen Testmethoden anhand einer Bewertung mittels Kappa- Koeffizienten schränken die Möglichkeit eines Vergleiches auf Augenhöhe stark ein. Da sich die Recherche auf den deutsch- und englischsprachigen Raum beschränkt hat, könnten natürlich anderweitig Studien vorhanden sein, die in der vorliegenden Arbeit nicht berücksichtigt wurden. Eine „language“ Bias wäre durchaus möglich.

6.2 Interpretation der Ergebnisse

Der errechnete Kappa- Koeffizient liegt bei der Vorlaufstestung im Stand bei $\kappa = 0,36$. Die Übereinstimmung der Ergebnisse in Prozenten zwischen Testerin A und Testerin B betragen beim Vorlaufstest im Stand 76,09%. Das dazugehörige 95% Konfidenzintervall liegt zwischen 0.057 und 0.668. Damit ist das Intervall nicht so groß und deutet darauf hin, dass die Schätzung mit einer höheren Genauigkeit behaftet ist. Der wahre Wert liegt somit wahrscheinlich näher an der Schätzung.

Nach der Analyse und Berechnung konnte, nach Klassifikation von Landis und Koch (1977), eine mäßig beziehungsweise ausreichende Inter- Rater- Reliabilität beim Vorlaufstest im Stand nachgewiesen werden und ist nur in begrenztem Maße konsistent. Die Alternativhypothese wird verworfen. Mit einem p- Wert von 0,8097 ist das Ergebnis außerdem statistisch nicht signifikant.

Der berechnete Kappa- Koeffizient für den Vorlauftest im Sitz beträgt $\kappa = 0,28$. In Prozenten wurde eine Übereinstimmung von 80,43% für den Vorlauftest im Sitz berechnet. Der Vorlauftest im Sitz ist daher ebenfalls nicht reliabel und weist genauso nur eine begrenzte Konsistenz auf. Das dazugehörige 95% Konfidenzintervall liegt zwischen -0,069 und 0,637. Das Intervall ist somit relativ breit und mit einer gewissen Unsicherheit behaftet. Der p- Wert beträgt 0,7800 und ist statistisch ebenfalls nicht signifikant. Die Alternativhypothese wird verworfen und die Nullhypothese angenommen.

Entsprechend der Aussage von Fjellner et al. (1999), sind die beiden errechneten Kappa- Werte aus der vorliegenden Studie mit $\kappa = 0,36$ und $\kappa = 0,28$ zu gering, um zumindest als akzeptabel gewertet zu werden. Diese Aussage wird von McHugh (2012) unterstrichen. Wünschenswert wäre das Erreichen eines Kappa- Wertes von $\kappa = 0,4$ (Fjellner et al., 1999; McHugh, 2012).

Die von Patijn (2019) empfohlene Berechnung von P_{index} liegt beim Vorlauftest im Stand bei 0,25 und beim Vorlauftest im Sitz bei 0,16. Bewegt sich der Wert von P_{index} außerhalb der Grenzen von 0,2- 0,9, kann es zu Verzerrungen des Kappa- Wertes kommen (Patijn 2019). Da der Wert immer erst nach Durchführung der Studie berechnet werden kann, ergibt sich ein großer Nachteil von Reliabilitätsstudien (Patijn, 2019). In der vorliegenden Studie kommt es, vor allem bei der Testung des Vorlauftests im Sitz, zu zu wenig positiven also auffälligen Übereinstimmungen, was wiederum den niedrigen P_{index} von 0,16 verursacht. Ein Kappa- Wert von mehr als $\kappa = 0,61$ ist daher bereits völlig ausgeschlossen. Der P_{index} mit 0,25 vom Vorlauftest im Stand befindet sich im untersten Bereich der Grenzen und könnte ebenso einer Bias unterliegen.

Ziel einer Reliabilitätsstudie mit dem Kappa- Koeffizienten nach Cohen ist ein P_{index} von 0,5. Der Kappa- Wert wäre dann an der Spitze der Kappa/ P_{index} Kurve. Patijn (2019) beschreibt, die 0,50- P_{index} Methode als Möglichkeit dieser Bias vorzubeugen. Bei dieser Methode untersuchen Tester A und Tester B jeweils gleich viele ProbandInnen. Die eine Hälfte mit Auffälligen und die andere Hälfte mit Unauffälligem Test. Die ProbandInnen werden nach der ersten Testung zum anderen Rater geschickt, der/die wiederum mit auffällig oder unauffällig bewertet. Dabei sollte beim Probelauf bereits ein P_0 von 0,85 angestrebt werden. Die 0,50- P_{index} Methode wäre die beste Lösung, um dieser Bias beziehungsweise diesem Nachteil von Reliabilitätsstudien mit zwei Ratern und zwei Antwortmöglichkeiten vorzubeugen (Patijn, 2019). In der vorliegenden Studie muss durch den geringen P_{index} , speziell im Falle des Vorlauftests im Sitz, der berechnete Kappa- Wert von 0,28 in Frage gestellt werden. Offensichtlich besteht ein Kappa- Paradoxon wie es auch Derksen et al. (2024) in ihrer Studie beschreiben. Ist die Fallverteilung sehr ungleichmäßig, kann der niedrige Kappa- Koeffizient ein gutes Ergebnis sein. So die Erklärung von Derksen et al. (2024) in ihrem Artikel zum Kappa- Paradox. Die unter Punkt 5.2 beschriebene Stichprobe nennt elf Symptomatische und

sieben Personen mit einer Grunderkrankung bei einer Stichprobengröße von $n=46$. Ab wann die Fallverteilung als ungleichmäßig gilt, wurde von Derksen et al. (2024) nicht genauer ausgeführt.

Interessant zeigt sich unter Punkt 5.4.2 in der explorativen Analyse, dass beim Vorlauftest im Stand vor der Pause, also in der Gruppe der ProbandInnen mit Kennzeichnung Großbuchstaben A bis X ($n=23$), der höchste Kappa- Wert $\kappa = 0,51$ errechnet wurde (siehe Tabelle 10). Damit wäre die Nullhypothese zu verwerfen und der Test reliabel. Nach der Pause hingegen wird nur mehr eine Übereinstimmung von $\kappa = 0,09$ berechnet. Es zeigt sich also in der zweiten Gruppe ($n=23$) eine völlig zufällige Übereinstimmung.

Beim Vorlauftest im Sitz zeigen sich in der Gruppe vor und der Gruppe nach der Pause nicht so deutliche Unterschiede. In der Gruppe Großbuchstaben A bis X wurde ein Kappa- Wert von $\kappa = 0,31$ und in der Gruppe Kleinbuchstaben a bis w ein Kappa- Wert von $\kappa = 0,23$ errechnet. In weiteren Berechnungen konnte festgestellt werden, dass sich der Kappa- Wert auf die Erhebung von symptomatisch und asymptomatisch interessant verhält. Mit $\kappa = 0,42$ ist der Kappa- Wert von den symptomatischen ProbandInnen (siehe Anhang J), laut Landis und Koch (1977) im mittelmäßigen Bereich eingeordnet und nach Fjellner (1999) und McHugh (2012) somit zumindest akzeptabel.

Hinsichtlich der Confidence- Rating- Scale liegt anders als bei der Studie von Cooperstein et al. (2010) in vorliegender Arbeit kein offensichtlicher Mehrwert vor. Mit einem $\kappa = 0,35$ beim Vorlauftest im Stand und einem Wert von $\kappa = 0,16$ beim Vorlauftest im Sitz sind ist das Ergebnis völlig unabhängig von der Sicherheit der Testerinnen. Auch in der Gruppe, in der zumindest eine Testerin unsicher war, gibt es kein auffälliges Ergebnis. Es gibt in Bezug auf die ganze Stichprobe beim Vorlauftest im Stand und beim Vorlauftest im Sitz nur eine geringe Korrelation mit der Sicherheit beziehungsweise Unsicherheit der Testerinnen (0,29/ 0,21). Insgesamt ($n=92$) kam es bei 26 Testungen zu Unsicherheiten, wobei nur bei einem einzigen Probanden beide Raterinnen unsicher waren. Generell scheint die CRS in der vorliegender Arbeit zu vernachlässigen, da diese zwar eine Zusatzinformation gibt, aber das Ergebnis der vorliegenden Arbeit weder primär noch sekundär rechnerisch beeinflusst.

6.3 Bedeutung der Ergebnisse im Kontext der Osteopathie

Eine mäßige Übereinstimmung des durchgeführten Vorlauftests im Stand und des Vorlauftests im Sitz mit einem Kappa- Wert von $\kappa = 0,36$ und $\kappa = 0,28$ ist nicht ausreichend und wird daher als nicht reliabel eingeordnet.

Es wird mit dem Vorlauftest zwar keine Diagnose gestellt, sondern der Vorlauftest im Stand und im Sitz dienen nur als ein Parameter zur Eingrenzung des Untersuchungsbereiches.

Der Vorlauftest ist ein einziges Instrument, welches erst in Kombination mit anderen Werkzeugen eine weitere Aussage ermöglicht. Es darf nicht vergessen werden, dass die

Ergebnisse einer gesamten osteopathischen Untersuchung zahlreich sind und nur ihre Interpretation und Einordnung im Kontext der osteopathischen Prinzipien der Erstellung einer osteopathischen Diagnose dienen (Fossum, 2010, S. 70).

Trotzdem sollte der Einsatz des Tests mit unzureichender Evidenz überlegt werden. Aber auch der Test als einzelner oder gerade dann sollte ein Test gute Evidenz liefern.

Wie unter Punkt eins erwähnt, wird der Vorlauftest im Stand und im Sitz in der globalen Untersuchung eingesetzt. Sämtliche Ergebnisse der vorliegenden Reliabilitätsstudie sind wissenschaftlich gesehen nicht signifikant. Es konnten weder beim Vorlauftest im Stand noch beim Vorlauftest im Sitz statistisch signifikante Werte erzielt werden. Es gibt also jedenfalls verbesserungsbedarf.

Eine klinische Relevanz könnte für OsteopathInnen dennoch gegeben sein. Die Übereinstimmung der Ergebnisse zwischen Testerin A und Testerin B betragen beim Vorlauftest im Stand 76,09% und für den Vorlauftest im Sitz 80,43%. Wie oben beschrieben, verzerrt paradoxon Ergebnisse. McHugh (2012) meint sogar, dass bei gut ausgebildeten und erfahrenen RaterInnen die prozentuelle Übereinstimmung eine sehr gute Aussagekraft besitzt. Hier stellen sich zwei weitere Fragen! Wer gilt als gut ausgebildet? Ab wann gilt man als erfahren?

6.4 Kritische Reflexion des Studienpersonals

Es wurde in der Auswahl der Testerinnen darauf geachtet, dass beide den Grundberuf als Physiotherapeutin erlernt haben und hinsichtlich ihrer Erfahrung als Osteopathin ähnliche Voraussetzungen aufweisen, um den Einflussfaktor Erfahrung möglichst gering zu halten. Anhand der Confidence-rating-scale hat sich herausgestellt, dass in Bezug auf die Sicherheit der Angaben des jeweiligen Testergebnisses sich Testerin A eindeutig häufiger unsicher war als Testerin B. Insgesamt gibt Testerin A 17-mal an in ihrem Ergebnis unsicher zu sein. Das ist fast doppelt so häufig wie Testerin B, die bei 92 Testungen nur 9-mal angibt unsicher zu sein. Testerin B weist als Physiotherapeutin deutlich mehr Berufserfahrung auf. 16 Jahre weniger Erfahrung im Grundberuf und ein Jahr weniger Erfahrung als Osteopathin könnten zu den häufigeren Unsicherheiten von Testerin A geführt haben. Wie unter Punkt 2.5.1 erklärt, ist die Palpation eine maßgebliche Voraussetzung. Mit einer deutlich längeren Berufserfahrung ist Testerin B dabei klar im Vorteil. Anhand der Korrelationsanalyse konnte diesbezüglich zwar kein Zusammenhang festgestellt werden, aber es gibt auch keinen „golden standard“, der das Testergebnis von Raterin A oder Raterin B als richtig oder falsch bestätigen würde. Man kann anhand der Datenanalyse feststellen, dass sich Testerin A und auch Testerin B beim Vorlauftest im Sitz insgesamt deutlich sicherer sind.

Es wäre interessant, wodurch sich die TesterInnen verunsichert fühlen, wenn sie subjektiv die Angabe „unsicher“ machen. Anhand der erhobenen Daten lassen sich jedoch keine

Häufigkeiten oder Merkmale erkennen, die zu einer Unsicherheit führen hätten können. Nur ein einziges Mal waren sich beide Testerinnen bei der gleichen Probandin unsicher. Die Probandin gibt im Fragebogen aktuelle Beschwerden im linken Knie nach einer Knieprothese links an. Die zusätzliche Angabe von Gründen seitens der Raterinnen wäre hilfreich, um weitere Forschung in Richtung Sicherheit zu betreiben. Eine Möglichkeit der Unsicherheit könnte der Einfluss der Beziehung zwischen ProbandIn und Testerin gewesen sein. Auch wenn der Zeitraum der Testuntersuchung nur ein sehr kurzer ist, oder vielleicht auch genau weil dieser Zeitraum des Kontaktes zwischen ProbandIn und RaterIn so kurz ist, könnte die Testdurchführung und Palpation schwieriger als im klassischen Praxisalltag sein. Och (2010, S. 73) schreibt über Veränderungen der Palpation auf Grund der Sympathie von PatientIn und TherapeutIn. Die Studienleiterin empfiehlt für zukünftige Studien ein homogeneres Bild der RaterInnen. Es wird empfohlen, dass nicht nur die Ausbildungsvoraussetzungen der RaterInnen gleich sein sollten, sondern auch deren Jahre der Berufserfahrung. Dies ist besonders im Bereich der Palpation wichtig, da die Erfahrung der Raterinnen eine entscheidende Rolle spielen könnte. Damit könnte eine höhere Vergleichbarkeit und Zuverlässigkeit der Ergebnisse gewährleistet werden.

Durch die Unaufmerksamkeit des Assistenten kam es zu einem Probandenausschluss auf Grund des Alters von 17 Jahren nach Durchführung der Studie. So ein Fehler wäre durch genaue Überprüfung der Daten zu vermeiden. Da genügend ProbandInnen rekrutiert wurden die alle Einschlusskriterien erfüllten, war der im Nachhinein erfolgte Ausschluss für die Ergebnisse nicht relevant. Für die Studienleiterin jedoch war es ein Hinweis, bei folgenden Forschungsprojekten die Aufmerksamkeit diesbezüglich zu erhöhen.

6.5 Kritische Reflexion der Stichprobe

Die Studienleiterin hat sich an die Vorgaben von Patijn (2019) gehalten und daher von einer genauen Fallzahlberechnung abgesehen. Bei den 46 ProbandInnen zeigt sich sowohl bei der Geschlechterverteilung als auch bei der Verteilung Symptomatisch und Asymptomatisch, dass es ein starkes Ungleichgewicht gibt. Es haben sich deutlich mehr Frauen (78,26%) als Männer (21,74%) als ProbandInnen zur Verfügung gestellt. Eine ungleiche Verteilung gibt es auch zwischen symptomatisch und asymptomatisch ProbandInnen. Insgesamt konnten nur elf (23,91%) von 46 ProbandInnen als symptomatisch zugeordnet werden. Dies entspricht nicht der täglichen Praxis von OsteopathInnen. Es würden kaum gesunde Menschen, also Asymptomatische ohne Grunderkrankung, zur Behandlung kommen. Die Anzahl der symptomatischen ProbandInnen ist aber zu gering um einen Schluss daraus zu ziehen. Womöglich hätte eine genaue Fallzahlberechnung die Größe der Stichprobe verändert, somit eventuell auch mehr Männer und symptomatische ProbandInnen in der Studie berücksichtigt. Eine bessere Homogenität anzustreben wäre für weitere Studien wünschenswert. Vielleicht

hätte eine größere Anzahl an ProbandInnen auch die Konfidenzintervalle verkleinert und die Aussagekraft der Ergebnisse verbessert. Laut Patijn (2019) kann eine größere Stichprobe außerdem den P_{index} verändern, was im Falle der vorliegenden Arbeit erstrebenswert wäre. Die ProbandInnenanzahl von 46 war in der durchgeführten Studie zu gering, um den P_{index} im Bereich von 0,5 zu halten. Natürlich kann der P_{index} erst im Nachhinein berechnet und somit auch nicht mehr beeinflusst werden.

6.6 Kritische Reflexion des Studienablaufes

Von einer Verblindung zwischen den Testerinnen und ProbandInnen während der Testuntersuchung im Sinne einer rein zu palpierenden Testung wurde aus zwei Gründen abgesehen. Zum einen erfordert der Vorlauftest im Stand und im Sitz die visuelle Kontrolle der Bewegung der Spinae iliacae posteriores superiores. Zum anderen ist es praxisfern und für die ProbandInnen deutlich unangenehmer, keinen Blickkontakt mit den Testerinnen zu haben. In der vorliegenden Arbeit wurde der Vorlauftest im Stand und im Sitz jeweils einmal von der Raterin durchgeführt. Diese Vorgabe könnte für weitere Studien vernachlässigt werden. Es besteht weder für den Vorlauftest im Stand noch für den Vorlauftest im Sitz die Möglichkeit einer Verzerrung des Ergebnisses durch Mobilisation, wie es im Falle einer Untersuchung eines bestimmten Gelenkes der Fall wäre. Im Praxisalltag würde man den Vorlauftest im Stand und im Sitz bei etwaiger Unsicherheit wahrscheinlich immer ein zweites, vielleicht sogar auch ein drittes Mal durchführen.

Ein durchaus berechtigter Kritikpunkt ist die Einhaltung der Vorgaben von Patijn (2019). Die vorliegende Studie weicht in einzelnen Aspekten von den Empfehlungen Patijns (2019) ab. Die sechste Regel aus den „Golden Rules“ besagt, wie unter Punkt 2.3.2 beschrieben, dass man immer nur eine Testung in einer Reliabilitätsstudie überprüfen soll (Patijn, 2019). Auf Grund der Nähe zum Praxisalltag wurden jedoch beide Testungen, der Vorlauftest im Stand und der Vorlauftest im Sitz, direkt hintereinander von einer Testerin an einem/einer ProbandIn durchgeführt. Eventuell hätten zwei unterschiedliche Testzeitpunkte, einer für den Vorlauftest im Stand und ein weiterer Testzeitpunkt für den Vorlauftest im Sitz gewählt werden sollen, um einen höheren Kappa- Wert zu erzielen. Rückblickend wäre es von Vorteil gewesen die 0,50- P_{index} Methode anzuwenden. Diese mögliche Bias wurde bei der Planung der vorliegenden Arbeit außer Acht gelassen.

7 Konklusio

7.1 Zusammenfassung

Die Untersuchung der Inter- Tester- Reliabilität des Vorlauftests im Stand und im Sitz zeigen sich in der vorliegenden Arbeit als mäßig. Die Kappa-Werte liegen laut Fjellner et al. (1999) und McHugh (2012) unter der akzeptablen Grenze und weit entfernt von beachtlich. Die Nullhypothese wird angenommen, die Alternativhypothese widerlegt. Der Vorlauftest ist weder im Stand noch im Sitz hinreichend reliabel. Die Ergebnisse stellen wissenschaftlich gesehen keine ausreichende Begründung für eine evidenzbasierte Anwendung dar.

Klinisch gesehen haben zwei Testerinnen bei einer Stichprobengröße von 46 ProbandInnen beim Vorlauftest im Stand 35 Mal (76,09%) und beim Vorlauftest im Sitz 37 Mal (80,43%) übereingestimmt. Der niedrige Kappa- Koeffizient weist auf ein Kappa Paradoxon hin.

Vom klinischen Aspekt weiter interessant ist der höhere Kappa- Wert von ausschließlich symptomatisch ProbandInnen bei der explorativen Datenauswertung. Die Anzahl der symptomatischen ProbandInnen ist in vorliegender Arbeit allerdings zu gering, um eine wissenschaftliche Aussage tätigen zu können.

In Bezug auf die Confidence- Rating- Scale können in dieser Arbeit keine wissenschaftlich relevanten Schlüsse gezogen werden.

7.2 Schlussfolgerung

Insgesamt zeigen die Ergebnisse, dass die Inter- Rater- Reliabilität beider Vorlauftestungen verbessert werden muss, um verlässliche und konsistente Testergebnisse zu erhalten.

Der sehr häufige und weit verbreitete Einsatz des Vorlauftests im Stand und im Sitz in der „Untersuchungsroutine“ sollte aus wissenschaftlicher Sicht hinterfragt werden. Auch das Lehren des Vorlauftests in der Grundausbildung der Osteopathie steht somit in Frage.

Der klinische Wert des Vorlauftests im Stand und im Sitz könnte trotzdem gegeben sein. Einerseits kann durch Optimierung der Studie die Auswirkung des Kappa Paradoxon auf die Inter-Tester-Reliabilität reduziert werden. Andererseits lassen die vorliegenden Ergebnisse eine veränderte Inter- Tester- Reliabilität an einer rein symptomatischen ProbandInnengruppe vermuten.

7.3 Ausblick

Um eine wissenschaftlich signifikante Aussagekraft des Vorlauftests im Stand und im Sitz zu erhalten, sollte bereits beim Probedurchlauf einen P_0 von 0,85 angestrebt werden, um einer Verzerrung des Kappa- Wertes durch einen zu geringen oder zu hohen P_{index} entgegenzuwirken. Es könnte überlegt werden, die $0,50-P_{index}$ Methode anzuwenden.

Außerdem sollte, wie auch schon von Patijn (2019) empfohlen, in folgenden Studien nur ein Test zu einem Zeitpunkt durchgeführt werden.

Die Confidence- Rating- Scale ist klinisch gesehen interessant. Genauere Angaben bei der subjektiven Angabe „unsicher“ wären für folgende Studien aus Sicht der Autorin wichtig um Vorkehrungen treffen zu können, etwaige Unsicherheiten seitens der TesterInnen bereits im Vorfeld verringern zu können.

Die Untersuchung der Intra- Rater- Reliabilität wäre für den Einsatz des Vorlauftests im Stand und im Sitz ebenso sehr interessant, da der Vorlauftest in der täglichen Praxis häufig als Wiederbefundungsparameter eingesetzt wird und dies den Nutzen des Vorlauftests in anderer Hinsicht wissenschaftlich belegen könnte.

Weitere Forschung und Reliabilitätsuntersuchungen im Bereich der osteopathischen Tests sind für die „Evidenced- based -practice“ auf jeden Fall unumgänglich und notwendig, um den OsteopathInnen Grundlagen in der Argumentation im medizinischen Bereich zu schaffen.

Literaturverzeichnis

- Biberschick, M. (2010). „Legt euch ein Schema zurecht!“. Die Routineuntersuchung in der Osteopathie. [Masterarbeit, Donau- Universität Krems]
<https://www.osteopathicresearch.com/files/original/50e652bbe212498038d974a52352de4fe6e2e87e.pdf>
- Boddenberg, M.-A. (2013). Ist ein positiver Vorlaufstest (Standing- flexion- test) durch das Einlegen von Zahnwatterollen zwischen die Zähne bei Patienten mit einer Kiefergelenkpathologie veränderbar? [Masterarbeit, Donau- Universität Krems]
<https://www.osteopathicresearch.com/files/original/749bc422f8277a67f4529e57ca0318d1dd118d4f.pdf>
- Bordoni, B. (2019). The Benefits and Limitations of Evidence-based Practice in Osteopathy. *Cureus*, 11(11). <https://doi.org/10.7759/cureus.6093>
- Ciranna- Raab, C. (2010). Das Tensegrity- Modell. In: Liem, T., Dobler, T. K.. Leitfaden Osteopathie. München: Urban & Fischer Verlag/ Elsevier GmbH.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46.
<https://doi.org/10.1177/001316446002000104>
- Cooperstein, R., Haneline, M., Young, M. (2010). Interexaminer reliability of thoracic motion palpation using confidence ratings and continuous analysis. *Journal of Chiropractic Medicine*, 9(3), 99-106. <https://doi./10.1016/j.jcm.2010.06.004>
- Corts, M. (2020). *Diagnoseleitfaden Osteopathie* (3. Aufl.). Georg Thieme Verlag KG
- Croibier, A. (2006). Diagnostik in der Osteopathie. Jena, München: Urban & Fischer Verlag

- Derksen, B. M., Bruinsma, W., Goslings, J. C., Schep, N. W. L. (2024). The Kappa Paradox Explained. *The Journal of hand surgery*. <https://doi.org/10.1016/j.jhsa.2024.01.006>
- Egan, D., Cole, J., & Twomey, L. (1996). The Standing Forward Flexion Test: An Inaccurate Determinant of Sacroiliac Joint Dysfunction. *Physiotherapy*, 82(4), 236–242. [https://doi.org/10.1016/S0031-9406\(05\)66879-4](https://doi.org/10.1016/S0031-9406(05)66879-4)
- Fjellner, A., Bexander, C., Faleij, R., Strender, L. E. (1999). Interexaminer reliability in physical examination of the cervical spine. *Journal of Manipulative and Physiological Therapeutics* 22. [https://doi.org/10.1016/s0161-4754\(99\)70002-3](https://doi.org/10.1016/s0161-4754(99)70002-3)
- Fossum, C. (2010). Die osteopathische Diagnosefindung. In: Liem, T., Dobler, T. K.. Leitfaden Osteopathie. München: Urban & Fischer Verlag/ Elsevier GmbH.
- Freyer, G., McPherson, H. C., & Keefe, P. O. (2005). The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. *International Journal of Osteopathic Medicine*, 8, 131-138. <https://doi.org/10.1016/j.ijosm2005.08.004>
- Hussy, W., Schreier, M., & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (2. Aufl.). Springer Verlag. <https://doi.org/10.1007/978-3-642-34362-9>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Liem, T. & Dobler, T. K., (2010). *Leitfaden Osteopathie: Parietale Techniken* (3. Aufl.). Urban & Fischer Verlag/Elsevier GmbH.

- Liem, T. (2014). Pitfalls and challenges involved in the process of perception and interpretation of palpatory findings. *International Journal of Osteopathic Medicine*, Volume 17, Issue 4, 243-249. <https://doi.org/10.1016/j.ijosm.2014.04.005>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. (online Zugriff am 12.10.2023)
- Meert, G. F. (2017). *Das Becken aus osteopathischer Sicht* (4. Auflage). Urban & Fischer Verlag/Elsevier GmbH.
- Myers, T. W. (2004). *Anatomy Trains: Myofasziale Meridiane* (1. Auflage). Urban & Fischer Verlag/Elsevier GmbH.
- Nolet, P. S., Yu, H., Côté, P., Meyer, A.-L., Kristman, V. L., Sutton, D., Murnaghan, K., & Lemeunier, N. (2021). Reliability and validity of manual palpation for the assessment of patients with low back pain: a systematic and critical review. *Chiropractic & Manual Therapies*, 29 (33). <https://doi.org/10.1186/s12998-021-00384-3>
- Patijn, J. (2019). Reproducibility protocol for diagnostic procedures in Manual/Musculoskeletal Medicine: Edition 2019. *Manuelle Medizin*, 57(6), 451–479. <https://doi.org/10.1007/s00337-019-00581-5>
- Paulsen, F. & Waschke, J. (2010). *Sobotta: Atlas der Anatomie des Menschen* (23. Aufl.). Urban & Fischer Verlag/ Elsevier GmbH.
- Ribeiro, R. P., Guerrero, F. G., Camargo, E. N., Pivotto, L. R., Aimi, M. A., Loss, J. F., & Candotti, C. T. (2021). Construct validity and reliability of tests for sacroiliac dysfunction: Standing flexion test (STFT) and sitting flexion test (SIFT). *Journal of Osteopathic Medicine*, 121(11), 849–856. <https://doi.org/10.1515/jom-2021-0025>

- Rittler, M. (2010). Ist das global Listening reliabel: Studie zur Intrarater- und Interraterreliabilität des Global Listenings [Masterarbeit, Donau- Universität Krems]
<https://www.osteopathicresearch.com/files/original/aa2eb6ea2714439ab8130506ad051ee11d24e18e.pdf>
- Scherfer, E., & Bossmann, T. (2011). *Forschung verstehen: Ein Grundkurs in evidenzbasierter Praxis* (2. Aufl.). Pflaum Verlag.
- Och, V. (2010). Palpation. In: Liem, T., Dobler, T. K.. Leitfaden Osteopathie. München: Urban & Fischer Verlag/ Elsevier GmbH.
- O´Haire, C., & Gibbons, P. (2000). Inter-examiner and intra- examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. *Manual Therapy, Volume 5, Issue 1, 13-20*. <https://doi.org/10.1054/math.1999.0203>
- Vincent- Smith, B., Gibbons, P. (1999). Inter- examiner and intra- examiner reliability of the standing flexion test. *Manual Therapy, 4 (2), 87-93*.
<https://doi.org/10.1054/math.1999.0173>
- Vincent, L. (2010). Diagnostik. In: Liem, T., Dobler, T. K.. Leitfaden Osteopathie. München: Urban & Fischer Verlag/ Elsevier GmbH.
- Weiß, C. (2013). *Basiswissen Medizinische Statistik*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-34261-5>

Tabellenverzeichnis

TABELLE 1 KLASSIFIKATION DER COHENS- KAPPA- KOEFFIZIENTEN ADAPTIERT NACH LANDIS UND KOCH (1977) ÜBERSETZT VON DER AUTORIN	5
TABELLE 2 ERGEBNISSE PROBELAUF	19
TABELLE 3 GESCHLECHTERVERTEILUNG	22
TABELLE 4 ALTERSVERTEILUNG	22
TABELLE 5 VERTEILUNG SYMPTOMATISCH/ ASYMPTOMATISCH.....	23
TABELLE 6 VERTEILUNG GRUNDERKRANKUNGEN	24
TABELLE 7 KLASSIFIKATION DER COHENS KAPPA KOEFFIZIENTEN- QUELLE: ADAPTIERT NACH LANDIS UND KOCH (1977).....	24
TABELLE 8 KREUZTABELLE GESAMTERGEBNIS VORLAUFTEST STAND	26
TABELLE 9 DARSTELLUNG ERGEBNISSE VORLAUFTEST STAND.....	26
TABELLE 10 ERGEBNISSE EXPLORATIVE ANALYSE VORLAUFTEST STAND.....	27
TABELLE 11 KREUZTABELLE GESAMTERGEBNIS VORLAUFTEST SITZ	27
TABELLE 12 DARSTELLUNG ERGEBNISSE VORLAUFTEST SITZ	27
TABELLE 13 ERGEBNISSE EXPLORATIVE ANALYSE VORLAUFTEST IM SITZ	28
TABELLE 14 KORRELATION PRIMÄRER UND SEKUNDÄRER ZIELPARAMETER	31

ABBILDUNGSVERZEICHNIS

ABBILDUNG 1 AKQUISITATION PROBANDINNEN	16
ABBILDUNG 2 ERGEBNIS IN ABSOLUTEN ZAHLEN VORLAUFTEST STAND	25
ABBILDUNG 3 ERGEBNIS IN ABSOLUTEN ZAHLEN VORLAUFTEST SITZ.....	25
ABBILDUNG 4 ERGEBNISSE CRS IN ABSOLUTEN ZAHLEN	29
ABBILDUNG 5 ÜBERSCHNEIDUNG SICHERHEIT GESAMT (N=92) TESTERIN A/ TESTERIN B ..	29
ABBILDUNG 6 ÜBERSCHNEIDUNG CRS VORLAUFTEST STAND (N=46).....	30
ABBILDUNG 7 ÜBERSCHNEIDUNG CRS VORLAUFTEST SITZ (N=46)	30

ABKÜRZUNGSVERZEICHNIS

CRS	Confidence Rating Scale
DM	Diabetes mellitus
EBP	Evidence based practice
κ	Kappa- Koeffizient nach Cohen
M.	Musculus
P_0	Relative Übereinstimmung
P_e	Zufällige Übereinstimmung
SD	Standardabweichung
SIPS	Spina iliaca posterior superior
SIPI	Spina iliaca posterior inferior

ANHANG A: LITERATURRECHERCHE

Die Literaturrecherche wurde von Februar 2023 bis Mai 2024 durchgeführt. Folgende Keywords wurden für die Recherche eingesetzt. Die Keywords „Reliability“, „Reliabilitätsstudie“, „Gütekriterien“, „validity“, „kappa koeffizient“, „Vorlaufstest im Stand“, „Vorlaufstest im Sitz“, „standing forward flexion test“, „sitting flexion test“, „Routineuntersuchung“, „clinical reasoning“, „global examination“ wurden dabei einzeln verwendet, aber auch unterschiedlich kombiniert. Es wurde in den Datenbanken PubMed (<https://pubmed.ncbi.nlm.nih.gov>), Google Scholar (<https://scholar.google.com>), Osteopathic research web (www.osteopathic-research.com), PEDro- Physiotherapie Evidenz Datenbank (www.pedro.org) und in diversen Fachzeitschriften nach Literatur gesucht. Ergänzend wurden Fachbücher aus der Medizin und Bücher zum Thema wissenschaftlichen Arbeiten und Statistik herangezogen. Die Recherche wurde während dem Verfassen der Master- Thesis kontinuierlich fortgesetzt. Verwendet wurden deutsch- und englischsprachige Arbeiten. Gefundene Literatur wurde mittels Schneeballsystem ergänzt.

ANHANG B: PROBANDINNENAUFTRUF

ICH, Tanja Flicker- Niederl, BSc.,



suche SIE!

Im Rahmen meiner Master- Thesis im Bereich Osteopathie suche ich ProbandInnen zur Durchführung meiner Studie. (Zwei kurze Tests am Becken.)

Zeitaufwand ca **15 Minuten**
KEINE KOSTEN!

Am **14.01.2024** **08.00** bis **12.30 Uhr**

In der **Praxis XUNDOS**
Josefa-Posch-Straße 3
8200 Gleisdorf

Physiotherapeutin und Osteopathin **Tanja Flicker- Niederl**
(0664/4160493).

Mehr Informationen finden Sie unter **www.natuerlich-bewegt.at**

ANHANG C: ZEITEINTEILUNG

	ProbandIn 1	ProbandIn 2		
08:00				
08:10				
08:20				
08:30				
08:40				
08:50				
09:00				
09:10				
09:20				
09:30				
09:40				
09:50				
Pause				
10:30				
10:40				
10:50				
11:00				
11:10				
11:20				
11:30				
11:40				
11:50				
12:00				
12:10				
12:20				

ANHANG D: GESUNDHEITSFRAGEBOGEN PROBANDIN

Gesundheitsfragebogen

	JA	NEIN
1) Ich habe in den letzten 6 Wochen eine Hüftprothese bekommen	<input type="checkbox"/>	<input type="checkbox"/>
2) Ich habe eine Grunderkrankung (Diabetes Mellitus, Rheuma, Asthma, Herz- Kreislauferkrankung, Osteoporose,...) _____	<input type="checkbox"/>	<input type="checkbox"/>
3) Ich habe derzeit Beschwerden (Schulterschmerzen, Kopfschmerzen, Infekt,...) Symptome: _____ _____	<input type="checkbox"/>	<input type="checkbox"/>

Inspektion (durch die Studienleitung)

Offene Wunden im Bereich des Beckens

JA

NEIN

ANHANG E: EINVERSTÄNDNISERKLÄRUNG PROBANDIN

EINVERSTÄNDNISERKLÄRUNG ZUR STUDIENTEILNAHME

DIE INTERTESTER-RELIABILITÄT DES VORLAUFTESTS IM STAND UND IM SITZ

Studienleiterin: Tanja Flicker- Niederl

Studienort: Praxis XUNDOS, Josefa-Posch-Straße 3, 8200 Gleisdorf

NAME:

GESCHLECHT:

GEBURTSDATUM:

Hiermit bestätige ich, ... (BITTE ANKREUZEN)

- ... dass ich das Informationsblatt zur Studie gelesen und verstanden habe.
- ... dass ich über den Ablauf der Studie Bescheid weiß und von der Studienleitung ausreichend aufgeklärt wurde.
- ... dass meine Teilnahme auf freiwilliger Basis beruht.
- ... dass meine personenbezogenen Angaben und der beiliegende Fragebogen wahrheitsgemäß beantwortet wurden.
- ... dass ich der anonymen Speicherung meiner persönlichen Daten zustimme.

Ort, Datum

Unterschrift ProbandIn

ANHANG F: ERGEBNISBOGEN GRUPPE 1 BIS ZUR
PAUSE

Testergebnisse TesterIn A

ProbandIn		Auffällig/ Positiv	Unauffällig/ Negativ	Sicher	Unsicher
A	<u>Stand</u>				
	<u>Sitz</u>				
B	<u>Stand</u>				
	<u>Sitz</u>				
C	<u>Stand</u>				
	<u>Sitz</u>				
D	<u>Stand</u>				
	<u>Sitz</u>				
E	<u>Stand</u>				
	<u>Sitz</u>				
F	<u>Stand</u>				
	<u>Sitz</u>				
G	<u>Stand</u>				
	<u>Sitz</u>				
H	<u>Stand</u>				
	<u>Sitz</u>				
I	<u>Stand</u>				
	<u>Sitz</u>				
J	<u>Stand</u>				
	<u>Sitz</u>				
K	<u>Stand</u>				
	<u>Sitz</u>				
L	<u>Stand</u>				
	<u>Sitz</u>				
M	<u>Stand</u>				
	<u>Sitz</u>				
N	<u>Stand</u>				
	<u>Sitz</u>				
O	<u>Stand</u>				
	<u>Sitz</u>				
P	<u>Stand</u>				
	<u>Sitz</u>				

Q	<u>Stand</u>				
	<u>Sitz</u>				
R	<u>Stand</u>				
	<u>Sitz</u>				
S	<u>Stand</u>				
	<u>Sitz</u>				
T	<u>Stand</u>				
	<u>Sitz</u>				
U	<u>Stand</u>				
	<u>Sitz</u>				
V	<u>Stand</u>				
	<u>Sitz</u>				
W	<u>Stand</u>				
	<u>Sitz</u>				
X	<u>Stand</u>				
	<u>Sitz</u>				

ANHANG G: ERGEBNISBOGEN GRUPPE 2 AB DER
PAUSE

Testergebnisse TesterIn A

ProbandIn		Auffällig/ Positiv	Unauffällig/ Negativ	Sicher	Unsicher
a	<u>Stand</u>				
	Sitz				
b	<u>Stand</u>				
	Sitz				
c	<u>Stand</u>				
	Sitz				
d	<u>Stand</u>				
	Sitz				
e	<u>Stand</u>				
	Sitz				
f	<u>Stand</u>				
	Sitz				
g	<u>Stand</u>				
	Sitz				
h	<u>Stand</u>				
	Sitz				
i	<u>Stand</u>				
	Sitz				
j	<u>Stand</u>				
	Sitz				
k	<u>Stand</u>				
	Sitz				
l	<u>Stand</u>				
	Sitz				
m	<u>Stand</u>				
	Sitz				
n	<u>Stand</u>				
	Sitz				
o	<u>Stand</u>				
	Sitz				
p	<u>Stand</u>				
	Sitz				

	<u>Sitz</u>				
q	<u>Stand</u>				
	Sitz				
r	<u>Stand</u>				
	Sitz				
s	<u>Stand</u>				
	Sitz				
t	<u>Stand</u>				
	Sitz				
u	<u>Stand</u>				
	Sitz				
v	<u>Stand</u>				
	Sitz				
w	<u>Stand</u>				
	Sitz				

ANHANG H: VARIABLEN UND KODIERUNG FÜR MICROSOFT EXCEL

ProbandInnenidentifikation: Die ProbandInnen bekamen im Rahmen ihrer Datenerhebung einen Groß- beziehungsweise Kleinbuchstaben zur Identifikation zugeteilt. In alphabetischer Reihenfolge wurden diese anhand ihres Eintreffens chronologisch zugeteilt. Es gab eine Verteilung von Großbuchstaben A bis X und eine Verteilung von Kleinbuchstaben a bis w.

Alter: Anhand des angegebenen Geburtsdatums auf dem Datenbogen wurde das Alter der ProbandInnen in Jahren zum Tag der Studiendurchführung berechnet. Das metrische Messniveau wurde in ganzen Jahren in einer Excel- Tabelle eingefügt.

Geschlecht: Die Frage nach dem Geschlecht konnte frei beantwortet werden. Es gab keinerlei Auswahlmöglichkeiten, wurde aber ausschließlich mit weiblich und männlich angegeben. Es besteht ein nominales Messniveau. Für die Berechnungen wurde weiblich mit der Zahl eins und männlich mit der Zahl zwei kodiert.

Symptomatisch/ Asymptomatisch: Anhand des kurzen Fragebogens wurden aktuelle Beschwerden abgefragt und erhoben. Bei der Beantwortung der Frage mit Nein galt die Person als asymptomatisch und wurde mit der Kodierung eins versehen. Wurde die Frage mit Ja beantwortet, wurde die Person als symptomatisch zugeordnet und mit zwei kodiert. Es gilt das nominale Messniveau.

Grunderkrankung: Die Möglichkeit mit Ja oder Nein zu antworten weist auch bei der Frage nach Grunderkrankungen das nominale Messniveau vor. Keine Grunderkrankung, also Antwort Nein, wurde mit der Zahl eins kodiert. Antwort Ja, Grunderkrankung vorhanden, wurde mit der Zahl zwei kodiert.

Testergebnis: Am Ergebnisbogen hatten die Testerinnen die Wahl das Ergebnis mit Unauffällig (Negativ) oder Auffällig (Positiv) zu bewerten. Unauffällig wurde mit der Zahl eins kodiert. Auffällig mit der Zahl zwei. Das nominale Messniveau wird angegeben.

Sicherheit des Ergebnisses der einzelnen Rater: Die Testerinnen mussten zusätzlich angeben, wie sicher sie sich in Bezug auf die Angabe ihres Testergebnisses sind. Auch hier wurde ein nominales Messniveau eingesetzt. Die Antwortmöglichkeit sicher wurde mit der Kodierungszahl eins versehen. Unsicher wurde die Zahl zwei zugeordnet.

Gesamtsicherheit: Zur Berechnung des Kappa Koeffizienten bei jenen Testungen, bei denen sich beide Rater sicher waren, sowie bei weiteren Konstellationen wurde eine eigene

Kategorie „Gesamtsicherheit“ erstellt. Die codierten Zahlen ergaben sich folgendermaßen: Waren sich beide Rater bei einem einer ProbandIn „sicher“ wurde die Gesamtsicherheit mit eins codiert. War sich ein Testender „sicher“, der andere „unsicher“, wurde die Zahl zwei vergeben. Waren sich beide Rater bei demselben/ derselben ProbandIn „unsicher“ wurde die Zahl drei vergeben. Das verwendete Messniveau ist nominal.

ANHANG I: KREUZTABELLENVORLAGE

Kreuztabelle					
		Testerin B		Gesamt	
		auffällig	unauffällig		
Testerin A	auffällig	Anzahl	a	b	a + b
		%	%	%	%
	Gesamtanzahl				c + d
	unauffällig	Anzahl	c	d	c + d
%		%	%	%	
Gesamtanzahl				%	
Gesamt	Anzahl		a + c	b + d	n
	%		%	%	%
	Gesamtanzahl				%
			Übereinstimmung	%	
			Keine Übereinstimmung	%	

ANHANG J: WEITERE BERECHNUNGEN

ERGEBNISSE EXPLORATIVE ANALYSE VORLAUFTEST STAND

	Cohens-Kappa Koeffizient [-1 bis +1]	Relative Übereinstimmung P_0 [%]	Zufällige Übereinstimmung P_e [%]	CRS Übereinstimmung [%]	p- Wert
Frauen n= 36	$\kappa= 0,41$	77,78%	62,5%	69,4%	nicht ermittelbar
Männer n=10	$\kappa= 0,21$	70%	62%	70%	p=0,6060
Symptomatisch n= 11	$\kappa= 0,42$	72,73%	52,89%	45,5%	p=0,1697
Asymptomatisch n= 35	$\kappa= 0,36$	77,14%	64,57%	77,1%	p=0,5692
Grunderkrankung n =7	$\kappa= 0,28$	57,14%	40,82%	71,4%	p=0,1088
Keine Grunderkrankung n= 39	$\kappa= 0,38$	79,49%	66,86%	69,2%	p=0,2620

ERGEBNISSE EXPLORATIVE ANALYSE VORLAUFTEST IM SITZ

	Cohens-Kappa Koeffizient [-1 bis +1]	Relative Übereinstimmung P_0 [%]	Zufällige Übereinstimmung P_e [%]	CRS Übereinstimmung [%]	p- Wert
Frauen n= 36	$\kappa= 0,37$	86,11%	78,09%	86,1%	p= 0,7205
Männer n=10	$\kappa= 0,09$	60%	56%	70%	p=0,3292
Symptomatisch n= 11	$\kappa= 0,0$	90,91%	90,91%	81,8%	p=0,3060
Asymptomatisch n= 35	$\kappa= 0,29$	77,14%	68%	82,9%	nicht ermittelbar
Grunderkrankung n =7	$\kappa= 1$	100%	59,18%	85,7%	nicht ermittelbar
Keine Grunderkrankung n= 39	$\kappa= 0,05$	76,92%	75,74%	82,1%	p=0,7471

ANHANG K: KORRELATIONEN

Der Zusammenhang von Variablen wird untersucht. Folgendes konnte berechnet werden.

ÜBERSICHT KORRELATIONEN

Korrelationsvariablen	Korrelationsfaktor (Vorlaufzeit im Stand)	Korrelationsfaktor (Vorlaufzeit im Sitz)
Übereinstimmung auffällig und unauffällig Tester A und B/ Übereinstimmung sicher und unsicher Tester A und B in der Gruppe A bis X	0,11	0,34
Übereinstimmung auffällig und unauffällig Tester A und B/ Übereinstimmung sicher und unsicher Tester A und B in der Gruppe a bis w	0,47	0,1
Testerin A / Geschlecht	0,08	0,07
Testerin B / Geschlecht	0,05	0,31
Tester A / Gruppe symptomatisch und asymptomatisch	0,28	0,24
Testerin B / Gruppe symptomatisch und asymptomatisch	0,08	0,12
Testerin A mit Gruppe der Grunderkrankungen oder keine Grunderkrankungen	0,05	0,16
Testerin B mit Gruppe der Grunderkrankungen oder keine Grunderkrankungen	0,47	0,12

ANHANG L: INFORMATIONSBLATT

Informationsblatt zur Studienteilnahme

Sehr geehrte Teilnehmerin! Sehr geehrter Teilnehmer!

Vielen lieben Dank, dass Sie sich dafür bereit erklärt haben, an der Studie zu meiner Masterarbeit mit dem Thema

„DIE INTERTESTER-RELIABILITÄT DES VORLAUFTESTS IM STEHEN UND IM SITZ “

teilzunehmen!

Die Ausführung des Vorlauftests ist für die Befundung von PatientInnen in der Osteopathie von Relevanz, da bereits zu Beginn Bereiche eingegrenzt und diese in weiterer Folge genauer untersucht werden können. Zusätzlich ist es eine gute Möglichkeit einer zeitsparenden Wiederbefundungsmaßnahme im Rahmen einer Behandlung. Im Rahmen dieser Arbeit soll herausgefunden werden, ob zwei OsteopathInnen bei der Testung zu übereinstimmenden Ergebnissen kommen.

ABLAUF:

Sie stellen sich hüftbreit hin. Die Osteopathin ertastet mit den Daumen zwei Knochenvorsprünge (SIPS) am Becken. Sie werden nun aufgefordert, die Hände Richtung Füße (Boden) zu bewegen. Danach wird das Ganze im hüftbreiten Sitz auf der Liege ausgeführt.

UM KEINEN EINFLUSS AUF DAS TESTERGEBNIS ZU NEHMEN, DÜRFEN SIE
BITTE NICHT MIT DEN OSTEOPATHINNEN SPRECHEN!

Vielen herzlichen Dank für Ihre Unterstützung!

Tanja Flicker- Niederl

ANHANG M: ENGLISCHE KURZFASSUNG

Inter- rater- reliability of the standing and sitting forward flexion test in symptomatic and asymptomatic adults

Tanja Flicker- Niederl, BSc

Andreas Brandl, MSc.

Danube University Krems – Department for Health Sciences and Biomedicine

Correspondence Address

Name: Tanja Flicker- Niederl, BSc

Address: Weizerzeil 152, 8160 Weiz

E-Mail address: tanja@natuerlich-bewegt.at

Phone number: 0664/6140493

Curriculum Vitae

The author Tanja Flicker- Niederl, BSc, attended the education for physiotherapy at the FH Joanneum Graz from 2008 until 2011. She started to work as a physiotherapist at the rehabilitation centre Judendorf- Straßengel in Styria in July 2012. Since 2016 she is selfemployed in her private practice for physiotherapy and osteopathy in 8184 Anger.

The author started the education for osteopathy in 2015 at the Vienna School of Osteopathy WSO. Currently, she is attending the master programme at the Danube University Krems.

ABSTRACT

Inter- rater reliability of the standing and sitting forward flexion test in symptomatic and asymptomatic adults

Objective: The standing and sitting forward flexion test is a common method in the global examination of osteopathic patients. Currently, the validity and reliability of the forward flexion test for evaluating dysfunction of the lumbosacral joint are not well- established. This study aims to assess the inter- rater reliability of the standing and sitting forward flexion test as a tool in the global examination, to determine the level of agreement between two independent osteopathic practitioners.

Methods: This study employed an inter- rater reliability design, where two independent osteopathic practitioners examined 47 symptomatic and asymptomatic adults using the standing and sitting forward flexion test. Each examiner also had to indicate their confidence in the test results using a rating scale of “confident” or “not confident”.

Results: Cohen’s Kappa coefficients were calculated for the standing and sitting forward flexion tests, resulting in $\kappa = 0.36$ for the standing forward flexion test and $\kappa = 0.28$ for the sitting flexion test, indicating fair agreement. There was a low correlation (0.29 and 0.21) between the test results and the confidence- rating- scale.

Conclusion: Based on the fair agreement observed in the standing and sitting forward flexion tests, neither can be recommended as a standalone test in the global examination of osteopathic practitioners. Further optimization of the study is needed to address the Kappa paradox. The results of the standing and sitting forward flexion test were not influenced by the confidence level of the raters.

Keywords: inter- rater reliability, standing forward flexion test, sitting forward flexion test, global examination, osteopathy

INTRODUCTION

For the osteopathic diagnosis, it is important to carry out a detailed anamnesis and a meaningful clinical examination. The clinical examination and thus the selection of tests remains the responsibility of the individual osteopath (Fossum, 2010).

The lack of evidence from the standing and sitting forward flexion test was chosen as an incentive for carrying out and writing the master's thesis. Nothing could be found in the literature that justifies the use of the standing and sitting forward flexion test.

It should be noted that the following work is not a specific test for the sacroiliac joint. It was discovered in 1996 and 1999 and in 2021 by Ribeiro et al. confirms that a positive forward flexion test is not sufficient evidence of dysfunction in the SIJ (Egan et al., 1996; Ribeiro et al., 2021; Vincent-Smith & Gibbons, 1999).

In addition to the primary target parameter, it should be found out whether there is a correlation between the self- assessment and the inter- tester- reliability. In addition, the raters' self- assessment of the tested result, with safe and uncertain, is also asked with the Confidence rating scale (CRS).

METHODS

Study design

The present study is a methodological study. More precisely it is an inter- rater- reliability study design. The agreement between two testers in assessing the abnormality of the standing and sitting forward flexion test is checked.

Sample

All people of all genders aged 18 years and over were included in the study. The prerequisite was the independent execution of the specified starting positions.

Symptomatic and asymptomatic individuals were admitted to the study.

People for whom palpation of the posterior superior spine iliac was not possible due to injuries (e.g. open wounds) and people with recent hip prostheses (up to 6 weeks postoperatively) would have been excluded from the study.

48 people were recruited. One test subject was in the wrong location. She was excluded from the study so that time management was not affected. Another person was subsequently excluded from the study due to age. The gender distribution and age distribution including standard deviation (SD) can be found in Tables 1 and 2.

TABLE 1 SEX

gender (n= 46)	n	%
female	36	78.26
male	10	21.74

TABLE 2 AGE

Age [years]	minimum	maximum	median	average	SD
(n = 46)	22	77	42	44.33	14.3

TABLE 3 SYMPTOMATIC/ ASYMPTOMATIC

(n=46)	n	[%]
symptomatic	11	23.91%
asymptomatic	35	76.09%

TABLE 4 UNDERLYING DISEASE

underlying disease	(n=46)	female	male
Yes	7	7	0
No	39	29	10

Raters

The prerequisite for the study director was that both testers had successfully completed their training as osteopaths at the Vienna School of Osteopathy. The basic occupation as a physiotherapist was also an inclusion criterion. Both osteopaths have completed the master's course "Osteopathy MSc." in cooperation with the Danube University Krems.

Assistant

The assistant's task was to provide the prepared questionnaires. Checking all the fields to be filled out in the data and questionnaire, as well as the signature, were also the assistant's job.

Blinding

The subjects and testers were instructed verbally and in writing not to communicate with each other more than necessary.

Blinding between the two testers during the room change in the aisle was ensured by the supervision of the study director.

Training period

Both osteopaths adapted the execution of the standing and sitting forward flexion test in daily practice to the instructions of the study director and up to the meeting practiced or used in everyday work.

The aim was to achieve palpation and execution that was as identical as possible. The standing and sitting forward flexion test were carried out and all details such as time management, blinding and randomization were discussed. An implementation that was as similar as possible to the actual study was encouraged.

The results of the test run were satisfactory both organizationally and mathematically.

A further training phase was therefore refrained from.

Examination

The forward flexion test is carried out by osteopaths first from standing position and then from sitting position. Every ten minutes two participants were scheduled. They were randomly assigned by the assistant to one of the two rooms.

The testers asked the subjects taking part in the study to take off their shoes and pants.

A one-time implementation of the test was agreed upon in advance with the testers.

The execution of the the standing and sitting forward flexion test adheres to the specifications described in the guidelines for osteopathy without any deviations (Liem & Dobler, 2010, pp. 397-400). The result is normal if both posterior superior iliac spines begin and stop moving superiorly simultaneously during the flexion movement. The result is noticeable when one side moves significantly further superiorly.

After carrying out the two tests, the test subject was asked to stand up and take a few steps across the room. In the meantime, the raters changed the two therapy rooms under the supervision of the study director.

Each rater carried his list of results with him when changing rooms.

The lists of results from the two testers were collected by the study director before the scheduled break and arranged directly in a folder. After the break, new result lists were handed out. After the last test, the result lists were again collected by the study director and immediately placed in a folder by.

STATISTICS

The data from the tests carried out in the present study are presented using a cross table. For the evaluation of a reliability study with a nominal measurement level, the reliability is specified with the kappa coefficient (Patijn, 2019).

The interpretation of the kappa coefficient follows the classification of Landis and Koch (1977).

RESULTS

All values in absolute numbers and percentages for the standing forward flexion test can be found in Tables 5 and 6. The 95% confidence interval is reported to be 0.057 to 0.668. Regarding the CRS, rater A is confident in 78.3% of her statements. Rater B is 84.8% confident in reporting the results.

For the sitting forward flexion test the results can be found in table 7 and 8. The 95% confidence interval is between -0.069 and 0.637. During the sitting forward flexion test, tester A is 87% confident in her statement of the result. In percentage terms, rater B is 95.7% certain about the result.

In absolute numbers it can be seen that overall, i.e. the standing and sitting forward flexion test (n=92), tester A is almost twice as likely to be unsure as tester B.

TABLE 5 CROSS TABLE OVERALL RESULTS STANDING FORWARD FLEXION TEST

Cross table Rater A / Rater B – overall results for standing forward flexion test					
		Rater B		overall	
		conspicuous	inconspicuous		
Rater A	conspicuous	Noticed number	6	6	12
		in %	13.04%	13.04%	26.08%
Rater A	inconspicuous	Noticed number	5	29	34
		in %	10.87%	63.04%	73.91%
Total		Noticed number	11	35	46
		in %	23.91%	76.09%	100%
		Agreement		76.09%	
		No match		23.91%	

TABLE 6 PRESENTATION OF RESULTS OF THE STANDING FORWARD FLEXION TEST

Kappa- coefficient	0.36
P ₀	76.09%
P _{Index}	0.25
p- value	0.8097

TABLE 7 CROSS TABLE OVERALL RESULTS SITTING FORWARD FLEXION TEST

Cross table Rater A / Rater B – overall results for sitting forward flexion test					
		Rater B		overall	
		conspicuous	inconspicuous		
Rater A	conspicuous	Noticed number	3	4	7
		in %	6.52%	8.7%	15.28%
	inconspicuous	Noticed number	5	34	39
		in %	10.87%	73.91%	84.78%
Total	Noticed number	8	38	46	
	in %	17.39%	82.61%	100%	
		Agreement		80.43%	
		No match		19.57%	

TABLE 8 PRESENTATION OF RESULTS OF THE SITTING FORWARD FLEXION TEST

Kappa- coefficient	0.28
P ₀	80.43%
P _{Index}	0.16
p- value	0.7800

DISCUSSION

Interpretation of results

The calculated Kappa coefficient for the standing forward flexion test is $\kappa = 0.36$. The agreement of results in percentage between Tester A and Tester B during the standing forward flexion test is 76.09%. The corresponding 95% confidence interval ranges from 0.057 to 0.668. This is not very wide, indicating that the estimation is more precise. After analysis and calculation, according to the classification by Landis and Koch (1977), a moderate agreement was demonstrated for the standing forward flexion test, showing limited consistency. The alternative hypothesis is rejected. With a p- value of 0.8097, the result is also not statistically significant.

The calculated Kappa coefficient for the sitting forward flexion test is $\kappa = 0.28$. An agreement of 80.43% was calculated for the sitting forward flexion test. Therefore, the sitting forward flexion test is also not reliable and shows limited consistency. The corresponding 95% confidence interval ranges from -0.069 to 0.637. The interval is relatively wide and subject to

some uncertainty. The p- value is 0.7800, also not statistically significant. The alternative hypothesis is rejected, and the null hypothesis is accepted.

The calculation of P_{index} recommended by Patijn (2019) is 0.25 for the standing forward flexion test and 0.16 for the sitting forward flexion test. If the P_{index} value falls outside the range of 0.2-0.9, it may lead to biases in the Kappa value (Patijn, 2019). Patijn (2019) describes the 0.50- P_{index} method as a way to prevent this bias. In the present study, due to the low P_{index} especially in the case of the sitting forward flexion test, the calculated Kappa value of 0.28 must be questioned. There is evidently a Kappa paradox as described by Derksen et al. (2024) in their study.

Regarding the Confidence- Rating- Scale, unlike the study by Cooperstein et al. (2010), there is no apparent added value in the present work. With $\kappa = 0.35$ for the standing forward flexion test and $\kappa = 0.16$ for the sitting forward flexion test, the result is completely independent of the confidence of the testers. Overall, CRS seems negligible in the present work, as it provides additional information but does not computationally influence the primary or secondary outcome of the study.

Interpretation of Results in the context of Osteopathy

The agreement of results between Rater A and Rater B is 76.09% for the standing flexion test and 80.43% for the sitting flexion test. As described above, this paradox distorts the results. McHugh (2012) even suggests that in well- trained and experienced raters, the percentage agreement holds great significance. This raises two additional questions! Who is considered well- trained? At what point does one become experienced?

Limiting factor concerning the raters

The study supervisor recommends a more homogeneous group of raters for future studies. It is suggested that not only should the educational qualifications of the raters be similar, but also their years of professional experience. This is particularly important in the field of palpation, as the raters' experience could play a crucial role. This could ensure higher comparability and reliability of the results.

Limiting factors concerning the sample size

Overall, only eleven (23.91%) out of 46 participants could be classified as symptomatic. This does not correspond to the daily practice of osteopaths. Healthy individuals, i.e., asymptomatic individuals without underlying conditions, would rarely seek treatment. However, the number

of symptomatic participants is too small to draw conclusions. Perhaps a precise calculation of sample size would have altered the sample size, potentially including more men and symptomatic participants in the study. It would be desirable to aim for better homogeneity in future studies. A larger number of participants may have also reduced the confidence intervals and improved the robustness of the results. According to Patijn (2019), a larger sample size can also alter the P_{index} , which would be desirable in the present study.

Critical reflection on the study process

A valid point of criticism is the adherence to the guidelines set by Patijn (2019). The present study, unlike the recommendation of Patijn (2019), was conducted differently. However, due to the proximity to clinical practice, both tests, the standing and the sitting forward flexion test, were performed consecutively by one tester on one participant. It is possible that two different testing time points, one for the standing forward flexion test and another for the sitting flexion test, should have been chosen to achieve a higher Kappa value.

Summary

The investigation of the inter- rater- reliability of the standing and sitting forward flexion test in the present study appears to be moderate. According to Fjellner et al. (1999) and McHugh (2012), the Kappa values are below the acceptable threshold and far from substantial. The null hypothesis is accepted, while the alternative hypothesis is rejected. The standing and sitting forward flexion test are not sufficiently reliable. From a scientific perspective, the results do not provide adequate justification for evidence- based application.

Clinically, two testers agreed 35 times (76.09%) during the standing forward flexion test and 37 times (80.43%) during the sitting forward flexion test, with a sample size of 46 participants. The low Kappa coefficient indicates a kappa paradox.

Of further clinical interest is the higher Kappa value observed exclusively in symptomatic participants during exploratory data analysis. However, the number of symptomatic participants in the present study is too small to make a scientific statement.

Regarding the Confidence- Rating- Scale, no scientifically relevant conclusion can be drawn in this study.

CONCLUSION

Overall, the results indicate that the inter- rater- reliability of both forward flexion tests needs to be improved in order to obtain reliable and consistent test results. The very frequent and

widespread use of the standing and sitting forward flexion test in the “examination routine” should be questioned from a scientific perspective. Additionally, teaching the forward flexion test in the basic training of osteopathy is therefore also questionable.

However, the clinical value of the standing and sitting forward flexion test may still be present. On one hand, by optimizing the study, the impact of the Kappa paradox on inter-rater-reliability can be reduced. On the other hand, the present results suggest a potential change in inter-tester-reliability in a purely symptomatic group of participants.

Disclosure

The author has no personal financial or institutional interest in any of the drugs , materials, or devices described in this article.

REFERENCES

- Egan, D., Cole, J., & Twomey, L. (1996). The Standing Forward Flexion Test: An Inaccurate Determinant of Sacroiliac Joint Dysfunction. *Physiotherapy*, 82(4), 236–242. [https://doi.org/10.1016/S0031-9406\(05\)66879-4](https://doi.org/10.1016/S0031-9406(05)66879-4)
- Fjellner, A., Bexander, C., Faleij, R., Strender, L. E. (1999). Interexaminer reliability in physical examination of the cervical spine. *Journal of Manipulative and Physio- logical Therapeutics* 22. [https://doi.org/10.1016/s0161-4754\(99\)70002-3](https://doi.org/10.1016/s0161-4754(99)70002-3)
- Fossum, C. (2010). Die osteopathische Diagnosefindung. In: Liem, T., Dobler, T. K.. Leitfaden Osteopathie. München: Urban & Fischer Verlag/ Elsevier GmbH.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Liem, T. & Dobler, T. K., (2010). *Leitfaden Osteopathie: Parietale Techniken* (3. Aufl.). Urban & Fischer Verlag/Elsevier GmbH.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. (online Zugriff am 12.10.2023)

Patijn, J. (2019). Reproducibility protocol for diagnostic procedures in Manual/Musculoskeletal Medicine: Edition 2019. *Manuelle Medizin*, 57(6), 451–479. <https://doi.org/10.1007/s00337-019-00581-5>

Ribeiro, R. P., Guerrero, F. G., Camargo, E. N., Pivotto, L. R., Aimi, M. A., Loss, J. F., & Candotti, C. T. (2021). Construct validity and reliability of tests for sacroiliac dysfunction: Standing flexion test (STFT) and sitting flexion test (SIFT). *Journal of Osteopathic Medicine*, 121(11), 849–856. <https://doi.org/10.1515/jom-2021-0025>

Vincent- Smith, B., Gibbons, P. (1999). Inter- examiner and intra- examiner reliability of the standing flexion test. *Manual Therapy*, 4 (2), 87-93. <https://doi.org/10.1054/math.1999.0173>