

Return to Activity Algorithmus

Erfüllen die drei quantitativen Hop-Tests die Gütekriterien
der Validität und Reliabilität?

Liechti, Ursina
16544926

Gfeller, Ariane
16544827

Departement: Gesundheit
Institut für Physiotherapie

Studienjahr: PT16

Eingereicht am: 26.04.2019

Begleitende Lehrperson: Meichtry André

**Bachelorarbeit
Physiotherapie**

Inhaltsverzeichnis

1	Einleitung	7
1.1	Begründung der Themenwahl	7
1.2	Problemstellung	8
1.3	Eingrenzung des Themas	8
1.4	Fragestellung	9
1.5	Zielsetzung	9
2	Theoretischer Hintergrund	9
2.1	Funktionsphase versus Wundheilungsphase	9
2.2	Funktionelle Assessments	10
2.3	Return to Activity Algorithmus	10
2.4	Definition der relevanten Sprungtests	12
2.4.1	Front Hop (for distance)	12
2.4.2	Side Hop	13
2.4.3	Square Hop	13
2.5	Wundheilungsphasen	13
2.5.1	Entzündungsphase	14
2.5.2	Proliferationsphase	14
2.5.3	Remodulierungsphase	14
2.5.4	Wundheilungsdauer	15
2.5.5	Assessments zur Bestimmung der Wundheilung	15
2.6	Einführung Reliabilität und Validität	15
2.7	Definition Reliabilität	15
2.8	Definition Validität	18
2.8.1	Augenscheinvalidität (face validity)	18

2.8.2	Inhaltsvalidität (content validity)	18
2.8.3	Konstruktvalidität (construct validity)	19
2.8.4	Kriteriumsvalidität (criterion-related validity) und Sensitivität, Spezifität versus positiv- und negativ-prädiktivem Wert	19
2.9	Relevante Kennwerte	21
2.9.1	Standardabweichung (SD = Standard Deviation).....	21
2.9.2	Standardmessfehler (SEM = Standard Error of Measurement) und kleinste erkennbare Veränderung (SDC = Smallest Detectable Change)	22
2.9.3	Konfidenzintervall (CI = Confidence Intervall)	22
2.10	Verwendete Verfahren.....	22
2.10.1	Korrelation nach Bravais-Pearson	22
2.10.2	Rangkorrelationsanalyse nach Spearman (Spearman Rank)	23
2.10.3	Intraklassen-Korrelations-Koeffizient (ICC = Intraclass Correlation Coefficient).....	23
2.10.4	Bland-Altman-Diagramm.....	23
2.10.5	Mann-Whitney-U-Test.....	24
2.10.6	ANOVA (analysis of variance)	24
2.10.7	T-Test für abhängige Stichproben (paired sample t-test).....	24
2.10.8	Wilcoxon- Test (Wilcoxon signed-rank test).....	24
2.10.9	Post-hoc-Tests.....	24
3	Methode.....	25
3.1	Literaturrecherche.....	25
3.2	Eingrenzen der gefundenen Literatur	26
3.3	Zusätzlich verwendete Literatur und Quellen.....	28
3.4	Arbeitsinstrument zur kritischen Würdigung der Studien	29
4	Ergebnisse	29
4.1	Studie 1 von Kockum und Heijne (2014)	29

4.1.1	Zusammenfassung.....	29
4.1.2	Würdigung.....	31
4.2	Studie 2 von Gustavsson et al. (2006).....	32
4.2.1	Zusammenfassung.....	32
4.2.2	Würdigung.....	35
4.3	Studie 3 von Haitz et al. (2014).....	36
4.3.1	Zusammenfassung.....	36
4.3.2	Würdigung.....	38
4.4	Studie 4 von Reid et al. (2007)	38
4.4.1	Zusammenfassung.....	38
4.4.2	Würdigung.....	41
5	Diskussion.....	41
5.1	Reliabilität	42
5.1.1	Kontrolle von Bias in den untersuchten Studien.....	42
5.1.2	Wichtige Masse für die Bestimmung der Reliabilität	44
5.1.3	Front Hop	45
5.1.4	Side Hop	46
5.1.5	Square Hop.....	48
5.2	Validität.....	48
5.2.1	Front Hop	50
5.2.2	Side Hop	52
5.2.3	Square Hop.....	52
6	Schlussfolgerung.....	53
6.1	Beantwortung der Fragestellung.....	53
6.2	Limitationen der Arbeit.....	53
6.3	Weiterführende Überlegungen.....	53
7	Praxistransfer.....	54

Literaturverzeichnis.....	56
Glossar	59
Abkürzungsverzeichnis	61
Abbildungsverzeichnis	62
Tabellenverzeichnis	62
Weitere Quellen	63
Deklaration der Wortanzahl	63
Danksagung.....	63
Eigenständigkeitserklärung.....	63
Anhang	64
Anhang 1: Neun relevante Studien nach der ersten Auswahl	64
Anhang 2: Suchmatrix.....	65
Anhang 3: AICA-Formulare.....	69

Abstrakt

Einleitung: Um die Frage zu klären, wann nach einer Verletzung wieder in den Sport zurückgekehrt werden darf, werden vermehrt funktionelle Assessmenttests verwendet. Ein solches Assessment entwickelten die Physiotherapeuten Matthias Keller und Oliver Schmidlein zusammen mit dem Sportwissenschaftler Eduard Kurz, den Return to Activity Algorithmus.

Ziel: Das Ziel dieser Bachelorarbeit ist, drei der insgesamt acht Tests des Algorithmus auf die Gütekriterien Reliabilität und Validität zu überprüfen. Dies betrifft den Front Hop, Side Hop und Square Hop Test.

Methode: Mittels systematischer Literaturrecherche wurde in den Datenbanken PubMed, CINAHL Complete, MEDLINE und AMED nach relevanter Literatur gesucht. Dabei wurden vorgängig definierte Ein- und Ausschlusskriterien beachtet. Die ausgewählten vier Studien wurden mithilfe des AICA-Formulars ausgewertet und beurteilt.

Ergebnisse: Der Intraklassen-Korrelations-Koeffizient von gesunden, jungen und sportlichen Personen, zeigt für alle Hop Tests eine akzeptable bis sehr gute Test-Retest-Reliabilität. Zur Validität liegen vereinzelte Ergebnisse vor, welche eine Beantwortung der hier relevanten Fragestellung nicht zulassen.

Schlussfolgerung: Den Ergebnissen nach sind die untersuchten Tests zuverlässige Messinstrumente. Die Resultate können mehrheitlich nur auf eine gesunde und sportliche Population übertragen werden.

Die Validität müsste mit einer klaren, auf diese Arbeit abgestimmten, Hypothese eruiert werden.

Keywords: Untere Extremität, Validität, Reliabilität, Sprungtests, Front Hop Test, Side Hop Test, Square Hop Test

Abstract

Introduction: Functional assessment tests are used more frequently to answer the question when to return to sports after injuries. Such an assessment was developed by the physical therapists Matthias Keller and Oliver Schmidlein and sports scientist Eduard Kurz, the “Return to Activity Algorithm”.

Objectives: The aim of this paper is, to verify three out of eight tests on the quality criteria reliability and validity. These three tests are the front hop, the side hop and the square hop.

Method: In the data bases PubMed, CINAHL Complete, MEDLINE and AMED relevant literature was searched. Criteria for inclusion and exclusion, that had previously been determined, were considered. The studies were assessed with the AICA form.

Results: Intraclass-correlation coefficient of young and healthy athletes shows acceptable, even excellent test-retest-reliability. As to validity, only few results were available. They don't answer the relevant question, that was attempted to be proved in this paper.

Conclusion: According to the results, the analysed tests are reliable measuring instruments mainly used on a young, sportive population.
With a clear hypothesis, adjusted to this paper, validity would have to be elicited.

Keywords: Lower extremity, Validity, Reliability, Hop Test, Front Hop Test, Side Hop Test, Square Hop Test

1 Einleitung

Wann darf man nach einer Verletzung der unteren Extremität wieder Sport machen? Welche Art von Sport ist wann wieder erlaubt? Solche und ähnliche Fragen beschäftigen laut Mykelbust und Bahr (2005) Patienten und Patientinnen sowie Therapeuten und Therapeutinnen gleichermaßen. Laut Keller, Kurz, Schmidlein, Welsch und Anders (2016) richtete man sich dabei lange Zeit nach Zeitangaben, die von der Wundheilungsphase abhängig gemacht wurden. Passive Tests bei Verletzungen der unteren Extremität weisen laut Keller, Schmidlein und Kurz (2012) keinen Zusammenhang mit dem funktionellen Outcome auf. Aus diesen Gründen werden seit einigen Jahren vermehrt funktionelle Belastungstests eingesetzt, um das Belastungsniveau der Patienten und Patientinnen einschätzen zu können. Die Physiotherapeuten Matthias Keller und Oliver Schmidlein entwickelten zusammen mit dem Sportwissenschaftler Eduard Kurz im Jahr 2012 einen funktionellen Test, der genau diese Fragen klären sollte, den Return to Activity Algorithmus (RTA-Algorithmus). Ziel des RTA-Algorithmus ist, das aktuelle Leistungsniveau eines Patienten oder einer Patientin zu bestimmen. Er dient auch als Entscheidungshilfe für eine Belastungssteigerung. Der RTA-Algorithmus besteht aus acht funktionellen Tests. Dazu gehören der Y-Balance, der Balance Squat, drei qualitative und drei quantitative Sprungtests. Eine ausführliche Beschreibung des RTA-Algorithmus ist unter 2.3 ersichtlich.

1.1 Begründung der Themenwahl

Bald stand fest, dass sich diese Bachelorarbeit mit einer muskuloskelettalen Thematik befassen würde. Durch ein Praktikum wurde das Interesse am RTA-Algorithmus geweckt. Da Verletzungen der unteren Extremität im Berufsalltag der zwei Autorinnen immer ein relevantes Thema sein würden, verstärkte sich das Interesse, mehr über dieses funktionelle Assessment herauszufinden.

Die meiste Literatur, welche sich mit dem RTA-Algorithmus befasst, untersuchte Patientinnen und Patienten mit Verletzungen des vorderen Kreuzbandes. Gemäss Unfallstatistik UVG (zit. nach der Gesundheitsdirektion des Kantons Zürich, 2009) gibt es in der Schweiz pro Jahr ca. 10'000-12'000 Knieverletzungen, bei denen das vordere Kreuzband beschädigt wird. Grundsätzlich sollte der RTA-Algorithmus laut Keller et al. (2016) aber auch bei anderen Verletzungen der unteren Extremität angewendet werden

können. Aus diesen Fakten geht hervor, dass ein funktionelles Assessment wie der RTA-Algorithmus von grosser Bedeutung ist.

1.2 Problemstellung

Der RTA-Algorithmus scheint ein optimales Assessment zur Bestimmung des aktuellen Leistungsniveaus zu sein und als Entscheidungshilfe für den geeigneten Zeitpunkt einer Belastungssteigerung zu dienen. Nun stellt sich die Frage, ob diese acht Tests auch die Gütekriterien (siehe 1.3) erfüllen und somit aussagekräftige Assessments sind. Seitens Matthias Keller, Oliver Schmidlein und Eduard Kunz wurden diesbezüglich keine Aussagen gefunden. Auch in den Datenbanken der Hochschulbibliothek der ZHAW wurde von den Autorinnen keine Literatur gefunden, die die Gütekriterien des RTA-Algorithmus untersucht. Das von Keller und Schmidlein gegründete Institut «Bewegung für Orthopädie und Sportmedizin» (OSINSTITUT), gab auf Anfrage im Juli 2018 bekannt, dass sich aktuell eine Masterarbeit mit der Validität der acht Tests des RTA-Algorithmus befasst. Unklar ist, wann diese Masterarbeit fertiggestellt wird. Der aktuelle Stand der Forschung über die Gütekriterien des RTA-Algorithmus zeigt somit auf, dass in diesem Bereich noch weitere Forschung sinnvoll ist.

1.3 Eingrenzung des Themas

Da die Frage nach den Gütekriterien aller acht Tests des RTA-Algorithmus den Rahmen einer Bachelorarbeit sprengen würde, musste die Fragestellung eingegrenzt werden. Von den acht Tests des RTA-Algorithmus ist der Y-Balance Test schon am meisten untersucht worden. Zudem existieren bereits zwei Bachelorarbeiten an der ZHAW, die diesen Test untersuchen. Also beschlossen die Autorinnen, diesen Test nicht in die Bachelorarbeit aufzunehmen. Übrig blieben sieben Tests, wovon sechs auf Hop-Tests basieren, das heisst, der grösste Teil des RTA-Algorithmus besteht aus Sprungtests. Die Vermutung liegt nahe, dass eine Arbeit über die Sprungtests mehr Aufschluss über den RTA-Algorithmus bringt, als über den Single Leg Squat alleine. Aufgrund dieser Überlegungen wurde das Thema auf die sechs Sprungtests eingegrenzt. Bei der Literatursuche nach passenden Studien bezüglich Validität und Reliabilität der Tests wurde festgestellt, dass über die drei qualitativen Sprungtests kaum Studien vorhanden sind. Sie wurden deshalb auch ausgeschlossen. Es bleiben die drei quantitativen «Hop-Tests».

Oesch et al. (2011) untersuchten die jeweiligen Assessments auf die klinische Relevanz und ihre Praktikabilität, sowie auf die klassischen Gütekriterien Reliabilität, Validität und Responsivität. Da die Überprüfung der drei quantitativen Sprungtests des RTA-

Algorithmus anhand aller fünf erwähnten Kriterien für diese Bachelorarbeit zu umfangreich wäre, beschränkte man sich auf die beiden Gütekriterien Validität und Reliabilität.

1.4 Fragestellung

Erfüllen die drei quantitativen Hop-Tests des Return to Activity Algorithmus die Gütekriterien der Validität und Reliabilität?

1.5 Zielsetzung

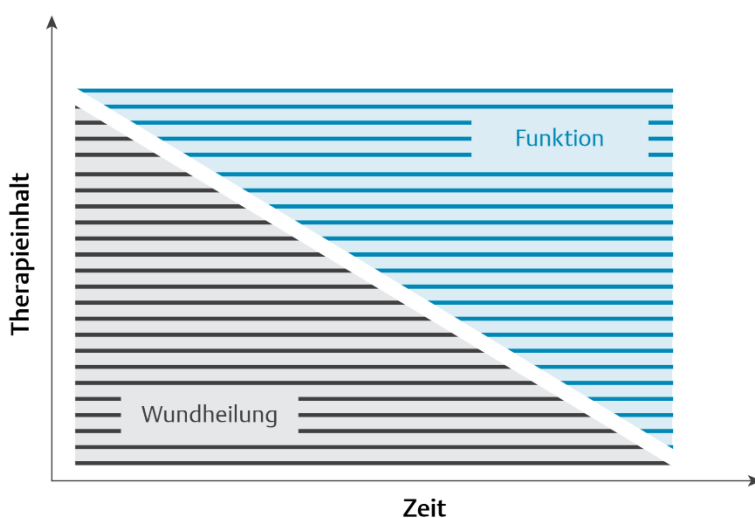
Das Ziel dieser Bachelorarbeit ist, anhand eines Literaturreviews die drei quantitativen Hop-Tests des RTA-Algorithmus, nämlich den Front Hop, den Side Hop und den Square Hop, auf deren Reliabilität und Validität zu überprüfen.

2 Theoretischer Hintergrund

2.1 Funktionsphase versus Wundheilungsphase

«Grundsätzlich lässt sich die Rehabilitation in eine Wundheilungsphase und eine Funktionsphase einteilen. (...) Beide Phasen gehen fließend und progressiv ineinander über» (Keller et al., 2016, S. 138). Zu Beginn steht die Wundheilung im Vordergrund und die Funktion beschränkt sich auf ein Minimum. Mit dem Voranschreiten der Wundheilung werden die Strukturen belastbarer und übernehmen immer mehr ihrer ursprünglichen Funktion (Keller et al., 2016). Dies verdeutlicht die Abbildung 1, die das Kontinuum der Wundheilung und Funktion darstellt.

Abbildung 1 Kontinuum der Wundheilung und Funktion (Keller et al., 2016)



Um bestimmen zu können, wie viel Belastung in der Funktionsphase wieder möglich ist, gibt es verschiedene funktionelle Assessments. Eines davon ist der RTA-Algorithmus, um den es in dieser Bachelor-Arbeit geht.

2.2 Funktionelle Assessments

Wirz et al. (2014) unterscheiden Assessments, welche Körperfunktion und Körperstrukturen untersuchen, von solchen, die Aktivitäten und Partizipation erfassen. Assessments, die Körperfunktion und Körperstrukturen umfassen, werden nachfolgend als strukturelle Assessments, solche die Aktivität und Partizipation umfassen als funktionelle Assessments bezeichnet.

Strukturelle Assessments sind in der Physiotherapie von grosser Bedeutung. So ist der Lachman-Test, in Kombination mit anderen VKB-Tests, ein preisgünstiges und schnelles Assessment zur Diagnosestellung einer VKB-Ruptur (DocCheck o.D.). Auch Muskelfunktionstests für den M. Quadriceps oder die Hamstrings vermitteln einen groben Eindruck über eine eventuelle Muskelschwäche. Strukturelle Assessments zeigen jedoch grundlegende Unterschiede zu funktionellen Assessments. So ist ein Merkmal der funktionellen Assessments, dass sie von mehreren Faktoren beeinflusst werden, wohingegen die strukturellen Assessments von weniger Faktoren abhängen (Wirz et al, 2014). Des Weiteren weisen passive strukturelle Assessments bei Verletzungen der unteren Extremität laut Keller et al. (2012) keinen Zusammenhang mit dem funktionellen Outcome auf. Auch Kong et al. (2011) weisen darauf hin, dass strukturelle Assessments wie der Lachman-Test, vordere Schubladentest, Pivot-Shift-Test, isokinetische Krafttest und weitere Tests, die funktionelle Stabilität des Knies bei Sportaktivitäten nicht verlässlich voraussagen können. Auch Keller und Kunz (2016) unterstrichen dies: «Isolierte Funktionsprüfungen, wie Kraft- und Beweglichkeitsmessungen, werden eingesetzt, um Faktoren zu identifizieren, die mit einem erhöhten Verletzungsrisiko assoziiert sind und um Erfolge im Rehabilitationsverlauf zu dokumentieren. Um komplexere Fähigkeiten zu prüfen arbeitet man mit «On Field Tests».» (S. 17)

Es gibt eine Vielzahl funktioneller Assessments, die in der Praxis angewendet werden, deren Erwähnung hier aber keinen Platz findet. Aus diesem Grund wird im Folgenden nur auf den RTA-Algorithmus eingegangen.

2.3 Return to Activity Algorithmus

Die Idee des RTA-Algorithmus besteht darin, die Rückkehr in den Sport nicht an der Anzahl posttraumatischer, respektive postoperativer Wochen entsprechend der Wundheilungsphasen zu richten, sondern der Patientin und dem Patienten dann grünes Licht zu geben, wenn er funktionell in der Lage ist, die betroffene Struktur zu stabilisieren und zu kontrollieren (Borgstedt, 2016). Wie Matthias Keller dem Online-Magazin

«Netzathleten» gegenüber erwähnt, besteht der Vorteil des Algorithmus darin, dass bereits während der Rehabilitationsphase objektive Parameter vorhanden sind, welche über eine Belastungssteigerung entscheiden (Borgstedt, 2016).

Der Algorithmus unterscheidet vier unterschiedliche Belastungsstufen, die laut Keller et al. (2012) schon in den Neunzigerjahren von anderen Autoren und Autorinnen definiert wurden. Keller et al. (2012) haben in der Literatur nach je einem qualitativen und einem quantitativen Test gesucht, welcher zu der jeweiligen Belastungsstufe passt. Daraus ergaben sich acht Tests, je zwei in jeder der vier Belastungsstufen. Auf jeder Belastungsstufe muss zuerst der qualitative Test bestanden werden, bevor der quantitative durchgeführt werden darf. Die Beurteilung der vier qualitativen Tests bezüglich «bestanden» oder «nicht bestanden» geschieht anhand vorgegebener Kriterien. Da sie aber nicht Teil dieser Arbeit sind, wird an dieser Stelle nicht weiter darauf eingegangen. Die quantitativen Tests werden anhand eines Limb Symmetrie Index (LSI) bestimmt, welcher den Seitenunterschied der betroffenen zur nicht betroffenen Seite in Prozent ausdrückt. Besteht der Patient/die Patientin den qualitativen Test und beim quantitativen wird ein LSI-Wert von über 90% im Seitenvergleich erreicht, ist das Level bestanden. Der Patient/die Patientin kann nun mit dem nächsten Level starten, sofern er/sie das Ziel eines höheren Belastungslevels anstrebt (Keller et al., 2016). Level 1 besteht aus zwei Balancetests, dem Balance Squat und den Y-Balance Test, welche ohne Beschleunigung stattfinden. Der Balance Squat, sprich eine einbeinige Kniebeuge, testet die Stabilisation des zu testenden Gelenkes in der Sagittalebene. Der Y-Balance Test ist ein funktioneller und dynamischer Balancetest, der die posturale Kontrolle im Einbeinstand quantifiziert (Keller et al., 2016). Level 2 besteht aus einem Balance Front Hop Test und dem Front Hop Test. Während beim erstgenannten die Qualität massgebend ist, kommt es beim Front Hop nur auf die Weite der Sprünge im Seitenvergleich an (Keller et al., 2016). Level 3 untersucht die frontale Ebene mittels Balance Side Hop und dem Side Hop Test. Level 4 inkludiert durch den 90° Balance Hop und den Square Hop Test die frontale und sagittale Ebene. Bei den quantitativen Tests aus Level 3 und 4 ist nicht die Distanz, sondern die Anzahl korrekter Sprünge massgebend (Keller et al., 2016). Dies wird im Kapitel 2.4.2 näher erläutert.

Eine Übersicht der acht Tests des RTA-Algorithmus ist in Tabelle 1 ersichtlich.

Tabelle 1 Aufbau des RTA-Algorithmus (angepasst nach Keller, M. & Kurz E. (2016))

	Qualitativer Test	Quantitativer Test
Level 1	Balance Squat	Y-Balance
Level 2	Balance Front Hop	Front Hop Test
Level 3	Balance Side Hop	Side Hop Test
Level 4	90° Balance Hop	Square Hop Test

Unter 2.4 ist eine genaue Definition der drei in dieser Arbeit verwendeten Sprungtests ersichtlich. Auf die genaue Durchführung der restlichen fünf Tests des RTA-Algorithmus wird in dieser Arbeit verzichtet.

Zur Standardisierung wird bei jedem Wiederbefund jeweils mit dem Level 1 gestartet. Dies, so Keller gegenüber dem Online-Magazin «Netzathleten», sei eine Sicherheitsmassnahme, damit Veränderungen der motorischen Kontrolle sofort erkannt werden können (Borgstedt, 2016).

Tabelle 2 richtet sich nach einer bereits existierenden Graphik von Daniel et al. (1994) welche von Keller et al. (2016) verwendet und modifiziert wurde. Sie zeigt, ab welchem Level welche Aktivität wieder möglich ist.

Tabelle 2 Einteilung der Aktivitätslevel (angepasst nach Keller, M., Schmidlein, O., Kurz, E., Welsch, G. & Anders C. (2016))

Level	Aktivitäten
I	<ul style="list-style-type: none"> • Sportarten ohne Sprünge (Golf) • Alltagsbewegungen
II	<ul style="list-style-type: none"> • Dynamische Sportarten ohne Dreh-/Stoppbewegungen (Joggen und Rennen) • Leichte körperliche Arbeit
III	<ul style="list-style-type: none"> • Seitliche Bewegungen und geringere Dreh-/ Stoppbewegungen als Level IV • Klettern • Harte körperliche Arbeit • Unebener Boden
IV	<ul style="list-style-type: none"> • Sprünge, schnelle Dreh-/ Stoppbewegungen (Fussball, Basketball) • Sport als Arbeitsbelastung (z.B. Berufsfeuerwehr)

2.4 Definition der relevanten Sprungtests

2.4.1 Front Hop (for distance)

Der Front Hop wurde schon 1990 von Barber, Noyes und Mangine, McCloskey und Hartman beschrieben. Dieser beschreibt einen Einbeinsprung, der eine möglichst grosse

Distanz umfassen soll. Dabei ist ein Armeinsatz erlaubt. Die Landung erfolgt mit dem gleichen Bein und muss sicher gehalten werden können. Gemessen wird die Distanz von der Startlinie bis zur Ferse in Zentimetern. Dies wird jeweils mit jedem Bein dreimal wiederholt. Mit dem besten der drei Versuche pro Bein wird anschliessend der LSI berechnet (Keller et al., 2016).

2.4.2 Side Hop

Für den Side Hop übernahmen Keller et al. (2016) das Protokoll von Gustavsson et al. (2006). Die zu untersuchende Person steht mit dem zu testenden Fuss lateral der Linie. Eine zweite Linie ist mit einem Abstand von 40cm auf den Boden geklebt. Nun geht es darum, innerhalb von 30 Sekunden möglichst oft über die Linien zu hüpfen, ohne diese dabei zu berühren. Gezählt werden die Anzahl korrekter Bodenkontakte. Sind mehr als 25% aller Sprünge fehlerhaft, muss der Test nach 3 Minuten wiederholt werden. Die korrekten Bodenkontakte werden anhand des LSI im Seitenvergleich ausgewertet.

2.4.3 Square Hop

Der Square Hop wurde von Keller et al. (2016) so definiert, dass ein Quadrat von jeweils 40cm Länge auf den Boden geklebt wird. Das Quadrat liegt vor der Patientin/dem Patienten, diese/dieser steht auf einem Bein und versucht während 30 Sekunden so viele Sprünge wie möglich ins Quadrat und wieder hinaus zu machen. Dabei wechselt jeweils die Kante des Quadrats, welche zu überhüpfen ist. Nach dem Sprung zur Mitte geht es als nächstes zur Seite wo die Grosszehe hinzeigt. Springt der Proband oder die Probandin mit dem rechten Bein, bedeutet dies wörtlich: Mitte – Links – Mitte – Oben – Mitte – Rechts – Mitte – Unten. Dann beginnt der Zyklus von vorne. Werden die Linien beim Sprung berührt, gilt dies als Fehler. Sind mehr als 25% der Sprünge fehlerhaft, muss der Test nach 3 Minuten wiederholt werden. Die korrekten Bodenkontakte werden anhand des LSI im Seitenvergleich ausgewertet.

2.5 Wundheilungsphasen

Wie unter 2.1 beschrieben, sollte die Wiederaufnahme der Funktion in direktem Zusammenhang mit der Wundheilung stehen. Die verschiedenen Strukturen des Körpers weisen mehrheitlich denselben Ablauf der Wundheilung auf: Entzündungsphase, Proliferationsphase, Remodulierungsphase. Nicht identisch hingegen ist die Dauer der jeweiligen Phasen, wie dies in Kapitel 2.5.4 besprochen wird.

2.5.1 Entzündungsphase

Die Entzündungsphase wird in zwei Phasen unterteilt, die vaskuläre und die zelluläre. Das Ziel der vaskulären Phase ist es, den Blutverlust möglichst gering zu halten. Dies geschieht mittels Vasokonstriktion, einer Gefäßverengung defekter Blutgefäße. Nach van den Berg (2011) erhöht sich zudem die Permeabilität der Kapillare, wodurch Flüssigkeit ins Gewebe austreten kann. Die typischen Entzündungszeichen Schwellung, Rötung und Erwärmung werden hervorgerufen. Zeitgleich zeigt sich bei den intakten Gefässen eine Vasodilatation, also eine Gefässerweiterung, wodurch Wundflüssigkeit des sekundär entstandenen Wundgebietes abtransportiert werden kann (van den Berg, 2011). In der folgenden zellulären Phase wandern Makrophagen in das Wundgebiet ein. Ihre Aufgabe ist es, nekrotisches Material aus dem Wundgebiet zu entfernen, damit neues, intaktes Gewebe aufgebaut werden kann. So kann bereits sehr früh ein provisorisches Fibrinnetzwerk über das Wundgebiet aufgebaut werden (Physiosupport, o.D.).

2.5.2 Proliferationsphase

Fibroblasten beginnen in dieser Phase mit der Produktion von Matrixbestandteilen (Van den Berg, 2011). Vom Wundrand ausgehend breiten sie sich immer mehr über das provisorische Fibrinnetzwerk ins gesamte Wundgebiet aus. Nach ca. 7-14 Tagen befinden sich immer weniger Monozyten, Leukozyten, Lymphozyten und Makrophagen im Wundgebiet, bis schlussendlich nur noch Fibroblasten vorhanden sind. Verläuft die Wundheilung physiologisch, sind in dieser Phase keine Entzündungszeichen mehr vorhanden. Halten die Entzündungszeichen jedoch an, kann das ein Indiz dafür sein, dass sich die Person zu wenig schont oder seine Verletzung bagatellisiert.

Während der Proliferationsphase hat die Synthese von Kollagen Typ 3 überhand. Bei angepasster Belastung kann sich das heranwachsende Gewebe sauber organisieren und ausrichten (Physiosupport, o.D.).

2.5.3 Remodulierungsphase

In dieser Phase wird das Kollagen Typ 3 in Kollagen Typ 1 umgebaut, wodurch das Gewebe stabiler wird. Zudem beginnen das Einwachsen der Nervenfasern sowie die Reorganisation des Gefässbettes. Das Gewebe darf nun mehr belastet werden. Am Ende der Remodulierungsphase ist das Gewebe wieder soweit stabil, dass es nahezu seine alte Stabilität wiedererlangt hat (Physiosupport, o.D.).

2.5.4 Wundheilungsdauer

In Tabelle 3 wird die Dauer der Wundheilungsphasen für verschiedene Gewebe aufgezeigt. Die Zeitangaben sind Richtwerte und können variieren (Physiosupport, o.D.).

Tabelle 3 Dauer der Wundheilung (angepasst nach Diemer F. & Sutor V. (2018))

	Muskeln	Sehnen	Ligamente	Knorpel	Knochen
Entzündung	2 Tage	Bis 1 Woche	2 - 5 Tage	Keine	2 - 5 Tage
Proliferation	14 - 21 Tage (Zerrung) 6 - 8 Wochen (Faserriss)	Bis 6 Wochen	Bis 6 Wochen	3 Monate	3 - 10 Wochen
Remodulierung	Ab 3 Wochen	Bis 1 Jahr	Bis 1 Jahr	6 - 9 Monate (je nach Verletzung bleibender Schaden)	Mind. 1 Jahr

2.5.5 Assessments zur Bestimmung der Wundheilung

Neben dem zeitlichen Verlauf der Wundheilung existieren auch Assessments, die eine Aussage zur Wundheilung machen. So verwenden Keller et al. (2016) zur Sicherstellung der Wundheilung die Schmerzerfragung mittels verschiedener Assessments, die Umfangmessung an vier Punkten (Kniegelenkspalt, 15 Zentimeter distal, 10 und 20 Zentimeter proximal des Kniegelenkspalts), die Messung der Hautoberflächentemperatur und das Bewegungsausmass nach der Neutral-Null-Methode. Diese Assessments werden jeweils mit der Gegenseite verglichen. Mit den funktionellen Tests wird erst begonnen, wenn die Werte seitengleich sind.

2.6 Einführung Reliabilität und Validität

Die Begriffe Reliabilität und Validität nehmen in dieser Arbeit einen grossen Stellenwert ein. Doch was bedeuten und beinhalten diese Begriffe eigentlich? Oft schmücken sich Studien mit der Aussage, ein Assessment sei valide oder reliabel. Beim näheren Betrachten der Studien wird dann schnell klar, dass die Begriffe Validität und Reliabilität nicht einfach zu verstehen sind und ganz verschiedene Ausmasse annehmen können. Aus diesem Grund folgt nun eine etwas längere Definition der zwei Begriffe.

2.7 Definition Reliabilität

Laut Oesch et al. (2011) sind Reliabilität, Reproduzierbarkeit und Zuverlässigkeit weitgehend identische Begriffe. Kommen wiederholte Messungen unter gleichen

Bedingungen zum gleichen Ergebnis, bezeichnet man das Messverfahren als reliabel. Nach Mokkink, Terwee und Patrick (2010, zitiert nach de Vet, Terwee, Mokkink & Knoll, 2011, S. 96) ist Reliabilität definiert als das Ausmass, in dem ein Messinstrument frei von Messfehlern ist. Messfehler können verschiedene Ursachen haben: das Messinstrument selbst; die Person, die das Messinstrument anwendet; Personen, bei denen das Messinstrument angewendet wird oder Umstände, unter denen die Messungen erfolgt sind. De Vet et al. (2011) definieren Reliabilität auf eine etwas komplexere Weise. So ist Reliabilität das Ausmass, in dem Werte für Patientinnen und Patienten, die sich nicht verändert haben, bei repetierten Messungen unter verschiedenen Bedingungen, gleich bleiben. Anhand dieser Definition werden vier verschiedene Arten der Reliabilität unterschieden: die interne Konsistenz, welche verschiedene Items aus dem gleichen Multi-Item Verfahren anwendet; die Test-Retest-Reliabilität, welche Messungen zu unterschiedlichen Zeitpunkten miteinander vergleicht; die Inter-Rater-Reliabilität, bei der verschiedene Personen die gleiche Begebenheit beobachten und die Intra-Rater-Reliabilität, bei der eine Person zu verschiedenen Zeiten eine Messung durchführt (de Vet et al., 2011).

In dieser Bachelorarbeit werden drei quantitative Tests untersucht, bei denen jeweils die Zentimeter oder die Anzahl Wiederholungen gemessen werden. Es kann davon ausgegangen werden, dass verschiedene Testende auf die gleiche Anzahl Zentimeter und Wiederholungen kommen. Aus diesem Grund wird die Intrarater- und Interrater-Reliabilität kaum Gegenstand der Untersuchung sein. Vielmehr wird die Test-Retest-Reliabilität den grössten Stellenwert einnehmen. Bei Test-Retest Studien ist die Quelle der Fehlervarianz nicht der Rater, sondern der Patient selber (Meichtry, 2018). So kann es vorkommen, dass die Tagesverfassung des Patienten/der Patientin zu einem anderen Ergebnis führt.

Nach de Vet et al. (2011) sieht die mathematische Beschreibung der Reliabilität folgendermassen aus:

Der beobachtete Wert einer Messung besteht immer aus dem wahren- und dem Fehlerwert der Messung. Daraus ergibt sich folgende Formel: $Y_i = \eta + \varepsilon_i$

Dabei steht Y für den beobachteten Wert einer Messung, η für den wahren Wert, ε für den Fehlerwert und i beschreibt die wiederholten Messungen.

Nun werden jedoch nicht die absoluten Werte angeschaut, sondern die Varianzen der absoluten Werte. Dies lässt sich wie folgt ausdrücken: $\sigma^2(Y_i) = \sigma^2(\eta) + \sigma^2(\varepsilon_i)$ Dabei bedeutet σ^2 die Varianz. Geht man von der Annahme aus, dass die Varianz aus

wiederholten Messfehlern konstant ist, liegt die Vermutung nahe, dass die Varianz von verschiedenen beobachteten Werten ebenfalls konstant ist. Somit kann das «i» in der Formel weggelassen werden. Der Einfachheit halber wird $\sigma^2(\varepsilon_i)$ ab nun als $\sigma^2(error)$ geschrieben und anstelle von $\sigma^2(\eta)$, also von der Varianz der wahren Werte, wird der Term $\sigma^2(p)$, also die Personenvarianz verwendet, da meistens Personen untersucht werden.

Somit lässt sich Reliabilität in folgender Formel ausdrücken:

$$\text{Reliabilität} = \frac{\sigma^2(p)}{\sigma^2(Y)} = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(error)}$$

Reliabilität ist somit der Anteil der Personenvarianz an der totalen Varianz. Wenn die Personenvarianz $\sigma^2(p)$ über die Varianz des beobachteten Wertes gleich 1 ist, ist die Messung vollkommen reliabel, denn sie beinhaltet keinen Messfehler $\sigma^2(error)$. Die Reliabilität wird also hoch, wenn der Messfehler klein ist. Die Reliabilität ist allerdings nicht nur vom Messfehler abhängig. Wie die Formel zeigt, steigt die Reliabilität mit der Heterogenität der Personen. Dies lässt sich an folgendem Beispiel erklären: Möchte man zwischen zehn etwa gleich grossen Personen unterscheiden, braucht man ein viel genaueres Messband, als wenn man zwischen zehn ganz unterschiedlich grossen Personen differenzieren will. Eine Messung ist dann reliabel, wenn der Messfehler relativ zur Heterogenität der Personen klein ist (Meichtry, 2018). Reliabilität drückt aus, wie exakt sich Patienten und Patientinnen trotz Messfehlern voneinander unterscheiden lassen (de Vet et al., 2011).

Somit kommt es bei der Reliabilität nicht nur auf das Messinstrument an, sondern auch auf die Population, bei welcher das Messinstrument angewendet wird (Meichtry, 2018).

Reliabilität zeigt einen Zusammenhang, eine Korrelation, zwischen verschiedenen Resultaten auf und wird daher laut Oesch et al. (2011) oft als Korrelationskoeffizienten angegeben. Ein Korrelationswert beschreibt die Stärke des Zusammenhangs zweier Testresultate und zeigt auf, wie nahe Messpaare (Test-Retest) bei grafischen Darstellungen auf einer Linie liegen. Ein Korrelationskoeffizient kann Werte zwischen -1.00 und 1.00 annehmen. 0 bedeutet, dass kein Zusammenhang zwischen den Variablen existiert. 1.00 beschreibt eine perfekte positive, -1.00 eine perfekte negative Korrelation, bei der alle Messpaare auf einer Geraden liegen. Abhängig von den verwendeten Variablen einer Studie wird der parametrische Pearson Korrelationskoeffizient (r_p), der nicht-parametrische Spearmans Rang-Korrelationskoeffizient (r_s) oder der parametrische

Intraklassen-Korrelations-Koeffizient (ICC) berechnet. Es gibt, so Meichtry (2018), verschiedene Arten des ICC. Die oben beschriebene Formel für Reliabilität ist gleichzeitig auch der erste ICC. Die Werte eines ICC können nur zwischen 0.00 und 1.00 liegen, da er anhand der Varianz berechnet wird und diese nur positiv sein kann (Meichtry, 2018). Genaueres zu den drei Korrelations-Koeffizienten ist unter 2.10 ersichtlich.

2.8 Definition Validität

Die Validität, so Oesch et al. (2011), beschreibt, wie gut ein Instrument misst, was es vorgibt zu messen. Es geht um die Frage, ob das Instrument verzerrt ist (Meichtry, 2018). Die Formel $Y_i = \eta + \varepsilon_i$ wurde unter 2.6.1 bereits erklärt. Bei der Validität geht es darum, ob η auch wirklich dem wahren Wert entspricht, oder ob das Messinstrument verzerrt ist. Die Validität untersucht das Messinstrument auf systematische Fehler, sogenannte potentielle Bias (b) (Meichtry, 2018). Die Formel für die Validität ist somit:

$$\text{Validität} = \frac{\sigma^2(p)}{\sigma^2(p) + b^2 + \sigma^2(\text{error})}$$

Wenn die Bias gegen Null gehen und somit vernachlässigbar sind, ist die Validität gleich mit der Reliabilität (Meichtry, 2018).

Es gibt laut Oesch et al. (2011) vier unterschiedliche Stufen der Validität, nämlich die Augenschein-, Inhalts-, Konstrukt- und Kriteriumsvalidität. Mit jeder Stufe nimmt die Validität, die Gültigkeit eines Assessments, zu.

2.8.1 Augenscheinvalidität (face validity)

Die niedrigste Stufe der Validität ist die Augenscheinvalidität, so Oesch et al. (2011). Bei der Augenscheinvalidität wird ein Test von einer Fachperson als valides Messinstrument akzeptiert, obwohl keine formelle Untersuchung dazu vorhanden ist. Die Augenscheinvalidität ist ein subjektives Assessment. Aus diesem Grund gibt es auch keine Standards, wie die Daten erfasst werden sollten und sie können auch nicht quantifiziert werden (de Vet et al., 2011).

2.8.2 Inhaltsvalidität (content validity)

Die Inhaltsvalidität beschäftigt sich laut Meichtry (2018) mit der Frage, ob das Messinstrument das Konstrukt gut abbildet. Wenn Experten der Meinung sind, dass der Inhalt der Messung das Problem gut erfasst, ist die Inhaltsvalidität gegeben. Dazu wird eine Gruppe von Fachleuten befragt, die dann eine Beurteilung zum Grad der Inhaltsvalidität des Tests angeben. Diese Stufe der Validität wird oft zur Erarbeitung eines

Fragebogens verwendet, indem die Fachleute angeben, welche Priorisierung die Items im Fragebogen haben sollen (Oesch et al., 2011).

Eine andere Bezeichnung für die Inhaltsvalidität ist laut Meichtry (2018) der Begriff der internen Konsistenz, welcher in Kapitel 2.7 bereits besprochen wurde.

2.8.3 Konstruktvalidität (construct validity)

Bei der Konstruktvalidität wird ein Instrument mit einem anderen Instrument, welches dasselbe Konstrukt misst, verglichen. Dabei weist das Vergleichsinstrument keinen Goldstandard auf (de Vet et al., 2011).

Die Konstruktvalidität wird weiter in Unterkategorien eingeteilt. Hier liegen kontroverse Definitionen vor: Oesch et al. (2011) teilen die Konstruktvalidität weiter in konvergente und divergente Validität ein. Konvergente (Konstrukt-) Validität ist gegeben, wenn das Messverfahren mit anderen Messverfahren, welche dasselbe Konstrukt messen, stark korreliert. Divergente (Konstrukt-) Validität ist gegeben, wenn das Messverfahren mit Messverfahren, welche andere Konstrukte messen, schwach, beziehungsweise nicht korreliert. Um eine hohe Konstruktvalidität zu erhalten, müssen beide Arten gegeben sein. De Vet et al. (2011) unterteilen die Konstruktvalidität dagegen in strukturelle Validität (structural validity), Hypothesen testende Validität (hypothesis testing validity) und kulturelle Validität (cross-cultural validity).

Die Konstruktvalidität wird mittels Korrelations-Koeffizienten (siehe Punkt 2.10.1, 2.10.2 und 2.10.3) oder Zwischengruppentests (siehe 2.10.5 – 2.10.9) quantifiziert (Meichtry, 2018).

2.8.4 Kriteriumsvalidität (criterion-related validity) und Sensitivität, Spezifität versus positiv- und negativ-prädiktivem Wert

Mokkink et al. (2010, zitiert nach de Vet et al., 2011, S. 159) definieren die Kriteriumsvalidität anhand der Übereinstimmung eines Messinstruments mit dem Goldstandard. So kann die Kriteriumsvalidität nur dann erhoben werden, wenn ein Goldstandard (bzw. ein Kriterium) verfügbar ist. Die Kriteriumsvalidität wird nach Oesch et al. (2011) weiter in prädiktive (predictive validity) und konkurrente Validität (concurrent validity) unterteilt. Die prädiktive Validität sagt aus, ob die vom Test vorhergesagten Prognosen auch wirklich eintreten, beispielsweise Rezidiv-Gefahr für VKB-Rupturen. Die konkurrente Validität beschreibt die Übereinstimmung der Messresultate mit den Werten eines zuvor festgelegten Kriteriums.

Zur Quantifizierung der Kriteriumsvalidität sind die Ausprägungen von Test und Kriterium relevant. Handelt es sich beim Test und beim Kriterium um eine dichotome Ausprägung, wird anhand der Sensitivität und Spezifität quantifiziert. Dichotom bedeutet, dass es nur zwei Ausprägungen gibt, wie positiv/negativ oder ja/nein. Handelt es sich um einen kontinuierlichen Test und ein dichotomes Kriterium, wird aus dem kontinuierlichen Test ein dichotomer Test generiert, der auch mittels Sensitivität und Spezifität quantifiziert wird (Meichtry, 2018). Handelt es sich jedoch beim Test und beim Kriterium um kontinuierliche Ausprägungen, wird mittels Korrelationen quantifiziert (siehe 2.10.1. und 2.10.2.). Sensitivität, Spezifität, positiv- und negativ-prädiktive Werte spielen bei der Quantifizierung von dichotomen Ausprägungen auf Stufe der Kriteriumsvalidität eine wesentliche Rolle. Die vier Begriffe lassen sich am besten anhand einer Vierfeldertafel erklären. Sowohl Tabelle 4, als auch die folgende Erklärung orientieren sich an der Vorlesung von Meichtry (2018) über Validität. Die Tabelle 4 stellt ein exemplarisches Beispiel von zwei dichotomen Ausprägungen dar.

Tabelle 4 Dichotomer Test und dichotomes Kriterium (angepasst nach Meichtry, A. (2018))

	+ Krank	- Gesund
+ Test positiv	a (richtig positiv)	b (falsch positiv)
- Test negativ	c (falsch negativ)	d (richtig negativ)

Die Sensitivität ist die Wahrscheinlichkeit eines positiven Tests bei Kranken und wird mit folgendem Bruch geschätzt: $a/(a+c)$. Die Spezifität wird mit $d/(d+b)$ berechnet und beschreibt die Wahrscheinlichkeit eines negativen Tests, wenn man gesund ist.

Laut Oesch et al. (2011) wären die Ergebnisse für die Sensitivität und Spezifität im Idealfall 1. Dies würde bedeuten, dass 100% als «richtig-positiv» und «richtig negativ» identifiziert wurden. Diese Anforderungen kann jedoch kaum ein diagnostisches Assessment erfüllen. Oft hat ein Test mit hoher Sensitivität eine niedrige Spezifität und umgekehrt.

Die Tabelle 5 wurde von den Autorinnen anhand der Tabelle 4 angepasst. Es wurden als Beispiel zwei dichotome Ausprägungen gewählt, die für diese Bachelorarbeit relevant sind.

Tabelle 5 Vierfeldertafel als Beispiel unserer Fragestellung

	+ Instabilität UEx	- Keine Instabilität UEx
+ Sprungtest nicht bestanden	a (richtig positiv)	b (falsch positiv)
- Sprungtest bestanden	c (falsch negativ)	d (richtig negativ)

Also wäre die Sensitivität die Wahrscheinlichkeit des Nichtbestehens eines Sprungtests bei Instabilität der unteren Extremität. Die Spezifität dagegen wäre die Wahrscheinlichkeit, eines bestandenen Tests bei Abwesenheit einer Instabilität.

In der Physiotherapie sind meist nicht die Sensitivität und die Spezifität von primärer Bedeutung, sondern hauptsächlich der prädiktive (voraussagbare) Wert eines positiven oder negativen Tests, so Oesch et al. (2011). Physiotherapeuten und Physiotherapeutinnen interessiert meistens mehr der positiv prädiktive Wert, nämlich wie gross die Wahrscheinlichkeit ist, dass wenn der Test positiv ausfällt, man auch krank ist. Der positive prädiktive Wert wird mit $a/(a+b)$ berechnet und beschreibt, auf unsere Fragestellung bezogen, die Wahrscheinlichkeit, dass eine Instabilität vorliegt, wenn der Sprungtest nicht bestanden wird. Der negative prädiktive Wert, der mit $d/(c+d)$ berechnet wird, beschreibt in unserem Fall die Wahrscheinlichkeit, dass keine Instabilität vorliegt, wenn der Sprungtest bestanden wird.

2.9 Relevante Kennwerte

Die Studien, die in dieser Arbeit verwendet werden, beinhalten verschiedene Kennwerte, die mit der Reliabilität und der Validität zusammenhängen. Für das optimale Verständnis der verwendeten Studien werden diese Kennwerte nun beschrieben.

2.9.1 Standardabweichung (SD = Standard Deviation)

Die Standardabweichung ist ein Mass für die Streuung der Messwerte. Sie sagt aus, wie weit sich einzelne Datenpunkte innerhalb einer Stichprobe vom Mittelwert oder der Stichprobe unterscheiden (MatheGuru, o.D.).

Laut de With (2017) wird die Standardabweichung (teilweise auch mit s abgekürzt), aus der Quadratwurzel der Varianz berechnet. Für die Varianz wiederum werden die Differenzen aller Werte vom Mittelwert (\bar{x}) jeweils quadriert, aufsummiert und durch die Anzahl aller Werte (n) geteilt. Der Mittelwert ist die Summe aller Werte, geteilt durch die

Anzahl aller Werte. Daraus ergibt sich für die Standardabweichung folgende Formel:

$$s = \sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2) / n}$$

Die Verwendung der drei Kennwerte Mittelwert, Varianz und Standardabweichung macht nur dann Sinn, wenn die betreffende Verteilung eingipflig und symmetrisch ist.

2.9.2 Standardmessfehler (SEM = Standard Error of Measurement) und kleinste erkennbare Veränderung (SDC = Smallest Detectable Change)

Der Standardmessfehler ist nichts anderes als die Wurzel aus der Fehlervarianz, also $SEM = \sigma(\text{error})$ (Meichtry, 2018).

Bei Verlaufsmessungen, welche die Veränderung einer Variablen zwischen zwei Messzeitpunkten aufzeigen soll, ist der Fehler grösser als bei einer Einzelmessung. Aus diesem Grund ist der Standardmessfehler einer Veränderung (SEM_D) = $\sqrt{2} \times SEM$ (Meichtry, 2018).

Wie gross muss eine Veränderung sein, dass sie wirklich als Veränderung erkennbar ist und nicht von einem Messfehler stammt? Genau dies berechnet die kleinste erkennbare Veränderung (SDC) (Meichtry, 2018).

$$SDC = 1.96 \times \sqrt{2} \times SEM$$

Um mit 5% Risiko sagen zu können, dass es sich wirklich um eine Veränderung und nicht um einen Messfehler handelt, muss die Veränderung grösser als $1.96 \times \sqrt{2} \times SEM$ sein, also fast dreimal so gross wie der Standardmessfehler.

2.9.3 Konfidenzintervall (CI = Confidence Intervall)

In der Forschung werden meist Stichproben untersucht. Dies führt dazu, dass die erhobenen Parameter selten die Grundgesamtheit widerspiegeln, sondern lediglich diejenige der Stichprobe. Konfidenzintervalle definieren einen Bereich, in dem sich der "wahre" Wert des Parameters mit einer gewissen Wahrscheinlichkeit befindet (StatistikGuru, o.D.).

2.10 Verwendete Verfahren

Gleich wie bei den Kennwerten, werden die in den untersuchten Studien angewendeten Verfahren genauer erläutert.

2.10.1 Korrelation nach Bravais-Pearson

Der lineare Zusammenhang von zwei intervallskalierten Variablen kann mit der Korrelation nach Bravais-Pearson berechnet werden. Voraussetzungen für die Korrelation nach Bravais-Pearson sind mindestens intervallskalierte, normalverteilte Variablen. Der

Zusammenhang zwischen den Variablen muss linear sein (UZH – Methodenberatung, o.D.).

2.10.2 Rangkorrelationsanalyse nach Spearman (Spearman Rank)

Die Rangkorrelationsanalyse nach Spearman berechnet den linearen Zusammenhang zweier mindestens ordinalskalierten Variablen. Der Spearmans Rang-Korrelationskoeffizient ist das nicht-parametrische Äquivalent des Pearson Korrelationskoeffizienten. Er wird angewendet, wenn die Voraussetzungen für ein parametrisches Verfahren nicht gegeben sind. Die verwendeten Daten müssen nicht normalverteilt und die Variablen lediglich ordinalskaliert sein. Der Spearmans Rang-Korrelationskoeffizient kann auch bei kleinen Stichproben berechnet werden (UZH – Methodenberatung, o.D.).

2.10.3 Intraklassen-Korrelations-Koeffizient (ICC = Intraclass Correlation Coefficient)

Die Intraklassen-Korrelation ist ein parametrisches Verfahren, das eingesetzt wird, wenn mehr als zwei Beobachtende vorhanden sind und/oder mehrere Beobachtungszeitpunkte miteinander verglichen werden sollten. Das zugehörige Mass ist der Intraklassen-Korrelationskoeffizient, der mindestens intervallskalierte Daten voraussetzt (Wikipedia, o.D.). Insgesamt lassen sich sechs verschiedene Arten des Intraklassenkorrelations-Koeffizienten (ICC) unterscheiden, welche sich alle an der Basisformel der Reliabilität (siehe 2.7) orientieren (de Vet et al., 2011). Drei ICCs beziehen sich auf Einzelmessungen, drei weitere auf Durchschnittsmessungen. Für diese Bachelorarbeit ist der ICC(1.1) wichtig, der bei der Test-Retest-Reliabilität berechnet wird. Dieser ist

derselbe wie die Basisformel für die Reliabilität:
$$ICC(1.1) = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(error)}$$

2.10.4 Bland-Altman-Diagramm

Das Bland-Altman-Diagramm ist eine graphische Darstellungsmethode und stellt Differenz versus Mittelwerte für den Vergleich zweier Messmethoden dar. Somit lässt sich mit diesem Diagramm beurteilen, wie hoch die Schwankungsbreite der Abweichungen ist, ob eine Messmethode höher oder tiefer misst als andere und ob die Abweichung der Methoden von der Höhe der Werte abhängig ist. Vor allem für den Vergleich neuer Messverfahren mit einem Goldstandard werden solche Diagramme gebraucht (Wikipedia, o.D.).

2.10.5 Mann-Whitney-U-Test

Der Mann-Whitney-U-Test prüft, ob Mittelwert, Modalwert und Median zweier unabhängiger Stichproben verschieden sind. Der Mann-Whitney-U-Test wird angewendet, wenn die Voraussetzungen für ein parametrisches Verfahren nicht erfüllt sind und somit der t-Test für unabhängige Stichproben nicht angewendet werden kann. Anders als beim t-Test für unabhängige Stichproben muss die abhängige Variable beim Mann-Whitney-U-Test nur ordinalskaliert sein und die Daten müssen nicht normalverteilt sein. Der Mann-Whitney-U-Test kann auch für kleine Stichproben berechnet werden (UZH – Methodenberatung, o.D.).

Bezogen auf diese Bachelor-Arbeit bedeuten zwei unabhängige Stichproben, dass zwei verschiedene Gruppen miteinander verglichen wurden (Meichtry, 2018).

2.10.6 ANOVA (analysis of variance)

Bei der ANOVA handelt es sich um eine Erweiterung des t-Tests für unabhängige Stichproben, bei der mehr als zwei Gruppen miteinander verglichen werden können (Meichtry, 2018).

2.10.7 T-Test für abhängige Stichproben (paired sample t-test)

Der t-Test für abhängige Stichproben sagt aus, ob die Mittelwerte zweier abhängiger Stichproben verschieden sind (UZH – Methodenberatung, o.D.). Zwei abhängige Stichproben bedeuten, dass der Vergleich zwischen der gleichen Personengruppe zu zwei unterschiedlichen Zeitpunkten gezogen wurde (Meichtry, 2018).

Die Voraussetzungen dafür sind, dass die abhängige Variable mindestens intervallskaliert ist und eine Normalverteilung vorliegt (UZH – Methodenberatung, o.D.).

2.10.8 Wilcoxon- Test (Wilcoxon signed-rank test)

Im Gegensatz zum Mann-Whitney-U-Test untersucht der Wilcoxon-Test, ob Mittelwert, Modalwert und Median zweier abhängiger Stichproben verschieden sind. Der Wilcoxon-Test wird ebenfalls dann verwendet, wenn die Voraussetzungen für ein parametrisches Verfahren und somit für einen t-Test für abhängige Stichproben nicht erfüllt sind. Gleich wie beim Mann-Whitney-U-Test müssen die Variablen beim Wilcoxon-Test lediglich ordinalskaliert und die Daten nicht normalverteilt sein (UZH – Methodenberatung, o.D.).

2.10.9 Post-hoc-Tests

Laut Wikipedia (2018a) handelt es sich bei den Post-hoc-Tests um Signifikanztests. Im Vergleich mit der einfaktoriellen ANOVA, bei der nur festgestellt werden kann, ob in einer

Gruppe von Mittelwerten signifikante Unterschiede vorhanden sind, geben die Post-hoc-Tests noch Auskunft darüber, welche Mittelwerte sich signifikant voneinander unterscheiden.

Es werden sechs verschiedene Post-hoc-Tests unterschieden. Beispielsweise der Scheffé-Test, welcher in einer der ausgewählten Studien verwendet wird.

3 Methode

3.1 Literaturrecherche

Die Arbeit ist in Form eines Literaturreviews verfasst. Dieses ist gegliedert in einen Einleitungs-, Methoden-, Ergebnis- und Diskussionsteil, sprich nach dem EMED-Format. In den Datenbanken «PubMed», «CINAHL Complete», «MEDLINE» und «AMED» wurde nach geeigneter und relevanter Literatur gesucht. Diese Datenbanken wurden deshalb ausgewählt, weil sie einen Grossteil der für die Physiotherapie relevanten Literatur abdecken. In den Datenbanken «PubMed» und «CINHAL Complete» wurde die Suche mit Medical Subject Headings (MeSH) gestartet. Die Datenbankrecherche mit MeSH Terms macht es möglich, dass mit einem Suchbegriff etliche Unterbegriffe mitenthalten sind. Abbildung 2 zeigt den exemplarischen Suchstamm eines MeSH Terms.

Abbildung 2 MeSH Term «Reproducibility of Results» (PubMed, 2018)



Dem MeSH Term «Reproducibility of Results» sind Begriffe wie «Data Accuracy» und «Dimensional Measurement Accuracy» untergeordnet. Diesen sind wiederum die Begriffe Reproducibility of Findings, Reliability of Results, Reliability (Epidemiology), Validity (Epidemiology), Validity of Results, Face Validity, Reliability and Validity, Validity and Reliability, Test-Retest Reliability und Reliabilities untergeordnet.

Die MeSH Terms wurden mit Keywords ergänzt. In Datenbanken, in denen es keine MeSH Terms gibt, wurde eine reine Keyword Suche durchgeführt. Die in dieser Arbeit verwendeten Keywords sind in Tabelle 6 ersichtlich.

Des Weiteren wurden sowohl Trunkierungen wie «validit*» oder «reliabilit*» als auch die

Boolschen Operatoren AND und OR verwendet. Die genaue Suchmatrix ist im Anhang ersichtlich.

Tabelle 6 Keywords und MeSH Terms

Schlüsselwörter	Synonyme, Unterbegriffe, Oberbegriffe	Englische Keywords	MeSH Terms
Untere Extremität		Lower extremity	«Lower Extremity» [MeSH Terms]
Validität		Validity Validities Accuracy	«reproducibility of results» [MeSH Terms]
Reliabilität	Reproduzierbarkeit Zuverlässigkeit Messgenauigkeit Übereinstimmung (z.B. bei Wiederholungen)	Reliability Reliabilities Reliability of Results Reproducibility Repeatability Precision Concordance Dependability	
Sprungtests		Hop Test	
Front Hop Test	Single Leg Hop Test Single Leg Hop Test (for distance) One Legged Hop (for distance)		
Side Hop Test			
Square Hop Test			

3.2 Eingrenzen der gefundenen Literatur

In den oben genannten Datenbanken wurden insgesamt 108 Treffer gefunden (Doppelnennungen inbegriffen). Von diesen Treffern wurde das Abstrakt gelesen und anhand dessen entschieden, ob sie für unsere Fragestellung relevant sind. Insgesamt wurden zwölf Studien als relevant eingestuft. Über die Referenzen vom OSINSTITUT wurde eine weitere Studie gefunden, die für unsere Fragestellung relevant ist und welche unter unseren Suchbegriffen auf den Datenbanken nicht gefunden wurde. Von diesen nun insgesamt dreizehn Studien wurden zwei ausgeschlossen, da sich kein Volltext finden liess, eine weitere, weil sie eine andere Version des Front Hop Tests untersuchte und eine vierte Studie, da sie den Front Hop nur nebensächlich analysiert. Die Begründungen für den Ausschluss dieser Studien sind in Tabelle 7 ersichtlich.

Tabelle 7 Ein- und Ausschlusskriterien

Einschlusskriterium	Ausschlusskriterium
<ul style="list-style-type: none"> • Vorhanden sein eines Volltextes in englischer oder deutscher Sprache • Der untersuchte Sprungtest muss die gleiche Definition haben wie die drei qualitativen Sprungtests des RTA-Algorithmus 	<ul style="list-style-type: none"> • Studie älter als 20 Jahre

Danach blieben neun Studien übrig, die im Anhang aufgelistet sind. Diese Studien wurden mittels eines von den Autorinnen selbst erstellten Rasters eingeteilt (siehe Tabelle 8). Anhand des Rasters erhielten die Studien jeweils einen Punkt, wenn sie Reliabilität und Validität untersuchten. Weitere Punkte wurden vergeben, wenn sie einen oder mehrere der quantitativen Hop Tests des RTA-Algorithmus beinhalteten, sowie einen Punkt, wenn Personen mit Verletzungen der unteren Extremität Teil der Studie waren. Die Autorinnen dieser Arbeit gehen davon aus, dass nur dann wirklich eine relevante Aussage über die Validität gemacht werden kann, wenn Personen auch eine Verletzung an der unteren Extremität aufweisen. Zwei weitere Punkte wurden verteilt, wenn bei der Reliabilität die Test-Retest-Reliabilität und bei der Validität die Kriteriums- oder die Konstruktvalidität untersucht wurde.

Tabelle 8 Einteilung der Studien

Studie	Reliabilität ist Teil der Studie	Validität ist Teil der Studie	Front Hop Test	Side Hop Test	Square Hop Test	Pers. mit einer Verletzung der unteren Extremität sind Teil der Studie	Untersucht die Test-Retest-Reliabilität	Die Kriteriums- oder die Konstruktvalidität ist Teil der Studie	Total
1	1	0	1	0	0	0	1	0	3/8
2	0	1	1	0	0	1	0	1	4/8
3	1	0	1	1	0	0	1	0	4/8
4	1	1	1	1	1	1	1	1	8/8
5	0	1	1	0	0	1	0	1	4/8
6	1	1	1	0	0	0	0	1	4/8
7	1	1	1	0	0	0	0	1	4/8
8	1	1	1	0	0	1	1	1	6/8
9	1	1	1	0	1	0	1	0	5/8

Da zu jedem Sprungtest mindestens zwei Studien vorliegen sollten, mussten die Studien **3, 4 und 9** ausgewählt werden, da sie als einzige den «Square Hop» und oder den «Side Hop» untersuchen. Von den restlichen Studien wurde die Studie 8 ausgewählt, da sie mit 6/8 Punkten die meisten Treffer der noch übriggebliebenen Studien erzielte. Somit liegen vier Studien zum «Front Hop Test», zwei Studien zum «Square Hop Test» und zwei Studien zum «Side Hop Test» vor.

Die für diese Bachelorarbeit verwendeten Studien sind in folgender Tabelle ersichtlich.

Tabelle 9 Übersicht der in dieser Arbeit analysierten Studien

Studie 1	Kockum, B., & Heijne, A. I. L. M. (2014). Hop performance and leg muscle power in athletes: Reliability of a test battery. <i>Physical Therapy in Sport</i> , S. 222–227.
Studie 2	Gustavsson, A., Neeter, C., Thomeé, P., Grävare Silbernagel, K., Augustsson, J., Thomeé, R., & Karlsson, J. (2006). A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. <i>Knee Surgery, Sports Traumatology, Arthroscopy</i> , 14(8), S. 778–788.
Studie 3	Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. <i>Journal of Orthopaedic & Sports Physical Therapy</i> , 44(12), S. 947–954.
Studie 4	Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. <i>American Physical Therapy Association</i> , 87(3), S. 337–349.

3.3 Zusätzlich verwendete Literatur und Quellen

Da die Arbeit auf dem RTA-Algorithmus basiert, wird zusätzlich Literatur verwendet, welche auf der Internetseite des OSINSTITUT zugänglich ist. Dort referenzierte Quellen werden ebenfalls zur Bearbeitung der Thematik herangezogen.

Zur Bearbeitung der Reliabilität und der Validität wurde den Autorinnen von ihrer begleitenden Lehrperson folgendes Buch empfohlen, das auch immer wieder verwendet wurde:

- Measurement in Medicine
Vet, H. C. W. de, Terwee, C. B., Mokkink, L. B., Knol, D. L., (2011)

Zudem erhielten die Autorinnen die Möglichkeit, bei ihrer begleitenden Lehrperson eine Vorlesung zur Reliabilität und Validität zu besuchen.

3.4 Arbeitsinstrument zur kritischen Würdigung der Studien

Der systematisch kritische Würdigungsprozess der ausgewählten Studien wird mittels AICA-Raster (Arbeitsinstrument für ein Critical Appraisal) festgehalten. Das AICA-Raster ist ein von Ris und Preusse-Bleuler (2015) erstelltes Arbeitsinstrument, um Studien systematisch zusammenzufassen und zu würdigen. Das Arbeitsinstrument ist nach dem EMED-Format aufgebaut. Das AICA-Raster wurde ausgewählt, weil es für die Beurteilung quantitativer Studien geeignet ist und im Rahmen der Vorbereitungsmodule für die Bachelorarbeit der Umgang damit schon geübt wurde.

4 Ergebnisse

Im folgenden Teil dieser Bachelorarbeit werden die zur Bearbeitung der Leitfrage wichtigen Ergebnisse der vier bearbeiteten Studien präsentiert und gewürdigt. Die ausführlichen Zusammenfassungen und Würdigungen im AICA-Raster sind im Anhang einsehbar.

4.1 Studie 1 von Kockum und Heijne (2014)

4.1.1 Zusammenfassung

Ziel: Das Ziel der Studie ist es, sowohl die absolute und relative Reliabilität, als auch die kleinste erkennbare Veränderung (SDC) bei drei häufig gebrauchten Sprungtests, zwei Krafttests und beim Single Leg Squat Jump herauszufinden.

Stichprobe: Die Stichprobe besteht aus 18 gesunden Sportlerinnen und Sportlern. Genauere Angaben dazu sind in Tabelle 10 ersichtlich.

Tabelle 10 Übersicht Stichprobe der Studie 1

	Stichprobe
Anzahl Teilnehmende	<ul style="list-style-type: none"> • 18 (9 Frauen/9 Männer)
Alter (Durchschnitt)	<ul style="list-style-type: none"> • 23.4 Jahre
Einschlusskriterien	<ul style="list-style-type: none"> • Aktive Sporttreibende, die seit mindestens drei Monaten sechs bis acht Trainingseinheiten pro Woche absolvieren • Keine Schmerzen (in den letzten drei Monaten) • Keine reduzierte Leistung (in den letzten drei Monaten)
Drop-Outs	<ul style="list-style-type: none"> • 4

Messverfahren: Die beiden für diese Arbeit wichtigen Hop Tests (Front Hop und Side Hop) werden nach Neeter et al. (2006), respektive nach Gustavsson et al. (2006) und Itoh et al. (1998) durchgeführt. Im Gegensatz zur Definition der Tests in dieser Arbeit, müssen die Studienteilnehmenden die Hände dabei in die Hüfte stützen.

Datenerhebung:

- Front Hop: Drei korrekte Sprünge pro Bein werden aufgezeichnet. Im Voraus können die Probandinnen und Probanden 3-5 Testsprünge pro Bein durchführen.
- Side Hop: Die Anzahl korrekter Sprünge innerhalb von 30 Sekunden wird gezählt. Die Probandinnen und Probanden dürfen vor dem zählenden Durchlauf 10-15 Testsprünge absolvieren.

Statistische Verfahren:

Die verwendeten statistischen Verfahren sind in Tabelle 11 ersichtlich.

Tabelle 11 Statistische Verfahren der Studie 1

Vergleich der Ergebnisse	<ul style="list-style-type: none"> • Zusammenhang zwischen den Tests: paired-sample t-Test • Signifikanzniveau: $p < 0.05$ • Abweichungen, Streuungen: Bland und Altman Diagramm
Reliabilität	<ul style="list-style-type: none"> • Relative Reliabilität: ICC(2.1) • Absolute Reliabilität: SEM, SEM%, SDC, SDC%

Resultate: Abgesehen vom Side Hop für das linke Bein kann bei den absoluten Werten kein signifikanter Unterschied von der ersten zur zweiten Testgelegenheit festgestellt

werden. Die ICC-Werte sind vor allem beim Front Hop sehr hoch (0.97 für beide Beine). Der Side Hop weist für die linke Seite ebenfalls einen sehr guten ICC-Wert (0.96) auf. Die rechte Seite liegt mit einem ICC von 0.84 tiefer.

Der SEM liegt beim Front Hop zwischen 5 – 5.3 cm. Der Side Hop schwankt zwischen 2.5 Wiederholungen für das linke Bein und 5.4 Wiederholungen für das rechte Bein. Somit liegt der SEM des Side Hops mit links ausgeführt bei 11.1%, was die anderen Prozentzahlen deutlich übersteigt. Dementsprechend verhält sich auch der SDC-Wert.

Schlussfolgerung: Laut Kockum und Heijne (2014) zeigen die ICC Werte für den Front Hop und den Side Hop gute bis absolute Übereinstimmung auf. Unbekannt ist, ob diese Testbatterie bei verletzten Personen durchführbar ist. Mit Personen, welche sich in einem bereits fortgeschrittenen Rehabilitationsstadium befinden, könnte man versuchen, die Testbatterie anzuwenden.

Limitationen: Als Limitation sehen die Forschenden einzig die kleine Stichprobengrösse, bestehend aus 14 Personen. Diese beziehe sich allerdings auf eine Empfehlung von Bruton et al. (2000), wonach diese Grösse der Stichprobe ausreichen soll.

4.1.2 Würdigung

Die Untersuchung der absoluten und relativen Reliabilität von verschiedenen funktionellen Tests ist für die Beantwortung der Fragestellung dieser Bachelorarbeit relevant und aufschlussreich.

Kritisch zu betrachten ist, dass die Studie keine Fragestellung oder Hypothese aufweist. Auch stellen die Autorinnen dieser Bachelorarbeit in Frage, ob es sinnvoller wäre, drei anstatt nur zwei Datenerhebungen durchzuführen. Laut der Studie von Reid et al. (2007) findet bei der ersten Testdurchführung ein intensives motorisches Lernen statt. Dies könnte zu einem Bias führen (siehe 5.1.1). Zudem ist nicht bekannt, ob die Probandinnen und Probanden über ihre Ergebnisse informiert wurden. Dies könnte eine weitere Quelle für einen Bias darstellen (siehe 5.1.1).

Weiter sollte die kleine Stichprobe kritisch betrachtet werden. Obwohl die Stichprobengrösse begründet wird, sind 14 Personen nach Ansicht der Autorinnen dieser Bachelorarbeit zu wenig aussagekräftig, um die Ergebnisse auf eine ganze Population anzuwenden. Auch wird die Population von den Forschenden nicht explizit definiert.

4.2 Studie 2 von Gustavsson et al. (2006)

4.2.1 Zusammenfassung

Die Grundlage der Studie von Gustavsson et al. (2006) bildet folgende Hypothese: Bei einer Testbatterie, die verschiedene Sprungqualitäten untersucht, ist die Wahrscheinlichkeit, eine Diskrepanz bei den Sprungtests aufzudecken, höher, als wenn nur ein einzelner Sprungtest durchgeführt wird. Zu den verschiedenen Sprungqualitäten gehören eine maximale Sprungleistung auf einem Bein, kombiniert mit einem Sprungtest, der auf Erschöpfung abzielt. Anhand dieser Hypothese wird das Ziel der Studie definiert.

Ziel: Das Ziel dieser Studie ist, eine Testbatterie aus Sprungtests zu entwickeln, anhand derer man zwischen der verletzten und unverletzten Seite bei Personen mit einer VKB-Verletzung unterscheiden kann. Dazu gehören eine hohe Test-Retest-Reliabilität, Sensitivität, Spezifität und Accuracy (Treffgenauigkeit). Um dieses Ziel zu erreichen, untersuchten Gustavsson et al. (2006) den Vertical Jump, Front Hop, Drop Jump followed by a Double Hop for distance, Square Hop und Side Hop Test.

Stichprobe: Die Stichprobe besteht aus insgesamt 80 Personen, unterteilt in drei Gruppen. Genauere Angaben dazu sind der untenstehenden Tabelle, sowie dem Anhang zu entnehmen.

Tabelle 12 Übersicht Stichprobe der Studie 2

	VKB konservativ	VKB operativ	Gesund (Kontrollgruppe)
Anzahl Teilnehmende	<ul style="list-style-type: none">• 30• (18m/12w)	<ul style="list-style-type: none">• 35• (25m/10w)	<ul style="list-style-type: none">• 15• (9m/6w)
Alter Männer	<ul style="list-style-type: none">• 28 ± 7 Jahre	<ul style="list-style-type: none">• 27 ± 8 Jahre	<ul style="list-style-type: none">• 29 ± 5 Jahre
Alter Frauen	<ul style="list-style-type: none">• 36 ± 8 Jahre	<ul style="list-style-type: none">• 27 ± 7 Jahre	<ul style="list-style-type: none">• 26 ± 4 Jahre
Anzahl Wochen postoperativ/ posttraumatisch	<ul style="list-style-type: none">• 48 ± 43	<ul style="list-style-type: none">• 28 ± 2	<ul style="list-style-type: none">• /
Ausschlusskriterien	<ul style="list-style-type: none">• Akute Schmerzen oder Schwellung• Vorgängige Operation an beiden Beinen		<ul style="list-style-type: none">• Rücken-, Hüft-, Knie- oder Fussbeschwerden

Messverfahren: Der Square Hop fand in einer modifizierten Version von Östenberg et al. (1998) statt. Der Side Hop wurde nach Itoh et al. (1998) modifiziert. Die zwei Tests

unterscheiden sich insofern von der Definition in dieser Arbeit, als dass die zu testenden Personen die Hände hinter dem Rücken halten mussten und diese nicht frei einsetzen durften.

Vor der Durchführung absolvierten die Studienteilnehmenden ein vorgegebenes Warm-Up. Weiter definiert wurde, dass die Personen mit einer VKB-Verletzung mit der nichtbetroffenen Seite zu starten hatten.

Datenerhebung: Die Daten der fünf Tests sind von allen Studienteilnehmenden erhoben worden. Für die folgenden drei, für diese Bachelorarbeit relevanten Tests, wird die Datenerhebung etwas genauer beschrieben.

- Front Hop: Drei bis fünf Probesprünge waren den Testpersonen erlaubt, bevor sie drei zählende Sprünge durchführten.
- Side Hop: Der Side Hop wurde nur einmal pro Seite getestet. Zwischen dem Seitenwechsel wurde eine dreiminütige Pause eingelegt.
- Square Hop: Dieser weist dieselbe Datenerhebung wie der Side Hop auf.

Statistische Verfahren:

Die verwendeten statistischen Verfahren sind in Tabelle 13 ersichtlich.

Tabelle 13 Statistische Verfahren der Studie 2

Test-Retest-Reliabilität	<ul style="list-style-type: none"> • ICC • Methodological Error • Wilcoxon’s signed rank Test • Mann-Whitney U-Test • Spearman Rank • Signifikanzniveau: $p \leq 0.05$
Validität	<ul style="list-style-type: none"> • LSI • Sensitivität, Spezifität und Accuracy

Resultate: Zwischen der zweiten und dritten Testgelegenheit wird kein signifikanter Unterschied festgestellt.

Die ICC-Werte zwischen der zweiten und der dritten Testgelegenheit liegen bei 0.95 (Front Hop), 0.85 (Square Hop) und 0.93 (Side Hop), was laut Gustavsson et al. (2006) auf eine hohe Reliabilität hindeutet (siehe auch 5.1.2).

Die LSI-Seitenunterschiede sind bei den Personen mit einer VKB-Verletzung grösser als bei der Kontrollgruppe. Einzig der Square Hop der konservativ behandelten Studienteilnehmenden zeigt keinen wesentlichen Unterschied auf.

Bezüglich Sensitivität, Spezifität und Accuracy gehen die Werte teils stark auseinander. Untenstehende Tabelle soll dies veranschaulichen. Mehr dazu ist in Kapitel 5.2.1 ersichtlich.

Tabelle 14 Ergebnisse für Sensitivität, Spezifität und Accuracy der Studie 2 (angepasst nach Gustavsson et al. (2006))

	Front Hop	Side Hop	Square Hop
Sensitivität (operativ) in %	63	69	51
Sensitivität (konservativ) in %	53	77	53
Spezifität (Kontrollgr.) in %	100	87	67
Accuracy (operativ) in %	74	74	56
Accuracy (konservativ) in %	69	80	58

Schlussfolgerung: Gustavsson et al. (2006) beschreiben, dass das zentrale Ergebnis ihrer Studie darin besteht, dass eine Testbatterie, bestehend aus dem Vertical Jump, dem Front Hop und dem Side Hop, eine hohe Fähigkeit aufweist, um zwischen der verletzten und nicht verletzten Seite von VKB-Patientinnen und -Patienten unterscheiden zu können. Für die genaue Aussagekraft seien jedoch weitere Untersuchungen nötig. Die Forschenden schlussfolgern, dass Testbatterien sinnvoll sind, da diese genauere Werte hervorrufen, als ein einzelner Hop Test.

Beängstigend sei, dass nach 6 Monaten postoperativ, respektive 11 Monaten posttraumatisch nur einer von zehn Patientinnen und Patienten einen normalen LSI zustande bringt.

Zu beachten für diese Bachelorarbeit gibt es zudem, dass der Square Hop aus verschiedenen Gründen nicht in die Testbatterie aufgenommen wird, unter anderem, da er

nicht in der Lage ist, bei verletzten und gesunden Personen einen Seitenunterschied aufzuzeigen.

Limitationen: Die Forschenden beschreiben lediglich, wo noch mehr Forschungsbedarf vorhanden ist. Ansonsten werden keine Limitationen genannt.

4.2.2 Würdigung

Diese Studie behandelt eine, sowohl für die Berufspraxis als auch für diese Bachelorarbeit wichtige Thematik. Es können interessante und wichtige Werte des Front Hop, Square Hop und Side Hop Tests entnommen werden.

Die Stichprobe kann auf eine breite Population angewendet werden, da sich unter den Teilnehmenden auch weniger sportliche Personen befinden. Allerdings wird die Population, auf die sich die Ergebnisse beziehen, von den Forschenden nicht explizit definiert.

Kritisch zu beäugen gibt es einige Aspekte. Einerseits ist dies sicherlich die Zeitspanne, die zwischen Operation/Trauma und dem Studienstart liegt. Diese liegt bei 4 – 44 Monaten. Den Autorinnen dieser Arbeit stellt sich die Frage, ob eine Person, die sich 44 Monate (entspricht 3 Jahren und 8 Monaten) postoperativ/posttraumatisch befindet, nicht bereits als ausgeheilt gelten sollte. Man beachte diesbezüglich das Kapitel 2.5 der Wundheilungsphasen, nach welcher die Remodulierungsphase bei Ligamenten bis zu einem Jahr dauert. Demnach müsste ein VKB bereits wieder vollständig geheilt sein. Ein weiteres Problem bezüglich Zeitspanne sehen die beiden Verfasserinnen dieser Bachelorarbeit darin, dass zwischen der ersten und der zweiten Testgelegenheit 3-13 Tage, zwischen der zweiten und der dritten Testgelegenheit 3-19 Tage verstreichen. Diese Werte gelten sicherlich für die Kontrollgruppe. Ob auch für die VKB-Gruppen, ist unklar. Sollte dies der Fall sein, wäre es möglich, dass bei einigen Personen die drei Testgelegenheiten innerhalb von sieben Tagen stattfanden, während bei anderen Personen 33 Tage verstrichen. Diese Zeitspanne kann massive Unterschiede hervorrufen, da der Heilungsprozess je nach dem noch im Gange ist, was zudem zu einem Bias führen könnte (siehe 5.1.1).

Mangelhaft erscheint zudem, dass nur die ICC-Werte der Kontrollgruppe aufgeführt werden. Jene der VKB-Gruppen sind unbekannt. Erst durch genaues Betrachten der Tabelle wird ersichtlich, von welcher Stichprobengruppe die Werte stammen.

4.3 Studie 3 von Haitz et al. (2014)

4.3.1 Zusammenfassung

Ziel: Das Ziel dieser Studie ist, die Augenscheinvalidität, die Test-Retest- und die Interrater-Reliabilität des Functional Lower Extremitiy Evaluation (FLEE) zu bestimmen. Auch wollen die Forschenden die FLEE-Werte für gesunde Personen dokumentieren, um für Patientinnen und Patienten in der Rehabilitation einen Vergleich zu haben. Der FLEE besteht aus 8 funktionellen Assessments, darunter auch der Front Hop und Square Hop.

Stichprobe: Die Stichprobe besteht aus 49 Sportstudierenden. Nach der ersten Testgelegenheit sind aus verschiedenen Gründen neun Drop-Outs zu verzeichnen. Somit beenden insgesamt 40 Personen die Studie. Die nachfolgende Tabelle zeigt genauere Werte auf.

Tabelle 15 Übersicht Stichprobe der Studie 3

	Männer	Frauen	Total
Anzahl Teilnehmende	• 22	• 18	• 40
Alter	• 18 – 24 Jahre	• 18 – 22 Jahre	• 18 – 24 Jahre
Einschlusskriterien	• Unverletzt • An der Stanford University		

Messverfahren: Die Augenscheinvalidität wird anhand eines Fragebogens evaluiert. Dieser beinhaltet die zwei Fragen, wie wichtig die einzelnen Tests des FLEEs eingeschätzt und wie häufig die Tests in der Praxis angewendet werden. Die Durchführung des Front Hops geschieht anhand der Definition von Noyes et al. (1991), jene des Square Hops nach Caffrey et al. (2009).

Datenerhebung: Bei zwei Testgelegenheiten (sieben Tage dazwischen) werden bei allen Personen, abgesehen von den neun Drop-Outs, die Daten des FLEEs erhoben. Vor Testbeginn ist ein kurzes Warm-Up angeordnet. Die Hop Test-Messungen werden zweimal erhoben. Zur Erhebung der Daten sind ein Physiotherapeut, ein Sporttherapeut und ein Forschungsassistent, welcher vorgängig intensiv in die Tests eingeführt wurde, vor Ort.

Den Fragebogen zur Ermittlung der Augenscheinvalidität versenden die Forschenden an sportmedizinisches und sportphysiotherapeutisches Fachpersonal und an Sporttrainerinnen und Sporttrainer mit minimal 13 Jahren Berufserfahrung.

Statistische Verfahren:

Die in der Studie berechneten Werte sind in Tabelle 16 ersichtlich.

Tabelle 16 Berechnete Werte der Studie 3

Reliabilität	<ul style="list-style-type: none"> • ICC(2.1) Interrater-Reliabilität • ICC(3.1) Test-Retest-Reliabilität • SEM • SDC
Validität	<ul style="list-style-type: none"> • LSI • Wichtigkeit der Tests in % • Verwendung der Tests in der Praxis in %

Resultate: Ca. 70% der befragten Fachpersonen stufen den Front Hop im Gebiet der «hohen Wichtigkeit» ein. Beim Square Hop sind es knapp 60%. Die Häufigkeit der Anwendung liegt deutlich tiefer (weiteres siehe 5.2.1).

Bezüglich der Reliabilität weist die Studie ein gutes Resultat auf. Der Test-Retest ICC liegt bei 0.91 für den Front Hop und bei 0.83 für den Square Hop. Die Interrater-Reliabilität liegt mit 1.00 und 0.98 sogar noch höher.

Schlussfolgerung: Das Ziel der Studie ist erreicht. Bezüglich der Augenscheinvalidität sind die Ergebnisse für eine «hohe Wichtigkeit» der Tests ca. eineinhalb bis zweimal grösser als die der regelmässigen Anwendung. Laut den Forschenden könnte der Grund dafür in der noch nicht erfassten Reliabilität, sowie fehlender Normwerte liegen.

Limitationen: Die Forschenden diskutieren, dass nur eine der drei Personen, die den Test durchführt hatten, den Probandinnen und Probanden Instruktionen gab, obwohl alle drei unabhängig voneinander Resultate abnahmen und Punkte vergaben. Im wirklichen Leben sei es nicht so und deswegen könnten die Ergebnisse verzerrt sein. In Bezug darauf, dass

die Normaldaten je nach Sportart variieren können, sei die Stichprobe zu klein. Daher können die Daten nicht auf jeden übertragen werden.

4.3.2 Würdigung

Die Studie weist einen klaren theoretischen Hintergrund auf und das Ziel ist für diese Arbeit relevant. Es wird in der Studie weder eine Fragestellung noch eine Hypothese formuliert. Die Stichprobenziehung erscheint den Verfasserinnen dieser Arbeit als angebracht, wenn man annimmt, dass sich die Testungen auf eine sehr sportliche Population beziehen. Die Population wird von den Forschenden nicht explizit definiert. Inwiefern die hohe Drop-Out Rate von ca. 20% die Ergebnisse beeinflusst, ist nicht ersichtlich.

Bezogen auf die Interrater-Reliabilität ist positiv zu bemerken, dass die testenden Personen die Ergebnisse der anderen nicht erfahren. Unbekannt ist, ob die Probandinnen und Probanden die Ergebnisse wissen und dadurch eine Quelle für einen Bias vorliegt (siehe 5.1.1). Wie bereits in der Studie von Analog zu Kockum und Heijne (2014) wird die Test-Retest-Reliabilität zwischen der ersten und zweiten Testgelegenheit bestimmt. Aus diesem Grunde gilt es, den potentiellen Bias zu beachten (siehe 5.1.1). Kritisch betrachtet werden sollte auch die Zählart des Square Hops. So sind die Teilnehmenden dazu aufgefordert, die Runden laut mitzuzählen. Die Testperson zählt währenddessen die fehlerhaften Sprünge. Nach eigener Erfahrung können die Autorinnen dieser Bachelorarbeit sagen, dass dies beim Square Hop eine koordinative Höchstleistung bedingt. Die Fehlerquelle könnte dadurch massgebend beeinflusst werden. Im Allgemeinen sind die Resultate eher oberflächlich gehalten. Viele Resultate aus den Tabellen werden im Fliesstext nicht erwähnt.

4.4 Studie 4 von Reid et al. (2007)

4.4.1 Zusammenfassung

Ziel: Ziel dieser Studie ist, die Reliabilität und longitudinale Konstruktvalidität von vier Sprungtests, darunter auch der Front Hop, bei Personen während der Rehabilitation nach einer VKB-Rekonstruktion herauszufinden.

Die Konstruktvalidität wird anhand zweier Theorien überprüft. Einerseits sollen die Veränderungen bei den Sprungtests beim operierten Bein signifikant grösser sein als beim

Nichtoperierten. Andererseits wird die Theorie aufgestellt, dass Veränderungen im LSI mit Veränderungen bei den unten beschriebenen Fragebögen korrelieren sollen.

Stichprobe: 42 Personen mit einer VKB-Rekonstruktion nahmen an der Studie teil. Die genaueren Angaben zur Stichprobe sind in Tabelle 17 ersichtlich.

Tabelle 17 Übersicht Stichprobe der Studie 4

	Frauen	Männer
Anzahl Teilnehmende	19	23
Alter	15 – 40	15 – 45
Meniskus repariert (ja/nein)	12/7	8/15
Ausschlusskriterien	<ul style="list-style-type: none"> • Zusätzliche Ruptur des hinteren Kreuzbandes • Zusätzliche Ruptur des medialen Seitenbandes • Sonstige muskuloskelettalen Probleme (Rücken, Hüfte oder Fuss), die die Person am Hüpfen hindern könnten • Fortgeschrittene degenerative Veränderungen • Nicht fähig, Englisch zu reden, lesen schreiben oder verstehen 	

Messverfahren: Der Front Hop wird so durchgeführt, wie er in dieser Arbeit definiert wird. Für die Ermittlung der Validität werden zwei Fragebögen eingesetzt. Der Lower Extremity Functional Scale (LEFS) ist ein Fragebogen mit 20 Fragen, der zwischen 0 und 80 Punkten variiert. Je höher die Ergebnisse, desto höher der funktionelle Status der Person. Der Global Rating of Change Fragebogen beinhaltet die Veränderung in den letzten 6 Wochen. Dabei können die Studienteilnehmenden auf einer Skala von 0-7 zwischen schlechter geworden, gleichgeblieben oder besser aussuchen.

Datenerhebung: Die Datenerhebung erfolgt zu vier verschiedenen Zeitpunkten. Die drei ersten Testgelegenheiten erfolgen in der 16. Woche nach der VKB-Rekonstruktion mit 24-48 Stunden zwischen den einzelnen Testdurchgängen. Die erste Durchführung ist dazu gedacht, dass die Studienteilnehmenden die Sprungtests motorisch lernen können. Die zweite und dritte Durchführung wird angewendet, um die Reliabilität bestimmen zu können. Die vierte Durchführung, die 6 Wochen später erfolgt, dient zur Bestimmung der longitudinale Validität. Die Daten der vier Sprungtests und des LEFS-Fragebogens werden

bei allen vier verschiedenen Testgelegenheiten erhoben. Der Global Rating of Change Fragebogen wird nur bei der letzten Testgelegenheit erhoben.

Statistische Verfahren:

Die verwendeten statistischen Verfahren sind in Tabelle 18 ersichtlich.

Tabelle 18 Statistische Verfahren der Studie 4

Vergleich der Ergebnisse	<ul style="list-style-type: none"> • ANOVA • Scheffe Post-hoc-Test
Reliabilität	<ul style="list-style-type: none"> • ICC(2.1) • SEM • SDC
Validität	<ul style="list-style-type: none"> • Gepaarter t-Test (zwischen der vierten und dem Mittelwert der zweiten und dritten Testgelegenheit) • Pearson-Korrelations-Koeffizient (r) (Korrelation zwischen den LSI- und den Fragebögen-Werten) • Dabei gilt $r > 0.5$ = gut, $r > 0.36$ = mittel, $r > 0.2 - 0.35$ = niedrig und $r < 0.2$ = kein Beweis.

Resultate: Der Vergleich zwischen den Sprungtests anhand der ANOVA und des Post-hoc-Tests, zeigt einen signifikanten Unterschied zwischen der ersten und der zweiten Testgelegenheit, sowohl beim operierten als auch beim nichtoperierten Knie. Daraus schliessen Reid et al. (2007) auf einen erheblichen motorischen Lernprozess. Dieser flacht bei der dritten Testgelegenheit ab. Zwischen der zweiten und der vierten Testgelegenheit erzielt der Front Hop wieder einen signifikanten Unterschied.

Der Front Hop ist von den vier untersuchten Tests derjenige mit der höchsten relativen Reliabilität. Zudem weist er den niedrigsten SEM auf.

Die Veränderung der absoluten Punktzahl der Sprungtests beim operierten Bein sind statistisch grösser, als die Veränderung beim nichtoperierten Bein. Die Korrelation zwischen dem Front Hop und dem LEFS beträgt 0.37 (lower limit 95% CI 0.11), die zwischen dem Global Rating of Change liegt bei 0.48 (lower limit 95% CI 0.24) (weiteres dazu siehe 5.2.1).

Schlussfolgerung: Das Forschungsziel ist erreicht. Die Ergebnisse weisen auf eine gute Reliabilität hin und auch die Konstruktvalidität wird teilweise bestätigt. Laut den Forschenden unterstützen diese Ergebnisse die Interpretation, dass die Sprungtests sowohl in der Forschung als auch in der Praxis Verwendung finden sollten.

Limitationen: Limitationen werden in dieser Studie weder erwähnt, noch diskutiert.

4.4.2 Würdigung

Die Studie leitet das Thema und die Problemstellung anhand verschiedener Literatur logisch her. Positiv ist auch, dass die Studienteilnehmenden ihre Sprungtestergebnisse nicht erfahren und der motorische Lernprozess beachtet wird.

Die Stichprobengrösse wird anhand einer Sample Size Calculation berechnet. Die ermittelte Grösse von 50 Personen kann nicht ganz eingehalten werden. Nach Einschätzung der Autorinnen dieser Arbeit liegt die Stichprobengrösse mit 35 Teilnehmenden für die Reliabilität und 39 für die Validität, immer noch in einem angemessenen Rahmen.

Kritisch zu betrachten ist der LEFS, da dieser nach der Meinung der Autorinnen dieser Bachelorarbeit nicht dasselbe Konstrukt misst wie der Front Hop. So kann eine Person mit einem guten LEFS-Score trotzdem Mühe beim Front Hop aufweisen. Zudem ist das Datenniveau des LEFS und des Global Rating of Change ordinalskaliert. Die Korrelation jedoch wird anhand des Pearson-Korrelations-Koeffizienten berechnet, der mindestens intervallskalierte Daten voraussetzt. Weiter wird ein Wert von $r > 0.5$ als gut und ein Wert von $r = 0.36-0.5$ als mittel eingestuft. Dies halten die Verfasserinnen dieser Arbeit für eine eher niedrige Korrelation (siehe auch 5.2.1).

5 Diskussion

In diesem Teil der Arbeit werden die für die Beantwortung der Fragestellung wichtigen Ergebnisse der vier analysierten Studien sowohl in Bezug zueinander als auch zum theoretischen Hintergrund gesetzt und kritisch diskutiert.

In den folgenden Kapiteln sind jeweils die Erkenntnisse der Reliabilität und der Validität zu jedem Sprungtest einzeln in Diskussion. Am Anfang der beiden Unterkapitel «Reliabilität» und «Validität» wird noch vertiefter auf die jeweilige Begriffsbedeutung eingegangen. Die

Autorinnen dieser Bachelorarbeit beschreiben zudem, was sie unter diesen Begriffen erwartet haben und wie die Studien mit diesen umgehen.

5.1 Reliabilität

In allen vier Studien, die zur Ermittlung der Reliabilität der quantitativen Sprungtests des RTA-Algorithmus ausgesucht wurden, wird die Test-Retest-Reliabilität untersucht. In der Studie 3 wird zudem die Interrater-Reliabilität des Front Hops und Square Hops ermittelt. In dieser Bachelorarbeit ist vor allem die Test-Retest-Reliabilität von Interesse. Aus diesem Grunde wird im Folgenden hauptsächlich auf diese eingegangen.

Um sich der Formel nach de Vet et al. (2011) für die Reliabilität erneut bewusst zu sein, ist diese an dieser Stelle nochmals erwähnt: Reliabilität = $\frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(error)}$

Laut Meichtry (2017, S. 33) ist, wie bereits erwähnt, eine Messung dann reliabel, «wenn der (nicht-systematische) Messfehler relativ zur Heterogenität der Personen klein ist». Dabei ist bei Test-Retest-Studien die Quelle der Fehlervarianz die Patientin/der Patient selber (Meichtry, 2018).

Den systematischen Messfehler kennt man vor allem von der Validität. Darum wird im Folgenden der systematische Messfehler, der sich auf die Reliabilität bezieht, als Bias bezeichnet. So ist bei Studien oft nicht ersichtlich, ob eine Prüfung der Modellannahmen stattgefunden hat, oder ob Bias enthalten sind. Ein solcher Bias wäre beispielsweise, wenn die Studienteilnehmenden ihre Ergebnisse der Sprungtests kennen. Dies könnte die Reliabilität stark beeinflussen, da sie dann bei der zweiten Testgelegenheit versuchen, genauso weit wie bei der ersten zu springen. Werden alle Bias kontrolliert, bleibt nur noch der zufällige Fehler $\sigma^2(error)$ übrig, der in der oben genannten Formel vorkommt. Im Folgenden wird für die verschiedenen Studien beschrieben, ob Bias kontrolliert wurden. Weiter werden die verschiedenen Ergebnisse der Studien für den Front Hop, den Side Hop und den Square Hop Test, miteinander verglichen und diskutiert.

5.1.1 Kontrolle von Bias in den untersuchten Studien

Die genaue Ursache eines Bias festzustellen, ist schwierig, da zahlreiche in Frage kommen. Die Autorinnen dieser Bachelorarbeit haben sich anhand der Studien und eigenen Überlegungen auf die drei untenstehenden Ursachen geeinigt. Die Einschlussgründe sind nachfolgend ebenfalls erwähnt. Dabei ist den Verfasserinnen dieser Arbeit bewusst, dass es etliche andere Ursachen gäbe.

1. Kenntnis über die Sprungtestresultate des vorhergehenden Durchlaufes
2. Motorischer Lerneffekt
3. Vergangene Zeit zwischen den zwei Testgelegenheiten

Das Wissen über die Testresultate könnte die Reliabilität stark verändern, da die Personen beim zweiten Mal, bereits voreingenommen, versucht wären genau gleich weit zu springen, wie beim ersten Mal (siehe auch Punkt 5.1).

Zur Auswahl des Kriteriums «motorischer Lerneffekt» hat die Studie von Reid et al. (2007) geführt. Diese beschreibt, dass zwischen den Testergebnissen der ersten und zweiten Testgelegenheit ein signifikanter Unterschied festgestellt und anhand des motorischen Lerneffekts erklärt wird. In den Studien 1 und 3 wurde die Reliabilität zwischen der ersten und der zweiten Testgelegenheit bestimmt, wodurch sich die Frage stellt, ob dort der motorische Lerneffekt zu einem Bias führt.

Das dritte Kriterium wurde gewählt, weil in der Studie 2 die Werte für die Test-Retest-Reliabilität mit einem Zeitabstand von 3-19 Tagen erhoben werden. 19 Tage, also fast 2 ½ Wochen, könnten eine Veränderung bei der Wundheilung mit sich bringen, wodurch die Ergebnisse einseitig gerichtet wären. Werden die Tests bereits nach 24 Stunden durchgeführt, wie dies in Studie 4 der Fall ist, besteht die Gefahr, dass die Testpersonen unter Muskelkater leiden, welcher von der ersten Testdurchführung herrührt. Die Ergebnisse würden so allgemein schlechter ausfallen. Die Autorinnen dieser Arbeit kommen nach abwägen der Faktoren zum Schluss, dass die Zeitspanne zwischen zweier Tests optimalerweise zwischen 4-7 Tagen liegt.

In der Tabelle sind die drei ausgewählten, potenziellen Ursachen für Bias anhand der Studien dargestellt.

Tabelle 19 Übersicht potenzielle Ursachen für Bias

	Kenntnis der Sprungtestresultate	motorischer Lerneffekt	Zeit zwischen den zwei Testgelegenheiten
Studie 1	Keine Angabe	ICC wurde zwischen 1. und 2. Testgelegenheit bestimmt.	7-10 Tage
Studie 2	Keine Angabe	ICC wurde zwischen 2. und 3. Testgelegenheit bestimmt.	3-19 Tage
Studie 3	Keine Angabe	ICC wurde zwischen 1. und 2. Testgelegenheit bestimmt.	7 Tage
Studie 4	Pat. hatten keine Kenntnisse.	ICC wurde zwischen 2. und 3. Testgelegenheit bestimmt.	24-48 Stunden

5.1.2 Wichtige Masse für die Bestimmung der Reliabilität

ICC: Ein häufig verwendetes Mass für die Test-Retest-Reliabilität ist der ICC. Für die Berechnung der Test-Retest-Reliabilität wäre eigentlich der ICC(1.1) oder der ICC(3.1) angemessen. In zwei der vier bearbeiteten Studien wird der ICC(2.1) für die Test-Retest-Reliabilität berechnet. In Studie 2 ist nicht ersichtlich, welcher ICC berechnet wird. Lediglich in der Studie 3 wird der ICC(3.1) für die Test-Retest-Reliabilität berechnet. Der ICC(2.1) macht bei der Test-Retest-Reliabilität wenig Sinn, da dieser im Nenner die Testerin oder den Tester als weitere Fehlerquelle einrechnet (Meichtry, 2019). Allerdings ist er deshalb auch der konservativste Wert.

Laut Nunnally & Bernstein (1994, zit. nach de Vet et al., 2011) gilt ein ICC von 0.70 als akzeptabel. Anhand dieser Aussage ist in dieser Bachelorarbeit ein ICC von 0.70 akzeptiert, ein Wert von 0.80 gilt als gut und ein Wert von 0.90 als sehr gut.

Dass die Reliabilität und somit auch der ICC abhängig von der Heterogenität der Personen ist, ist laut de Vet et al. (2011) nicht unbedingt ein Nachteil. So ist die Heterogenität einfach ein Bestandteil der Reliabilität. Allerdings sollte beachtet werden, dass aus diesem Grunde der ICC nur bei ähnlichen Populationen angewendet wird, so de Vet. et al. (2011) weiter.

SEM: Bei Verlaufsmessungen ist laut Meichtry (2017) nicht die Reliabilität, und somit auch nicht der ICC-Wert von Interesse, sondern der SEM. Denn dieser ermöglicht es, zwischen dem Vorher und Nachher zu unterscheiden (Meichtry, 2017). Der SEM wird laut de Vet et al. (2011) immer in der Einheit der Messung angegeben. Aufgrund dessen ist es unmöglich, eine generelle Empfehlung abzugeben, ab wann ein SEM-Wert akzeptabel ist, so de Vet et al. (2011) weiter. Grundsätzlich kann gesagt werden, je kleiner der SEM, desto besser. Wenn Fachpersonen das Messinstrument kennen, werden sie ein Gefühl dafür haben, ob der SEM klein oder gross ist (de Vet et al., 2011).

SDC: Wird das Messinstrument, in unserem Falle die Sprungtests, für wiederholte Messungen eingesetzt, ist ein weiterer Parameter von grosser Bedeutung. Dies ist die kleinste erkennbare Veränderung (SDC), die bereits unter 2.9.2 erläutert wurde. So ist es sehr wichtig zu wissen, ab wann eine Veränderung als wirkliche Veränderung gilt und nicht einfach vom Messfehler stammen könnte.

5.1.3 Front Hop

In der untenstehenden Tabelle sind die wichtigsten Werte (ICC, SEM und SDC) für die Test-Retest-Reliabilität des Front Hops festgehalten. Zudem beinhaltet die Tabelle die Art der Stichprobe, bei der die Werte erhoben werden.

Tabelle 20 Ergebnisse zur Reliabilität des Front Hops

Studie 1	Studie 2	Studie 3	Studie 4
Stichprobe			
Gesunde, junge und sportliche Personen			Personen mit einer VKB-Ersatzplastik
ICC für die Test-Retest-Reliabilität			
ICC(2.1) re = 0.97 ICC(2.1) li = 0.97	ICC(?) alle Testpersonen = 0.95 (lower limit 95% CI 0.9) ICC(?) Männer = 0.86 (lower limit 95% CI 0.67) ICC(?) Frauen = 0.98 (lower limit 95% CI 0.93)	ICC(3.1) = 0.91 (lower limit 95% CI 0.87)	ICC(2.1) = 0.92
SEM und SEM%			
SEM re = 5 cm SEM li = 5.3 cm SEM% re = 3.5% SEM% li = 3.6%	Es wird kein SEM berechnet.	SEM = 7.7 cm	SEM% = +/- 3.49% (upper limit 95% CI 4.37)
SDC und SDC%			
SDC re = 13.9 cm SDC li = 14.6 cm SDC% re = 9.7% SDC% li = 10.1%	Es wird kein SDC berechnet.	SDC = 21.3 cm	SDC% (90% CI) = +/- 8.09% SDC% (95% CI) = +/- 9.67%

Interpretation der ICC-Werte: Die ICC-Werte der Studien 1-3 werden anhand einer gesunden, jungen und sportlichen Stichprobe erhoben. Die ICC-Werte werden je nach Studie anders präsentiert. Es werden jeweils auch verschiedene ICCs berechnet (siehe auch 5.1.2). Aus diesen Gründen und auch aus dem Grund, dass sich der ICC immer auf die jeweilige Stichprobe bezieht, ist ein direkter Vergleich der Ergebnisse nicht möglich. Praktisch alle ICCs liegen über 0.90. Einzig der ICC für Männer aus der Studie 2 liegt bei 0.86 (lower limit 95% CI 0.67). Allgemein kann gesagt werden, dass die Ergebnisse des

Front Hops für gesunde und junge Sportlerinnen und Sportler, eine gute bis sehr gute Test-Retest-Reliabilität aufweisen. Die Studie 4 ist die einzige, welche die Test-Retest-Reliabilität bei Personen mit einer VKB-Rekonstruktion ermittelt. Auch bei dieser Stichprobengruppe kann anhand des ICCs von 0.92 von einer sehr guten Test-Retest-Reliabilität ausgegangen werden.

Interpretation der SEM-Werte: Studie 1 und 3 berechneten den SEM für eine Stichprobe von gesunden und jungen Sportlerinnen und Sportlern. Dieser variiert zwischen 5.0 - 7.7 cm, bei einer durchschnittlichen Sprungdistanz von 140 cm. In Anbetracht dessen erachten die Autorinnen dieser Bachelorarbeit diese SEM-Werte als akzeptabel. In Studie 1 wird zudem die Prozentzahl des SEM, berechnet auf die gesamte Sprungdistanz, angegeben. Diese liegt bei 3.5-3.6%. In Studie 4 wurde der SEM% für eine Stichprobe von Personen nach einer VKB-Rekonstruktion berechnet. Dieser liegt mit 3.49% in einem ähnlichen Bereich.

Interpretation der SDC-Werte: Da der Front Hop auch für Verlaufsmessungen eingesetzt werden soll, ist der SDC ein wichtiges Mass. Anhand der Ergebnisse von Studie 1 und 3 kann man erst nach 13.9 - 21.3 cm von einer wirklichen Veränderung ausgehen. Für die Berufspraxis ist dies eine wichtige Angabe. So scheint eine Veränderung von 21 cm als sehr gross. Dennoch könnte diese anhand der Ergebnisse nur von Messfehlern stammen. In der Studie 4 wurde der SDC für Patientinnen und Patienten mit einer VKB-Rekonstruktion anhand eines 90% CIs berechnet. Die Autorinnen dieser Bachelorarbeit haben den SDC auch anhand eines 95% CIs umgerechnet, damit sich der Wert mit den anderen Werten vergleichen lässt. Der SDC% (95% CI) beträgt bei Studie 4 +/- 9.67% und ist somit mit dem SDC% der Studie 1 vergleichbar.

5.1.4 Side Hop

In der untenstehenden Tabelle sind die wichtigsten Werte (ICC, SEM und SDC) für die Test-Retest-Reliabilität des Side Hops und die Art der Stichprobe, bei der die Werte erhoben werden, festgehalten.

Tabelle 21 Ergebnisse zur Reliabilität des Side Hops

	Studie 1	Studie 2
Stichprobe	Gesunde, junge und sportliche Personen	
ICC für die Test-Retest-Reliabilität	ICC(2.1) re = 0.84 ICC(2.1) li = 0.96	ICC(?) alle Testpersonen = 0.93 (lower limit 95% CI 0.87) ICC(?) Männer = 0.78 (lower limit 95% CI 0.51) ICC(?) Frauen = 0.95 (lower limit 95% CI 0.83)
SEM und SEM %	SEM re = 5.4 Sprünge SEM li = 2.5 Sprünge SEM% re = 11.1% SEM% li = 5.5%	Es wird kein SEM berechnet.
SDC und SDC %	SDC re = 14.8 Sprünge SDC li = 7.0 Sprünge SDC% re = 30.7% SDC% li = 15.2%	Es wird keine SDC berechnet.

Interpretation der ICC-Werte: Die ICC-Werte, die für den Side Hop in Studie 1 und 2 berechnet werden, beziehen sich alle auf eine gesunde, junge und sportliche Stichprobe. Die meisten ICCs liegen über 0.90. Einzig der ICC aus Studie 1 für das rechte Bein und jener aus Studie 2 für die Männer liegen mit 0.84 und 0.78 tiefer. Alle Werte weisen darauf hin, dass die Ergebnisse des Side Hops für gesunde und junge Sportlerinnen und Sportler eine akzeptable bis sehr gute Test-Retest-Reliabilität aufweisen.

Interpretation der SEM-Werte: Lediglich Studie 1 berechnet den SEM für eine Stichprobe von gesunden, jungen und sportlichen Personen. Dieser variiert zwischen 2.5 - 5.4 Sprüngen, bei durchschnittlich 47 Sprüngen in 30 Sekunden. Daher erachten die Autorinnen dieser Bachelorarbeit diesen SEM als gut.

Interpretation der SDC-Werte: Wird der Side Hop als Verlaufsmessung eingesetzt, kann anhand der Ergebnisse von Studie 1 erst nach 14.8 Sprüngen von einer wirklichen Veränderung ausgegangen werden. Daraus wird ersichtlich, dass ein SEM-Wert, den die Autorinnen als gut einstufen, bei einer Verlaufsmessung in Form des SDC-Werts als sehr hoch erscheinen kann. Für die Praxis ist zu beachten, dass 14.8 Sprünge einem Drittel des Gesamtergebnisses entsprechen.

5.1.5 Square Hop

In der untenstehenden Tabelle sind die wichtigsten Werte (ICC, SEM und SDC) für die Test-Retest-Reliabilität des Square Hops und die Art der Stichprobe, bei der die Werte erhoben wurden, festgehalten.

Tabelle 22 Ergebnisse zur Reliabilität des Square Hops

	Studie 2	Studie 3
Stichprobe	Gesunde, junge und sportliche Personen	
ICC für die Test-Retest-Reliabilität	ICC(?) alle Testpersonen = 0.85 (lower limit 95% CI 0.71) ICC(?) Männer = 0.78 (lower limit 95% CI 0.5) ICC(?) Frauen = 0.89 (lower limit 95% CI 0.67)	Test-Retest-Reliabilität: ICC(3.1) = 0.83 (lower limit 95% CI 0.75)
SEM	Es wird kein SEM berechnet.	SEM = 5.8 Sprünge
SDC	Es wird kein SDC berechnet.	SDC = 16 Sprünge

Interpretation der ICC-Werte: Aufgrund der obenstehenden Werte kann ausgesagt werden, dass die Ergebnisse des Square Hops für gesunde und junge Sportlerinnen und Sportler eine akzeptable, bis gute Test-Retest-Reliabilität aufweisen.

Interpretation der SEM-Werte: Lediglich Studie 3 berechnet den SEM für eine Stichprobe von gesunden, jungen und sportlichen Personen. Dieser beträgt 5.8 Sprünge. Beim Square Hop liegt der Durchschnitt bei 62 Sprüngen in 30 Sekunden. Daher erachten die Autorinnen dieser Bachelorarbeit diesen SEM als gut.

Interpretation der SDC-Werte: Wird der Square Hop bei Verlaufsmessungen eingesetzt, kann anhand der Ergebnisse von Studie 3 erst nach 16 Sprüngen von einer wirklichen Veränderung ausgegangen werden. Prozentual ausgedrückt sind dies 25% mehr Sprünge als durchschnittlich. Dies zeigt wiederum auf, dass ein guter SEM-Wert bei Verlaufsmessungen einen tendenziell hohen SDC mit sich ziehen kann.

5.2 Validität

Wie unter 2.8 erläutert, beschreibt die Validität, wie gut ein Instrument misst, was es vorgibt zu messen (Oesch et al., 2011). Ein Messinstrument kann nur dann darauf geprüft werden, ob es misst, was es vorgibt zu messen, wenn die Forschenden klar beschreiben,

welches Konstrukt es messen soll (de Vet et al., 2011). So wie die Autorinnen dieser Bachelorarbeit die Aussage des RTA-Algorithmus verstehen, ist der jeweilige Sprungtest das Messinstrument, anhand dessen bestimmt wird, wann eine Person wieder in den Sport zurückkehren darf. Weiter wurde angenommen, dass die Sprungtests die Stabilität, Kraft und Koordination messen. Dadurch wird eine Aussage möglich, welche Belastung oder Bewegung ab wann wieder erlaubt sein wird. Anhand dieser Annahmen werden Untersuchungen der Validität erwartet, die auf folgender Hypothese basieren: Liegt der LSI-Wert des Front Hop bei Personen mit einer Verletzung der unteren Extremität unter 90%, ist die Gefahr einer Wiederverletzung bei dynamischen Belastungen ohne Dreh- und Stoppbewegungen signifikant höher, als wenn der LSI-Wert über 90% liegt. Hiermit wird deutlich, dass zur Bestimmung der Validität eine Hypothese notwendig ist und dass diese Hypothese so spezifisch wie möglich sein soll (de Vet et al., 2011). Dass es ganz unterschiedliche Hypothesen gibt, was ein Messinstrument vorgibt zu messen, ist erst im Verlaufe der Arbeit ersichtlich geworden. Die untenstehende Tabelle zeigt, welche Validitäts-Stufen in den Studien untersucht werden, was das Ziel oder die Hypothese ist und anhand welches Ansatzes die Studien die Validität prüfen.

Tabelle 23 Übersicht zu den Validitäts-Stufen und zum Studienansatz

Studie 2	Studie 3	Studie 4
Um welche Validitäts-Stufe handelt es sich?		
Kriteriumsvalidität	Augenscheinvalidität	Konstruktvalidität
Wie lautet der Ansatz, mit dem die Studie die Validität prüft?		
Das Ziel ist, verschiedene Sprungtests auf die Sensitivität, Spezifität und Accuracy zu untersuchen, damit besser zwischen VKB verletzten und gesunden Personen unterschieden werden kann.	Die Augenscheinvalidität wird anhand zweier Fragen geprüft, die von verschiedenen Fachpersonen beantwortet werden. Bei der Frage 1 geht es darum, wie wichtig die Tests eingestuft werden. Bei Frage 2 geht es darum, wie oft die Tests in der Praxis angewendet werden.	Die Überprüfung der Validität basiert auf zwei Hypothesen: 1.) Veränderungen des LSIs bei den Sprungtests sollten beim operierten Bein signifikant grösser sein, als beim Nichtoperierten. 2.) Veränderungen des LSIs sollten mit Veränderungen beim "LEFS" Fragebogen und beim Global Rating of Change Fragebogen korrelieren.

Wird der Fokus auf die Hypothesen zur Bestimmung der Validität gesetzt, fällt einem sofort auf, dass nur die Studie von Reid et al. (2007) (Studie 4) eine klare Hypothese aufstellt. Studie 2 erwähnt nicht, dass die Validität untersucht werden sollte.

In der Studie 3 ist die Rede davon, dass die Tests anhand einer Augenscheinvalidität überprüft werden. Allerdings beziehen sich die Fragen zur Überprüfung der Augenscheinvalidität nicht darauf, was die Tests eigentlich messen sollten, sondern auf die Wichtigkeit, die den Tests zugestanden wird und auf den Gebrauch der Tests. Aus Sicht der Autorinnen dieser Bachelorarbeit wird damit nicht die Validität geprüft.

5.2.1 Front Hop

Die Tabelle zeigt die wichtigsten Ergebnisse zur Validität des Front Hop Tests.

Tabelle 24 Ergebnisse zur Validität des Front Hops

Studie 2	Studie 3	Studie 4
Welche Kennwerte wurden berechnet? Welche Ergebnisse erhielten die Forschenden?		
<ul style="list-style-type: none"> • Sensitivität (VKB konservativ): 53% • Sensitivität (VKB operativ): 63% • Spezifität (gesund): 100% • Accuracy (VKB konservativ): 69% • Accuracy (VKB operativ): 74% 	<ul style="list-style-type: none"> • Aussage, diese Testaufgaben sind sehr wichtig: ca. 70% der Befragten • Aussage, Tests regelmässig zu nutzen: ca. 40% aller Befragten 	<p>1) Die Veränderung der absoluten Werte vom operierten Bein sind signifikant grösser, als die Veränderungen vom nichtoperierten Bein: paired t = 6.4, P<0.001</p> <p>2) Die Korrelation zwischen der Veränderung des LSI und des «LEFS»: r = 0.37 (lower limit 95% CI 0.11). Korrelation zwischen der Veränderung des LSI zum Global Rating of Change: r = 0.48 (lower limit 95% CI 0.24).</p>

Obwohl die Studie 3 den Front Hop nicht wirklich auf die Validität überprüft, zeigen die Ergebnisse, dass dieser Test als sehr wichtig eingeschätzt wird (ca. 70%). In der Praxis wird er von 40% der befragten Personen regelmässig verwendet und ist somit der Test, der, von den in dieser Arbeit untersuchten Tests am häufigsten in der Praxis eingesetzt wird.

Die Werte des Front Hops erzielen in der Studie 2 eine Spezifität von 100%. Das bedeutet, von den gesunden Probandinnen und Probanden erreichen alle einen Wert von über 90% des LSIs. Bei der Sensitivität wird bei den VKB-Patientinnen/Patienten ein Wert von 53-63% erreicht, was auf eine schlechte Sensitivität hindeutet. Ein Grund für diese tiefe Sensitivität könnte sein, dass sich die Patientinnen und Patienten teilweise bereits 44 Monate postoperativ oder posttraumatisch befinden und somit als «gesund» gelten könnten (siehe 4.2.2). Die Accuracy liegt bei 69-74%.

Die erste Hypothese aus der Studie 4 bestätigt sich, indem die Veränderungen vom operierten Bein signifikant höher sind, als die des nichtoperierten Beines. Allerdings bedeutet dies nur, dass die Korrelation nicht gleich Null ist (Meichtry, 2019). Wie stark die Korrelation ist, wird nicht geprüft. Dennoch schliessen Reid et al. (2007) aus dem Ergebnis, dass sich die Sprungleistung beim operierten Bein über die sechs Wochen verbessert. Was genau zum besseren Ergebnis des operierten Beines führt, die Wundheilung, die Physiotherapie oder etwas anderes, ist unklar. Die zweite Hypothese, dass Veränderungen des LSIs mit Veränderungen des LEFS und des Global Rating of Change einhergehen sollten, wird in den Augen der Autorinnen dieser Bachelorarbeit nicht bestätigt. So liegen die Werte mit $r = 0.37$ (lower limit 95% CI 0.11) für den LEFS und $r = 0.48$ (lower limit 95% CI 0.24) für den Global Rating of Change sehr tief. Reid et al. (2007) legen zwar einen Wert von $r > 0.5$ als gut und einen Wert von $r = 0.36 - 0.5$ als akzeptabel fest. Allerdings scheint diese Einstufung sehr tief zu liegen. So schreibt de Vet et al. (2011), dass es zwar schwer sei, einen akzeptablen Wert einer Korrelation festzulegen, da dieser immer von der Situation abhängt, dass aber grundsätzlich eine Korrelation von 0,7 akzeptiert wird. Zudem entsprechen die statistischen Verfahren nicht dem Datenniveau der erhobenen Variablen und der LEFS scheint nicht dasselbe Konstrukt wie der Front Hop zu messen (siehe Kapitel 4.4.2).

Zusammengefasst kann gesagt werden, dass es sich beim Front Hop um einen wichtigen Test handelt, der von den in dieser Arbeit untersuchten Tests auch am häufigsten in der Praxis verwendet wird. Auch scheint sich die Sprungleistung beim operierten Bein über die sechs Wochen verbessert zu haben. Die Werte deuten ausserdem auf eine exzellente Spezifität hin.

5.2.2 Side Hop

Die Tabelle zeigt die wichtigsten Ergebnisse zur Validität des Side Hops.

Tabelle 25 Ergebnisse zur Validität des Side Hops

Studie 2
Welche Kennwerte werden berechnet?
Welche Ergebnisse erhalten die Forschenden?
<ul style="list-style-type: none"> • Sensitivität (VKB konservativ): 77% • Sensitivität (VKB operativ): 69% • Spezifität (gesund): 87% • Accuracy (VKB konservativ): 80% • Accuracy (VKB operativ): 74%

Lediglich Studie 2 untersucht den Side Hop auf die Validität. Die Ergebnisse der Sensitivität sind mit 69-77% nicht überwältigend. Wird dazu bedacht, dass sich unter den Studienteilnehmenden solche befinden, die in der Wundheilung schon stark fortgeschritten sind, könnte der Wert noch höher liegen. Die Spezifität liegt bei 87%, die Accuracy bei 74-80% und werden von den Autorinnen dieser Arbeit als akzeptabel eingestuft.

5.2.3 Square Hop

Die Tabelle zeigt die wichtigsten Ergebnisse zur Validität des Square Hops.

Tabelle 26 Ergebnisse zur Validität des Square Hops

Studie 2	Studie 3
Welche Kennwerte werden berechnet?	
Welche Ergebnisse erhalten die Forschenden?	
<ul style="list-style-type: none"> • Sensitivität (VKB konservativ): 53% • Sensitivität (VKB operativ): 51% • Spezifität (gesund): 67% • Accuracy (VKB konservativ): 58% • Accuracy (VKB operativ): 56% 	<ul style="list-style-type: none"> • Aussage, diese Testaufgaben sind sehr wichtig: 50 - 60% der Befragten • Aussage, Tests regelmässig zu nutzen 20 – 30% aller Befragten

Die Ergebnisse des Square Hops liegen in der Studie 2 in allen Bereichen zwischen 51-67% und weisen somit eine sehr schlechte Sensitivität und Accuracy, sowie eine schlechte Spezifität auf. So erreichen nur etwas mehr als die Hälfte der gesunden Personen einen LSI von über 90%. Auch die Ergebnisse aus der Studie 3 deuten darauf hin, dass dieser Test den Weg in die Praxis noch nicht gefunden hat. So stuft nur die Hälfte der befragten

Personen den Test als sehr wichtig ein und lediglich 20-30% nutzen ihn regelmässig in der Praxis.

6 Schlussfolgerung

6.1 Beantwortung der Fragestellung

Anhand der vorliegenden Arbeit kommen die Autorinnen zum Schluss, dass das Gütekriterium der Reliabilität für die drei quantitativen Tests zumindest für eine junge, sportliche Population gegeben ist. Wie sie bei älteren, unsportlichen oder verletzten Personen ausfallen würde, ist unbekannt. Bezüglich der Validität, die von den Autorinnen anhand des RTA-Algorithmus erwartet wird, kann keine Aussage getroffen werden, da sie nicht in diesem Sinne untersucht wurde. Dies ist bedauerlich, da laut de With (2017) die Validität aus methodischer Sicht als wichtigstes Gütekriterium gilt.

6.2 Limitationen der Arbeit

Limitationen sind sicherlich bei der Studienwahl vorhanden. So konnten nicht alle zur Thematik vorhandenen Studien analysiert werden, da dies den Rahmen einer Bachelorarbeit sprengen würde. Da ausserdem zum Side Hop und zum Square Hop fast keine Studien existieren, mussten sich die Autorinnen dieser Arbeit mit den wenig vorhandenen zufriedengeben, was teils gewisse Abweichungen der Testdurchführungen mit sich zog. Die Übersetzung der Studien aus dem Englischen ins Deutsche und das Erarbeiten eines für die Autorinnen dieser Arbeit weitgehend unbekanntes Gebietes in der Statistik, stellen weitere Limitationen dar. Letztendlich kann die Aussage bezüglich Reliabilität und Validität nur auf tendenziell sportliche VKB-Patientinnen und Patienten, und vor allem auf eine gesunde sportliche Population angewendet werden.

6.3 Weiterführende Überlegungen

Laut dem OSINSTITUT ist der RTA-Algorithmus auf jegliche Population mit einer Verletzung der unteren Extremität anwendbar. Interessant wäre nun, Reliabilität und Validität der Tests für unterschiedliche Populationen zu bestimmen. Beispielsweise unterschiedliche Alters- und Fitnessstufen oder verschiedene Krankheitsbilder wie Hüftprothese, Meniskusläsion oder Supinationstrauma. Zusätzlich könnten Folgestudien die Validität im Sinne der folgenden Hypothese untersuchen: Liegt der LSI-Wert des Front Hops bei Personen mit einer Verletzung der unteren Extremität unter 90%, ist die Gefahr einer Wiederverletzung bei dynamischen Belastungen ohne Dreh- und Stoppbewegungen

signifikant höher, als wenn der LSI-Wert über 90% liegt. Allerdings müsste diese Hypothese noch spezifischer formuliert werden und es bräuchte noch zusätzliche Parameter, wie sie überprüft werden könnte. Einerseits sollte der Begriff «Verletzungen der unteren Extremität» genauer definiert werden. Denn dieser beinhaltet verschiedene Verletzungsmuster, was das Zusammenstellen einer repräsentativen Stichprobe erschwert. Weiter sollten Überlegungen zur Validitätsstufe stattfinden. Eine Kriteriumsvalidität kommt aus Sicht der Autorinnen dieser Bachelorarbeit nicht in Frage, da diese ethisch nicht vertretbar wäre.

Denn dabei müsste die Kontrollgruppe, trotz $LSI < 90\%$, ohne jegliche Intervention in den Sport zurückkehren. Nur dadurch könnte die Rerupturrate mit der Interventionsgruppe verglichen werden. Daher wäre am ehesten eine Konstruktvalidität angebracht, wie es in Studie 4 gemacht wurde. Allerdings sollte das Assessment, mit dem die Sprungtests verglichen werden, auch wirklich das gleiche Konstrukt messen. Die Schwierigkeit besteht darin, ein solches Assessment zu finden.

Tendenziell schloss der Square Hop bei den untersuchten Studien schlecht ab. Daraus resultiert die Frage, ob dieser überhaupt in eine Testbatterie aufgenommen werden sollte. In Studie 2 wird er deshalb ausgeschlossen, weil er zu stark mit dem Side Hop korreliert.

7 Praxistransfer

Die Autorinnen dieser Arbeit möchten gerne den Praxistransfer aus zwei verschiedenen Blickwinkeln beschreiben. Erstens den Transfer von der Theorie zur Praxis anhand eigener Erfahrungen und zweitens die Empfehlung für die Praxis anhand der vorliegenden Arbeit.

1. Die Autorinnen dieser Bachelorarbeit erhielten die Möglichkeit, an einem zweitägigen Seminar des OSINSTITUT in Zuzenhausen (Deutschland) teilzunehmen. Das Seminar 2a befasste sich intensiv mit dem RTA-Algorithmus. Die vertiefte praktische Auseinandersetzung mit diesem Algorithmus und die Meinung von Experten auf diesem Gebiet, wurden als grossen Mehrgewinn empfunden.
2. Für die Praxis können die drei untersuchten Tests insofern empfohlen werden, als dass sie alle eine gute Test-Retest-Reliabilität aufweisen. Für Verlaufsmessungen zu beachten sind die SDC-Werte (siehe 5.1.3 - 5.1.5). Erst wenn dieser Wert überschritten wurde, hat tatsächlich eine Veränderung stattgefunden. Bezüglich der

Validität liegen wenige Ergebnisse vor. Zudem untersucht keine der in dieser Arbeit verwendeten Studien, ob die Tests eine Wahrscheinlichkeit einer Reruptur voraussagen können. Daher sollte mit der Aussage der Tests bezüglich der Validität vorsichtig umgegangen werden.

Literaturverzeichnis

- Barber, SD., Noyes, FR., Mangine, RE., McCloskey JW., Hartman W. (1990). Quantitative assessment of functional limitations in normal and anterior cruciate ligament-deficient knees. Abgerufen am 09. April 2019 von https://www.researchgate.net/publication/233859877_Quantitative_Assessment_of_Functional_Limitations_in_Normal_and_Anterior_Cruciate_Ligament-Deficient_Knees
- Borgstedt, N. (2016, 8. April). Der Return to Activity Algorithmus (RTAA) - was ist das?. Abgerufen am 21. November 2018, von <https://www.netzathleten.de/fitness/richtig-trainieren/item/6096-der-return-to-activity-algorithmus-rtaa-was-ist-das>
- DocCheck Medical Services GmbH. (o.D.). Lachman-Test. Abgerufen am 4. Dezember, 2018, von <https://flexikon.doccheck.com/de/index.php?title=Lachman-Test>
- Gesundheitsdirektion des Kantons Zürich (2009). Ruptur des vorderen Kreuzbandes: operative oder konservative Behandlung? Abgerufen am 2. Juli 2018, von http://www.medical-board.ch/fileadmin/docs/public/mb/fachberichte/2009_06_30_ruptur_kreuzband_bericht.pdf
- Gustavsson A., Neeter C., Thomeé P., Silbernagel KG., Augustsson J., Thomeé R., & Karlsson J. (2006). A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. *Knee Surgery, Sports Traumatology, Arthroscopy*, 14(8), S. 778–788. <https://doi.org/10.1007/s00167-006-0045-6>
- Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. *Journal of Orthopaedic & Sports Physical Therapy*, 44(12), S. 947–954.
- Keller, M., & Kurz, E. (2016). Zurück zum Pre Injury Level nach Verletzungen der unteren Extremität – eine Einteilung funktioneller Assessments. *Manuelle Therapie*, 20, S. 16–18. Abgerufen von <https://www.osinstitut.de/publikationen>
- Keller, M., Schmidlein, O., & Kurz, E. (2012). Return to Activity Testing – Bedeutung funktioneller Tests für die Rehabilitation. Heruntergeladen am 2. Juli 2018, von <https://www.osinstitut.de/download/cms/22/56.pdf>
- Keller, M., Kurz, E., Schmidlein, O., Welsch, G., & Anders, C. (2016). Interdisziplinäre Beurteilungskriterien für die Rehabilitation nach Verletzungen an der unteren

- Extremität: Ein funktionsbasierter Return-to-Activity Algorithmus. *Sportverl Sportschad* 2016; 30: S. 38-46. <https://dx.doi.org/10.1055/s-0042-100966>
- Kockum, B., & Heijne, A. I. L. M. (2014). Hop performance and leg muscle power in athletes: Reliability of a test battery. *Physical Therapy in Sport*, S. 222–227.
- Kong, D. H., Yang, S. J., Ha, J. K., Jang S. H., Seo J. G., & Kim J. G. (2012). Validation of Functional Performance Test after Anterior Cruciate Ligament Reconstruction. *Knee Surgery & Related Research*, 24(1), S. 40-45.
<https://doi.org/10.5792/ksrr.2012.24.1.40>
- MatheGuru. (o.D.). Signifikanzniveau. Abgerufen am 14. März 2019, von <https://matheguru.com/stochastik/signifikanz-signifikanzniveau.html>
- MatheGuru. (o.D.). Standardabweichung. Abgerufen am 4. Dezember 2018, von <https://matheguru.com/stochastik/standardabweichung.html>
- Meichtry, A. (2017). *Statistik. Handbuch für Therapeuten* [elektronische Version]. Doi: 10.1055/b-005-143306
- Meichtry, A. (2018). Vorlesung über Reliabilität und Validität. Schulungsunterlagen Masterstudiengänge Departement Gesundheit ZHAW.
- Myklebust, G., & Bahr, R. (2005). Return to play guidelines after anterior cruciate ligament surgery. *British Journal of Sports Medicine*, 39(3), 127–131.
<https://doi.org/10.1136/bjism.2004.010900>
- Oesch, P., Hilfiker, R., Keller, S., Kool, J., Luomajoki, H., Schädler, S., Tal-Akabi, A., Verra, M., & Widmer Leu, C. (2011). Assessments in der Rehabilitation. Band 2: Bewegungsapparat, 2., *vollständig überarbeitete und erweiterte Auflage*, S. 25-35.
- Physiosupport. (o.D.) Wundheilung. Abgerufen am 04. Dezember 2018, von <http://www.physiosupport.org/wundheil.htm>
- Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. *American Physical Therapy Association*, 87(3), S. 337–349.
- Ris, I. & Preusse-Bleuler, B. (2015). AICA: Arbeitsinstrument für ein Critical Appraisal eines Forschungsartikels. Schulungsunterlagen Bachelorstudiengänge Departement Gesundheit ZHAW.
- StatistikGuru. (o.D.) Konfidenzintervall / StatistikGuru. Abgerufen am 12. November 2018, von <https://statistikguru.de/lexikon/konfidenzintervall.html>

- UZH – Methodenberatung – Zentrale Tendenz. (o.D.). Abgerufen am 21. November 2018, von https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral.html
- Van den Berg, F., & Cabri, J. (2011). *Angewandte Physiologie: Das Bindegewebe des Bewegungsapparates verstehen und beeinflussen* (3. Aufl.). Stuttgart, Deutschland: Theime, S. 68-72.
- Vet, H. C. W. de, Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*, S. 96-201.
- Wikipedia. (o.D.). Bland-Altman-Diagramm+ - Wikipedia. Abgerufen am 13. November 2018, von <https://de.wikipedia.org/wiki/Bland-Altman-Diagramm>
- Wikipedia. (o.D.). Drop-Out. Abgerufen am 14. März 2019, von [https://de.wikipedia.org/wiki/Drop-out_\(klinische_Studie\)](https://de.wikipedia.org/wiki/Drop-out_(klinische_Studie))
- Wikipedia. (o.D.). Intra-Klassen-Korrelation- Wikipedia. Abgerufen am 21. November 2018, von <https://de.wikipedia.org/wiki/Intra-Klassen-Korrelation>
- Wikipedia. (2018a, 24. September). Signifikanztest. Abgerufen am 21. November 2018, von <https://de.wikipedia.org/wiki/Post-hoc-Test>
- Wikipedia. (2018b, 24. November). Neutral-Null-Methode – Wikipedia. Abgerufen am 26. November 2018, von <https://de.wikipedia.org/wiki/Neutral-Null-Methode>
- Wirz, M., Köhler, B., Marks, D., Koll, J., Sattelmayer, M., Oesch, P., Hilfiker, R., Rogan, S., Schädler, S., Verra, M., & Lüthi, H. (2014). *Lehrbuch Assessments in der Rehabilitation* (1. Aufl.), S. 10.
- With, E. de. (2017). Basale Konzepte zu Quantitativen Verfahren. Schulungsunterlagen Bachelorstudiengänge Departement Gesundheit ZHAW.

Glossar

Begriff	Definition
Accuracy	Treffgenauigkeit = Anzahl Patienten und Patientinnen, die den Test nicht bestanden haben und Anzahl gesunde Personen, die den Test bestanden haben, geteilt durch alle gesunden Personen und alle Patienten und Patientinnen
Assessment	Verfahren, um Merkmale und Eigenschaften auf systematische Weise zu erfassen. (Wirz et al., 2014)
Bias	Statistik: Verzerrung/systematischer Fehler
Drop-Out	Probandinnen und Probanden, die ursprünglich für eine Studie rekrutiert wurden, diese aber nicht beenden (Wikipedia, o.D.)
Eingipflig	Statistik: beschreibt, dass eine Verteilung nur eine Wölbung hat, bzw. glockenförmig ist
Fibrin	Nicht wasserlösliches Protein, dient der Blutstillung
Global Rating of Change	Fragebogen, bei dem die Patientinnen und Patienten von einer Skala von 0-7 ankreuzen konnten, ob ihrer Meinung nach die funktionelle Leistung besser, gleich geblieben, oder schlechter geworden, ist
Goldstandard	Allgemeingültiges, bestes zu diesem Zeitpunkt verfügbares Instrument zur Messung einer Begebenheit
Hamstrings	M. biceps femoris, M. semitendinosus, M. semimembranosus (Muskeln am hinteren Oberschenkel)
Heterogenität	Die Individuen unterscheiden sich stark voneinander
Intervallskaliert	Skalenniveau aus der Statistik: Die Abstände der Werte sind klar definiert und beinhalten die gleichen Abstände, aber es ist kein Nullpunkt bestimmbar, z.B. Temperatur (de With, 2017)
Isokinetischer Krafttest	Spezieller Krafttest, Widerstand wird durch den Patienten/die Patientin selbst erzeugt
Items	Bestandteil, Element
Lachman-Test	Test zur Diagnosenstellung einer VKB-Ruptur
Leukozyten	Weisse Blutkörperchen, wehren Krankheitserreger ab
Lower limit	Statistik: Beschreibt den unteren Bereich des CI.
Lymphozyten	Gehören zu den Leukozyten, sind die eigentlichen Abwehrzellen
Matrix	Substanz zwischen Zellen und Gewebe (= Zwischenzellsubstanz)
Makrophagen	Entstehen aus Monozyten, wandern in Wundgebiet ein und «fressen» das nekrotische Gewebe
Median	Kennwert aus der Statistik: teilt eine Stichprobe hälftig. (de With, 2017)
Mittelwert	Kennwert aus der Statistik: Summe aller Werte, geteilt durch die Anzahl aller Werte (de With, 2017)

Modalwert	Kennwert aus der Statistik: die Ausprägung, die am häufigsten vorkommt (de With, 2017)
Modellannahmen	Beispielsweise die Überprüfung der Verteilung und des Datenniveaus einer Variablen
Monozyten	Abwehrzellen im Blut, können sich in Makrophagen umwandeln
Nekrotisch	Abgestorben
Neutral-Null-Methode	Standardisierter, orthopädischer Bewertungs- und Dokumentationsindex für die Beweglichkeit von Gelenken (Wikipedia, 2018b)
Normalverteilt	Verteilungsform aus der Statistik: symmetrische, eingipflige Verteilungsform numerischer Daten, z.B. Körpergrösse (de With, 2017)
Ordinalskaliert	Skalenniveau aus der Statistik: eindeutige Rangreihe der Werte, aber keine Informationen über die Abstände bestimmbar, z.B. «Vanilleeis schmeckt besser als Lebertran» (de With, 2017)
Parametrisch	Parametrische, statistische Testverfahren: Es ist im Gegensatz zu nonparametrischen Verfahren eine explizite Form der Verteilung der beteiligten Variablen erforderlich, z.B. Normalverteilung (de With, 2017)
Permeabilität	Durchlässigkeit
Pivot-Shift-Test	Test zur Überprüfung der Kniestabilität, v.a. VKB
Population	Statistik: Grundgesamtheit, über die eine Aussage gemacht werden kann
Posturale Kontrolle	Gewährleistet die Kontrolle der Körperhaltung im Raum zur Sicherung von Orientierung und Stabilität
Sample Size Calculation	Statistik: Berechnung der optimalen Stichprobengrösse
Signifikanzniveau	Statistik: «(...) Gibt an, wie hoch das Risiko ist, das man bereit ist einzugehen, eine falsche Entscheidung zu treffen.» (MatheGuru, o.D.)
Upper limit	Statistik: Beschreibt den oberen Bereich des CI
Vorderer Schubladentest	Test zur Diagnosestellung einer VKB-Ruptur

Abkürzungsverzeichnis

Abkürzung	Definition
b	Bias
ε	Fehler-Wert
i	Index für wiederholte Messungen
n	Anzahl aller Werte
η	Wahrer Wert
σ^2	Varianz
$\sigma^2(p)$	Personenvarianz
p	Beschreibt das Signifikanzniveau
r	Korrelation, in dieser Arbeit Pearson-Korrelations-Koeffizient
s	Mathematische Abkürzung für Standardabweichung
x	Messwert (in dieser Arbeit: Differenzen aller Werte vom Mittelwert)
Y	Beobachteter Wert
AICA	Arbeitsinstrument für ein Critical Appraisal
CI	Confidence Intervall = Konfidenzintervall
EMED	Einleitung, Methode, Ergebnis, Diskussion
FLEE	Funktional Lower Extremity Evalvau
ICC	Intraclass Correlation Coefficient = Intraklassen-Korrelations-Koeffizient
LEFS	Lower Extremity Functional Scale
LSI	Limb Symmetrie Index
MeSH	Medical Subject Headings
OSINSTITUT	Institut «Bewegung für Orthopädie und Sportmedizin» gegründet von O. Schmidlein und M. Keller
RTA-Algorithmus	Return to Activity Algorithmus
SD	Standard Deviation = Standardabweichung
SDC	Smallest Detectable Change = kleinste erkennbare Veränderung
SEM	Standard Error of Measurement = Standardfehler
SEM _D	Standard Error of Measurement Difference = Standardfehler für eine Veränderung
UEx	Untere Extremität
UVG	Unfallversicherungsgesetz
VKB	Vorderes Kreuzband
ZHAW	Zürcher Hochschule für Angewandte Wissenschaften

Abbildungsverzeichnis

Abbildung 1 Kontinuum der Wundheilung und Funktion (Keller et al., 2016)	9
Abbildung 2 MeSH Term «Reproducibility of Results» (PubMed, 2018)	25

Tabellenverzeichnis

Tabelle 1 Aufbau des RTA-Algorithmus (angepasst nach Keller, M. & Kurz E. (2016))	12
Tabelle 2 Einteilung der Aktivitätslevel (angepasst nach Keller, M., Schmidlein, O., Kurz, E., Welsch, G. & Anders C. (2016))	12
Tabelle 3 Dauer der Wundheilung (angepasst nach Diemer F. & Sutor V. (2018))	15
Tabelle 4 Dichotomer Test und dichotomes Kriterium (angepasst nach Meichtry, A. (2018))	20
Tabelle 5 Vierfeldertafel als Beispiel unserer Fragestellung.....	21
Tabelle 6 Keywords und MeSH Terms.....	26
Tabelle 7 Ein- und Ausschlusskriterien	27
Tabelle 8 Einteilung der Studien	27
Tabelle 9 Übersicht der in dieser Arbeit analysierten Studien	28
Tabelle 10 Übersicht Stichprobe der Studie 1	30
Tabelle 11 Statistische Verfahren der Studie 1	30
Tabelle 12 Übersicht Stichprobe der Studie 2	32
Tabelle 13 Statistische Verfahren der Studie 2	33
Tabelle 14 Ergebnisse für Sensitivität, Spezifität und Accuracy der Studie 2 (angepasst nach Gustavsson et al. (2006))	34
Tabelle 15 Übersicht Stichprobe der Studie 3	36
Tabelle 16 Berechnete Werte der Studie 3	37
Tabelle 17 Übersicht Stichprobe der Studie 4	39
Tabelle 18 Statistische Verfahren der Studie 4	40
Tabelle 19 Übersicht potenzielle Ursachen für Bias	43
Tabelle 20 Ergebnisse zur Reliabilität des Front Hops.....	45
Tabelle 21 Ergebnisse zur Reliabilität des Side Hops	47
Tabelle 22 Ergebnisse zur Reliabilität des Square Hops.....	48
Tabelle 23 Übersicht zu den Validitäts-Stufen und zum Studienansatz	49
Tabelle 24 Ergebnisse zur Validität des Front Hops	50

Tabelle 25 Ergebnisse zur Validität des Side Hops.....	52
Tabelle 26 Ergebnisse zur Validität des Square Hops	52

Weitere Quellen

Meichtry André

Deklaration der Wortanzahl

Diese Bachelorarbeit umfasst ohne Abstract, Abbildungen, Tabellen und deren Legenden, Inhalts- und Literaturverzeichnis, Eigenständigkeitserklärung, Danksagung und Anhänge 11960 Wörter.

Der deutsche Abstract enthält 200 Wörter, der Englische 199.

Danksagung

Wir bedanken uns bei Herrn André Meichtry für die Unterstützung beim Erklären schwieriger Begriffe, für die Möglichkeit des Besuches einer Vorlesung über Reliabilität und Validität, sowie für das Begleiten der Bachelorarbeit.

Dem OSINSTITUT, im Speziellen Matthias Keller und Oliver Schmidlein, gebührt Dank für die Ermöglichung des Seminarbesuchs über den RTA-Algorithmus, sowie für die Offenheit und investierte Zeit zur Beantwortung spezifischer Fragen.

Zudem möchten wir uns von ganzem Herzen bei unseren Eltern und unseren Freundinnen und Freunden für das Korrekturlesen der Arbeit, die psychische Unterstützung und wertvollen Anregungen bedanken.

Eigenständigkeitserklärung

«Wir erklären hiermit, dass wir die vorliegende Arbeit selbstständig, ohne Mithilfe Dritter und unter Benutzung der angegebenen Quellen verfasst haben.»

Winterthur / 26.04.2019




Ort / Datum

Ursina Liechti

Ariane Gfeller

Anhang

Anhang 1: Neun relevante Studien nach der ersten Auswahl

1. Munro, A. G., & Herrington, L. C. (2011). Between-session Reliability of four Hop Tests and the Agility T-Test. *Journal of Strength and Conditioning Research*, 25(5), S. 1470–1477.
2. Logerstedt, D., Grindem, H., Lynch, A., Eitzen, I., Engebretsen, L., Risberg, M. A., Snyder-Mackler, L. (2012). Single-Legged Hop Tests as Predictors of Self-Reported Knee Function After Anterior Cruciate Ligament Reconstruction. *The American Journal of Sports Medicine*, 40(10), S. 2348–2356. <https://doi.org/10.1177/0363546512457551>
3. Kockum, B., & Heijne, A. I. L. M. (2014). Hop performance and leg muscle power in athletes: Reliability of a test battery. *Physical Therapy in Sport*, S. 222–227.
4. Gustavsson, A., Neeter, C., Thomeé, P., Grävare Silbernagel, K., Augustsson, J., Thomeé, R., & Karlsson, J. (2006). A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. *Knee Surgery, Sports Traumatology, Arthroscopy*, 14(8), S. 778–788. <https://doi.org/10.1007/s00167-006-0045-6>
5. Järvelä, T., Kannus, P., Latvala, K., & Järvinen, M. (2002). Simple Measurement in Assessing Muscle Performance After an ACL Reconstruction. *Orthopedics and Clinical Science*, 23, S. 196–201.
6. Kamonseki, D. H., Cedin, L., Tavares-Preto, J., & Calixtre, L. B. (2018). Reliability, validity, and minimal detectable change of Side Hop Test in male children and adolescents. *Physical Therapy in Sport*, 34, S. 141–147. <https://doi.org/10.1016/j.ptsp.2018.09.009>
7. Walbright, P. D., Walbright, N., Ojha, H., & Davenport, T. (2017). Validity of functional screening Tests to predict lost-time lower quarter Injury in a cohort of female collegiate Athletes. *The International Journal of Sports Physical Therapy*, 12(6), S. 948–959.
8. Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. *American Physical Therapy Association*, 87(3), S. 337–349.
9. Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. *Journal of Orthopaedic & Sports Physical Therapy*, 44(12), S. 947–954.

Anhang 2: Suchmatrix

Farbig sind jeweils die Studien, die schon 1x aufgelistet wurden.

Stichwörter Schlagwörter (Englisch)	Daten- bank	Anzahl Treffer	Anzahl relevante Abstracts	Relevante Literatur
«Reproducibility of Results» [Mesh] AND «hop test»	PubMed	20	6	<p>Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. <i>Journal of Orthopaedic & Sports Physical Therapy</i>, 44(12), S. 947–954.</p> <p>Thelen, M. D., Koppenhaver, S. L., Hoppes, C. W., Shutt, C., Musen, J. L., & Williams, M. K. (2015). Reliability of a Novel Return to Duty Screening Tool for Military Clinicians. <i>U.S. Army Medical Department Journal</i>, S. 14–23.</p> <p>Ortiz, A., Olson, S. L., Roddey, T. S., & Morales, J. (2005). Reliability of Selected Physical Performance Tests in Young Adult Women. <i>The Journal of Strength and Conditioning Research</i>, 19(1), S. 39-44. https://doi.org/10.1519/14163.1</p> <p>Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. <i>American Physical Therapy Association</i>, 87(3), S. 337–349.</p> <p>Mani, K., Brechue, W. F., Friesenbichler, B., & Maffiuletti, N. A. (2017). Validity and reliability of a novel instrumented one-legged hop test in patients with knee injuries. <i>The Knee</i>, 24(2), S. 237–242. https://doi.org/10.1016/j.knee.2016.09.004</p> <p>Munro, A. G., & Herrington, L. C. (2011). Between-session Reliability of four Hop Tests and the Agility T-Test. <i>Journal of Strength and Conditioning Research</i>, 25(5), S. 1470–1477.</p>
«Reproducibility of Results» [Mesh] AND «Lower Extremity» [Mesh] AND «hop test»	PubMed	5	3	<p>Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. <i>Journal of Orthopaedic & Sports Physical Therapy</i>, 44(12), S. 947–954.</p> <p>Thelen, M. D., Koppenhaver, S. L., Hoppes, C. W., Shutt, C., Musen, J. L., & Williams, M. K. (2015). Reliability of a Novel Return to Duty Screening Tool for Military Clinicians. <i>U.S. Army Medical Department Journal</i>, S. 14–23.</p> <p>Ortiz, A., Olson, S. L., Roddey, T. S., & Morales, J. (2005). Reliability of Selected Physical Performance Tests in Young Adult Women. <i>The Journal of Strength and Conditioning</i></p>

				<i>Research</i> , 19(1), S. 39-44. https://doi.org/10.1519/14163.1
«single leg hop» AND validity	PubMed	9	0	
Validity OR reliability AND «square hop test»	PubMed	2	0	
«balance front hop» OR «single leg hop» AND validity AND reliability	CINHAL	3	0	
((MH «Validity») OR (MH «Reliability»)) AND «hop test»	CINHAL	3	0	
(MH «Reliability and Validity») AND «hop test»	CINAHL	12	2	Walbright, P. D., Walbright, N., Ojha, H., & Davenport, T. (2017). Validity of functional screening Tests to predict lost-time lower quarter Injury in a cohort of female collegiate Athletes. <i>The International Journal of Sports Physical Therapy</i> , 12(6), S. 948–959. Ortiz, A., Olson, S. L., Roddey, T. S., & Morales, J. (2005). Reliability of Selected Physical Performance Tests in Young Adult Women. <i>The Journal of Strength and Conditioning Research</i> , 19(1), S. 39-44. https://doi.org/10.1519/14163.1
Balance side hop AND validity AND reliability	CINHAL	17	3	Kockum, B., & Heijne, A. I. L. M. (2014). Hop performance and leg muscle power in athletes: Reliability of a test battery. <i>Physical Therapy in Sport</i> , S. 222–227. Ortiz, A., Olson, S. L., Roddey, T. S., & Morales, J. (2005). Reliability of Selected Physical Performance Tests in Young Adult Women. <i>The Journal of Strength and Conditioning Research</i> , 19(1), S. 39-44. https://doi.org/10.1519/14163.1 Walbright, P. D., Walbright, N., Ojha, H., & Davenport, T. (2017). Validity of functional screening Tests to predict lost-time lower quarter Injury in a cohort of female collegiate Athletes. <i>The International Journal of Sports Physical Therapy</i> , 12(6), S. 948–959.
Validity OR reliability AND «square hop test»	CINHAL	1	0	
((«validit*» or «accuracy» or «reliabilit*» or «reproductibility» or «reability of results» or «repeatability» or «precision» or «concordance» or	MEDLINE	36	10	Kamonseki, D. H., Cedin, L., Tavares-Preto, J., & Calixtre, L. B. (2018). Reliability, validity, and minimal detectable change of Side Hop Test in male children and adolescents. <i>Physical Therapy in Sport</i> , 34, S. 141–147. https://doi.org/10.1016/j.ptsp.2018.09.009 Walbright, P. D., Walbright, N., Ojha, H., & Davenport, T. (2017). Validity of functional screening Tests to predict lost-time lower quarter Injury in a cohort of female collegiate Athletes. <i>The International Journal of Sports Physical Therapy</i> , 12(6), S. 948–959. Mani, K., Brechue, W. F., Friesenbichler, B., & Maffiuletti, N. A. (2017). Validity and

<p>«dependability») and «hop test»)</p>				<p>reliability of a novel instrumented one-legged hop test in patients with knee injuries. <i>The Knee</i>, 24(2), S. 237–242. https://doi.org/10.1016/j.knee.2016.09.004</p> <p>Thelen, M. D., Koppenhaver, S. L., Hoppes, C. W., Shutt, C., Musen, J. L., & Williams, M. K. (2015). Reliability of a Novel Return to Duty Screening Tool for Military Clinicians. <i>U.S. Army Medical Department Journal</i>, S. 14–23.</p> <p>Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. <i>Journal of Orthopaedic & Sports Physical Therapy</i>, 44(12), S. 947–954.</p> <p>Munro, A. G., & Herrington, L. C. (2011). Between-session Reliability of four Hop Tests and the Agility T-Test. <i>Journal of Strength and Conditioning Research</i>, 25(5), S. 1470–1477.</p> <p>Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. <i>American Physical Therapy Association</i>, 87(3), S. 337–349.</p> <p>Ortiz, A., Olson, S. L., Roddey, T. S., & Morales, J. (2005). Reliability of Selected Physical Performance Tests in Young Adult Women. <i>The Journal of Strength and Conditioning Research</i>, 19(1), S. 39-44. https://doi.org/10.1519/14163.1</p> <p>Kong, D. H., Yang, S. J., Ha, J. K., Jang, S. H., Seo, J. G., & Kim, J. G. (2012). Validation of Functional Performance Tests after Anterior Cruciate Ligament Reconstruction. <i>Knee Surgery & Related Research</i>, 24(1), S. 40–45. https://doi.org/10.5792/ksrr.2012.24.1.40</p> <p>Logerstedt, D., Grindem, H., Lynch, A., Eitzen, I., Engebretsen, L., Risberg, M. A., Snyder-Mackler, L. (2012b). Single-Legged Hop Tests as Predictors of Self-Reported Knee Function After Anterior Cruciate Ligament Reconstruction. <i>The American Journal of Sports Medicine</i>, 40(10), S. 2348–2356. https://doi.org/10.1177/0363546512457551</p>
<p>(validity or reliability) and «side hop test»)</p>	<p>MEDLINE</p>	<p>3</p>	<p>2</p>	<p>Ortiz, A., Olson, S. L., Roddey, T. S., & Morales, J. (2005). Reliability of Selected Physical Performance Tests in Young Adult Women. <i>The Journal of Strength and Conditioning Research</i>, 19(1), S. 39-44. https://doi.org/10.1519/14163.1</p> <p>Kamonseki, D. H., Cedin, L., Tavares-Preto, J., & Calixtre, L. B. (2018). Reliability, validity, and minimal detectable change of Side Hop Test in male children and adolescents. <i>Physical Therapy in Sport</i>, 34, S. 141–147. https://doi.org/10.1016/j.ptsp.2018.09.009</p>
<p>Validity OR reliability AND square hop test</p>	<p>MEDLINE</p>	<p>0</p>	<p>0</p>	

Hop Test AND validity	AMED	4	2	<p>Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. <i>American Physical Therapy Association</i>, 87(3), S. 337–349.</p> <p>Järvelä, T., Kannus, P., Latvala, K., & Järvinen, M. (2002). Simple Measurement in Assessing Muscle Performance After an ACL Reconstruction. <i>Orthopedics and Clinical Science</i>, 23, S. 196–201.</p>
Validity OR reliability AND «square hop test»	AMED	1	0	
Über Referenz vom OSINSTITUT gefunden				<p>Gustavsson, A., Neeter, C., Thomeé, P., Grävare Silbernagel, K., Augustsson, J., Thomeé, R., & Karlsson, J. (2006). A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. <i>Knee Surgery, Sports Traumatology, Arthroscopy</i>, 14(8), S. 778–788. https://doi.org/10.1007/s00167-006-0045-6</p>

Anhang 3: AICA-Formulare

In den folgenden vier AICA-Rastern, einem von Ris und Preusse-Bleuler (2015) entwickelten Arbeitsinstrument zur systematischen Zusammenfassung und Würdigung quantitativer Studien, würde am Ende jeweils eine Einschätzung der Gütekriterien Objektivität, Reliabilität und Validität in Bezug zur Studie folgen. Da dies das Hauptthema dieser Bachelorarbeit, sowie zentrales Thema des gesamten Diskussionsteils ist, wird an dieser Stelle darauf verzichtet.

In den Zusammenfassungen und Würdigungen sowie in der ganzen Bachelorarbeit wird für die Begriffe Single-Leg Hop for distance, One-Leg Hop for distance und Single Hop for distance jeweils das Synonym «Front Hop (Test)» verwendet.

Studie 1: Kockum, B., & Heijne, A. I. L. M. (2014). Hop performance and leg muscle power in athletes: Reliability of a test battery. Physical Therapy in Sport, S. 222–227

Zusammenfassung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion															
<p>Hintergrund: Die Autorinnen beschreiben, dass Verletzungen (speziell auch Rerupturen) der unteren Extremität im Sport, vor allem auch im Spitzensport häufig vorkommen. Sie gehen davon aus, dass es bereits gute Tests gibt, um zwischen verletzten und nicht verletzten Personen reliabel unterscheiden zu können. Die Tests müssen jedoch auch im Setting von Sportlern, Sportlerinnen, Spitzensportlern und Spitzensportlerinnen genügend sensitiv sein, um feinste Unterschiede erkennen zu können. Anhand solcher Tests könnte dann beurteilt werden, ob ein Athlet/eine Athletin für ein höheres Trainingslevel bereit ist oder nicht. Der Goldstandard zur Bestimmung der Muskelkraft war bis anhin ein isokinetischer Muskelfunktionstest. Es sei jedoch so, dass dieser keinen Zusammenhang mit der funktionellen Performance habe. Deswegen wurde nach anderen geeigneten Tests, zur Bestimmung der Muskelkraft und der funktionellen Performance, gesucht.</p> <p>Ziel: Das Ziel der Studie ist, die absolute und die relative Reliabilität und die kleinste</p>	<p>Design: Methodische Studie, Test-Retest Design mit jeweils 7-10 Tagen zwischen den Messungen (zwei Testgelegenheiten)</p> <p>Stichprobe: 18 gesunde Sporttreibende (9 Frauen, 9 Männer) mit einem Durchschnittsalter von 23.4 Jahren (SD: 2.4), einem Durchschnittsgewicht von 72.1kg (SD: 15.3) und einer Durchschnittsgrösse von 172cm (SD: 11.0) wurden angefragt, um an der Studie teilzunehmen. Einschlusskriterien: <ul style="list-style-type: none"> - Aktive Sporttreibende, die seit mindestens drei Monaten 6-8 Trainingseinheiten pro Woche absolvieren - Keine Schmerzen (in den letzten drei Monaten) - Keine reduzierte Leistung (in den letzten drei Monaten) Zwei Männer und zwei Frauen konnten nicht an der Studie teilnehmen wegen einer Nerveneinklemmung, Krankheit und wegen Verletzungen. Also nahmen 14 Personen (7 Frauen, 7 Männer) an der Studie Teil. Die Personen befanden sich laut eigener Aussage auf verschiedenen Levels bezüglich ihrer Vertrautheit mit Langhantelkniebeugen und Krafttraining. So waren 4 Personen Neulinge, wohingegen 10 Personen schon vertraut waren mit solchen Trainings. Die Probanden waren alle Studierende der gleichen Sportschule. Es wird nichts darüber geschrieben, wie die Probanden erreicht wurden. Somit wird von einem Non-Probability Sampling ausgegangen. Es wird nicht beschrieben, warum diese Personen ausgewählt wurden. Es ist davon auszugehen, dass sie ausgewählt wurden, weil die Autorinnen eine Aussage über eine sportliche Zielgruppe treffen wollten. Warum aber genau die Studierenden dieser Sportschule ausgewählt wurden, ist nicht ersichtlich. Die Autoren schreiben, dass keine Power-Calculation (vermutlich Sample Size Calculation) durchgeführt wurde. Die Anzahl der Studienteilnehmenden richtet sich nach einer Beschreibung von Bruton, Conway, & Holgate (2000), die besagt, dass die Stichprobengrösse aus der Anzahl der Probandinnen und Probanden mal der Anzahl der Testdurchführungen ermittelt wird. Diese Zahl sollte über 25 sein. Mit 14 Probanden x 2 Testdurchführungen kommt man somit auf 28.</p> <p>Datenerhebung: Die Testbatterie wurde den Probandinnen und Probanden vor dem Test eingehend erklärt. Zudem führten sie ein Aufwärmprogramm durch. Kurz vor dem Vertical</p>	<p>Ergebnisse: Vergleicht man die absoluten Werte der verschiedenen Tests miteinander, werden keine signifikanten Differenzen festgestellt. Einzig beim Side-Hop Test beim linken Bein ($p=0.01$).</p> <p>ICC Der ICC(2.1) variiert zwischen 0.84 – 0.98 bei allen Tests für das linke und rechte Bein. In der Tabelle sind die einzelnen Werte, die für unsere Arbeit wichtig sind, ersichtlich.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Test</th> <th>ICC(2.1)</th> <th>95%CI (lower-upper)</th> </tr> </thead> <tbody> <tr> <td>Front Hop (Re)</td> <td>0.97</td> <td>0.9-0.99</td> </tr> <tr> <td>Front Hop (Li)</td> <td>0.97</td> <td>0.89-0.99</td> </tr> <tr> <td>Side Hop (Re)</td> <td>0.84</td> <td>0.58-0.94</td> </tr> <tr> <td>Side Hop (Li)</td> <td>0.96</td> <td>0.79-0.99</td> </tr> </tbody> </table> <p>SEM und SEM% Der SEM% variiert zwischen 3.4 – 11.1% bei den vier Sprungtests und zwischen 8.1 – 12.4% bei den Krafttests. In der Tabelle sind die einzelnen Werte, die für unsere Arbeit wichtig sind, ersichtlich.</p>	Test	ICC(2.1)	95%CI (lower-upper)	Front Hop (Re)	0.97	0.9-0.99	Front Hop (Li)	0.97	0.89-0.99	Side Hop (Re)	0.84	0.58-0.94	Side Hop (Li)	0.96	0.79-0.99	<p>Diskussion und Interpretation der Ergebnisse: Die Autoren beschreiben, dass dies die erste Studie sei, die die absolute Reliabilität und den SRD für die Knie-Flexion, Knie-Extension und den Single Leg Squat-Jump Test berechne. Die absolute Reliabilität für diese Tests, berechnet anhand des ICCs, sei nach Vincent (1994) als hoch einzustufen. Auch der ICC für den Vertical Jump, den Front Hop und den Side Hop Test weisen gute, bis absolute Übereinstimmungen auf. In der Diskussion erläutern die Autorinnen, dass ein hoher ICC auch von der Heterogenität der Stichprobe abhängen kann. Aus diesem Grund haben sie den SEM berechnet, der nicht von der Heterogenität der Personen abhängt. Die Standardmessfehler (SEM) dieser Studie sind für den Vertical Jump, den Front Hop und den Side Hop Test sehr ähnlich wie die Standardmessfehler bereits existierender Studien von Neeter et al. (2006). Es wird diskutiert, dass es in der Literatur keine Klarheit bezüglich eines akzeptablen SEM gebe und dass deswegen immer überlegt werden müsse, in welchem Kontext die Messung verwendet wird und ob der SEM dort klinisch akzeptabel ist. Es wird diskutiert, ob es sich bei 14 Versuchspersonen um eine zu kleine Testgruppe handelt und dass dies eine mögliche Limitation der Studie darstellt. Allerdings beziehen sich die Autorinnen auf eine Empfehlung von Bruton et al. (2000).</p>
Test	ICC(2.1)	95%CI (lower-upper)																
Front Hop (Re)	0.97	0.9-0.99																
Front Hop (Li)	0.97	0.89-0.99																
Side Hop (Re)	0.84	0.58-0.94																
Side Hop (Li)	0.96	0.79-0.99																

wirkliche Differenz (SRD) bei drei häufig gebrauchten Sprungtests, zwei Krafttests und beim Single Leg Squat Jump herauszufinden.

Die Testbatterie besteht aus den folgenden Tests:

Vertical Jump, Front Hop, Side Hop, Single Leg Squat Jump, Knie-Flexions- und Knie-Extensions-Kraft

Jump und dem Front Hop durften die Testpersonen 3-5 Versuche auf jedem Bein durchführen, gefolgt von 1 Minute Pause vor dem zählenden Test. Vor dem Side-Hop Test durften die Testpersonen 10-15 Sprünge machen, um mit dem Test vertraut zu werden. Auch für die anderen Tests durften die Probandinnen und Probanden eine Übungseinheit durchführen, die aber an dieser Stelle nicht genau beschrieben wird, da diese Tests nicht Teil der Bachelorarbeit sind.

Art der erhobenen Daten (für unsere BA relevant):

- Front Hop: cm
Es wurden jeweils pro Testgelegenheit drei korrekte Sprünge pro Bein aufgezeichnet.
- Side-Hop: Anzahl korrekter Sprünge innerhalb 30 Sekunden

Die Daten wurden zweimal mit einer Pause von 7-10 Tagen dazwischen erhoben.

Messverfahren & oder Intervention:

Für den Vertical Jump und die Muskelkrafttests wurden spezielle Messinstrumente verwendet. Da diese für die zwei in dieser BA relevanten Tests nicht eingesetzt wurden, wird an dieser Stelle nicht weiter darauf eingegangen.

Datenanalyse:

Datenniveau der erhobenen Variablen aus dem Front Hop und dem Side Hop Test:

- Cm = Proportionalskaliert
- Anzahl korrekter Sprünge in 30 Sekunden = Proportionalskaliert

Für die Daten der Teilnehmenden wurden Mittelwert und Standardabweichung (SD) berechnet.

Um den Zusammenhang zwischen den zwei Tests aufzuzeigen wurde ein gepaarter t-Test verwendet. Um systematische Abweichungen, Ausreisser und Streuungen zu überprüfen, wurden die Daten anhand eines Bland und Altman Diagramms dargestellt.

Für die relative Reliabilität wurde der ICC(2.1) berechnet. Die absolute Reliabilität wurde anhand des Standardmessfehlers (SEM) und dem SEM% bestimmt. Zudem wurde die Smallest Real Difference (SRD) und die SRD% berechnet.

Es wurde ein Signifikanzniveau von $p < 0.05$ festgelegt. Die Übereinstimmungsgrenze von 95% für alle Testvariablen wurde als mittlere Differenz zwischen Studienversuchen ± 1.96 SD berechnet.

Ethik:

Von den Forscherinnen werden keine ethischen Fragen thematisiert. Sie weisen aber darauf hin, dass die Studie vom Lokalkomitee, bestehend aus den Forschenden des Karolinska Instituts in Schweden (Stockholm) genehmigt wurde und dass ethische Überlegungen in der Studie anhand der Empfehlungen von Harris und Atkinson (2011) durchgeführt wurden.

Test	SEM	SEM%
Front Hop (Re)	5.0cm	3.5%
Front Hop (Li)	5.3cm	3.6%
Side Hop (Re)	5.4 Wdh.	11.1%
Side Hop (Li)	2.5 Wdh.	5.5%

SRD und SRD%

Anhand der Beschreibung der Autorinnen handelt es sich beim SRD um die, in unserer BA beschriebene Smallest Detectable Change (SDC). Die SRD% variierte zwischen 9.3 -30.7% bei den Sprungtests und zwischen 24.1– 34.3% bei den Krafttests. In der Tabelle sind die einzelnen Werte, die für unsere Arbeit wichtig sind, ersichtlich.

Test	SRD	SRD%
Front Hop (Re)	13.9cm	9.7%
Front Hop (Li)	14.6cm	10.1%
Side Hop (Re)	14.8 Wdh.	30.7%
Side Hop (Li)	7.0 Wdh.	15.2%

Die Ergebnisse werden in Tabellenform und als verständlichen Fliesstext präsentiert.

Das Forschungsziel kann bestätigt werden. So liegen nun Daten sowohl für die absolute und relative Reliabilität als auch für die SRD vor.

Schlussfolgerung, Anwendung und Verwertung in der Praxis:

Die absolute Reliabilität der Tests zeigt gute bis exzellente ICC-Werte und einen Standardmessfehler von ca. 10%.
-Ob diese Testbatterie bei verletzten Personen durchführbar ist, ist unbekannt. Jedoch kann diese Testbatterie empfohlen werden, um herauszufinden, wann Sportlerinnen und Sportler nach einer Verletzung wieder in den Sport zurückkehren kann, da einige dieser Tests sehr sportspezifisch sind. Zudem könne der sportspezifischste aller Tests, nämlich der Single-Leg Squat Jump, eingesetzt werden, um Kraft und Sprungqualitäten darzustellen. Beim Fazit im Abstract lauten die Aussagen der Autorinnen wie folgt: Diese Instrumente können zur Bestimmung der Funktion in Bezug auf die Kraft bei gesunden Sporttreibenden oder zu einem **späten** Rehabilitationsprozess empfohlen werden. Dabei sollten die methodischen Fehler der Tests berücksichtigt werden und das Standardisieren der Tests sollte mit Vorsicht getätigt werden.

Würdigung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Die Studie behandelt mit ihrer Zielsetzung eine für unsere BA wichtige Fragestellung. Allerdings sind von den sechs Tests nur zwei für uns relevant.</p> <p>Die Autorinnen definieren das Ziel der Studie klar und eindeutig (siehe oben). Eine konkrete Fragestellung oder Hypothese fehlt jedoch.</p> <p>Das Problem wird aus verschiedenen Perspektiven hergeleitet.</p>	<p>Design:</p> <p>Grundsätzlich ist das gewählte Design logisch und nachvollziehbar. Da die Forscherinnen die Test-Retest-Reliabilität der sechs Tests ermitteln wollten, sind zwei Testungen durchaus logisch. Allerdings stellt sich die Frage, ob nicht drei Testungen sinnvoller gewesen wären. Wenn man der Studie von Reid et al. (2007) vertraut, findet in der ersten Testung ein intensives motorisches Lernen statt, wodurch eigentlich erst die zweite und dritte Testgelegenheit zur Ermittlung der Reliabilität eingesetzt werden sollte.</p> <p>Die Forschenden versuchten die Gefahren der internen Validität durch folgende Schritte zu kontrollieren:</p> <ul style="list-style-type: none"> - Die Probandinnen und Probanden durften während 48 Stunden vor der Testung keinen intensiven Kraftsport und keine Sprünge machen (ansonsten sollten sie ihr normales Training und ihre Wettkämpfe absolvieren). - Die zwei Testungen wurden im gleichen Labor und unter gleicher Raumtemperatur durchgeführt. - Die Testungen wurden am Tag durchgeführt (dabei konnte die Zeit zwischen Morgen und Nachmittag variieren). - Die Testungen wurden von der gleichen Person durchgeführt. - Es wurden bei allen Testungen dieselben Erklärungen und Ermutigungen gegeben. <p>Es wird nichts darüber geschrieben, ob die Versuchspersonen über ihre Testergebnisse informiert wurden. Wenn dies der Fall war, dann wäre dies ein kritischer Punkt an der Studie, da dadurch die Ergebnisse deutlich beeinflusst werden könnten.</p> <p>Stichprobe:</p> <p>Wie genau die Stichprobenziehung durchgeführt wurde ist nicht ersichtlich. Es ist nicht eindeutig klar, auf welche Zielpopulation eine Aussage zielt. Die Testpersonen sind alle um die 23 Jahre alt und an einer Sportschule, wodurch sich die Aussage auf eine eher junge sportliche Population übertragen lässt. Die Autorinnen behaupten am Ende der Studie, dass diese Testbatterie für Athletinne und Athleten empfohlen werden kann, um herauszufinden, wann sie nach einer Verletzung wieder in den Sport zurückkehren können.</p> <p>In der Einleitung, der Diskussion und der Schlussfolgerung werden Personen mit einer VKB-Verletzung und Personen in einer späten Rehabilitation ins Spiel gebracht. Für diese Population ist die Stichprobe allerdings nicht repräsentativ. Die Stichprobengrösse scheint etwas klein zu sein. Da keine Sample Size Calculation durchgeführt wurde, ist nicht klar, was die optimale Stichprobengrösse gewesen wäre. Die Autorinnen thematisieren dies ebenfalls und begründen ihre Stichprobengrösse (siehe Methodenteil in der Zusammenfassung).</p>	<p>Die Ergebnisse werden präzise dargestellt. Zudem werden Tabellen als Ergänzung zum Text eingesetzt. Die Tabellen sind präzise und vollständig.</p>	<p>Diskussion und Interpretation der Ergebnisse:</p> <p>Es werden die meisten, allerdings nicht alle Resultate diskutiert. Speziell für diese BA wäre die Interpretation der absoluten Werte beim Side Hop Test interessant gewesen. Diese sind nämlich signifikant anders, wodurch sich die Frage stellt, ob dieser Test nun doch reliabel ist aufgrund des guten ICCs oder eben nicht. Ansonsten stimmen die Interpretationen mit den Resultaten überein.</p> <p>Die Autorinnen vergleichen ihre Ergebnisse mit jenen anderer Studien und setzen sie in Bezug zu ihrer Zielsetzung.</p> <p>Schlussfolgerung Anwendung und Verwertung in der Praxis:</p> <p>In Bezug auf die Zielsetzung, wurde die Studie sinnvoll durchgeführt. Allerdings kann sie nur die Frage nach der absoluten und relativen Reliabilität beantworten. Eine Stärke dieser Studie ist, dass sie die Reliabilität sowohl anhand des ICCs, als auch anhand des SEM und SRD berechnet. Als Limitation diskutieren die Forscherinnen lediglich die etwas geringe Anzahl Teilnehmender. Ganz am Ende der Studie, im Fazit, schreiben die Autorinnen, dass diese Testbatterie empfohlen werden kann, um herauszufinden, wann ein Sportler/eine Sportlerin nach einer Verletzung wieder in den Sport zurückkehren kann, da einige dieser Tests sehr sportspezifisch sind. Die Autorinnen dieser Arbeit erachten diese Aussage als inkorrekt. An der Studie nahmen nur gesunde Sporttreibende teil, wodurch keine Aussage über verletzte Sportlerinnen und Sportler getroffen werden kann. Zudem untersuchten die Autorinnen diese Tests lediglich auf die Reliabilität, wodurch keine Aussage dazu gemacht werden kann, wann die Athleten und</p>

	<p>Während der Studie gab es keine Drop-Outs. Bei den vier Testpersonen, die von Anfang an nicht teilnehmen konnten, werden die Gründe genannt.</p> <p>Datenerhebung: Die Datenerhebung für die Fragestellung nach der absoluten und relativen Reliabilität ist nachvollziehbar. Die Daten wurden bei allen Teilnehmenden gleich und von allen komplett erhoben.</p> <p>Messverfahren: Da für unsere Bachelorarbeit aus dieser Studie nur der Front Hop und der Side Hop Test relevant sind, wird an dieser Stelle nur auf die Messverfahren für diese beiden Tests eingegangen. Beim Front Hop wurden die Zentimeter, beim Side Hop die Anzahl Wiederholungen innerhalb 30 Sekunden gezählt. Es ist nicht ersichtlich, wer die Zentimeter und die Wiederholungen zählte und ob die Testpersonen über ihre Werte informiert wurden oder nicht. Beim Front Hop mussten die Hände an die Hüfte gehalten werden, was ihn etwas von den Vorgaben aus dem RTA-Algorithmus unterscheidet. Zudem wird beim Side Hop nichts darüber geschrieben, mit welchem Bein gestartet wird (was bei allen anderen Tests der Fall ist). Auch wird nicht erwähnt, dass das zweite Bein ebenfalls getestet wird. Dies ist erst aus den Tabellen entnehmbar. Des Weiteren wird nicht erwähnt, welche Zählmethode zum Eruiere der Fehl- und korrekten Sprünge des Side Hops angewendet wurde, um Fehler zu vermeiden.</p> <p>Datenanalyse: Es ist anzunehmen, dass die Datenerhebung von den Prüfern und Prüferinnen durchgeführt wurde. Es ist jedoch nicht klar, ob die Probanden und Probandinnen ihre Ergebnisse erfuhren oder nicht. Die verwendeten statistischen Verfahren sind sinnvoll. Die Autorinnen haben sich merklich mit der Bedeutung der Reliabilität auseinandergesetzt und deswegen nicht nur den ICC bestimmt, sondern auch den SEM. Warum allerdings der ICC(2.1) berechnet wurde, wird nicht beschrieben. So wäre für die Test-Retest-Reliabilität die Berechnung des ICC(1.1) oder (3.1) angebrachter gewesen, da diese den Rater nicht einbeziehen. Der ICC(2.1) ist der Konservativste von allen, da er im Nenner zusätzlich den Rater als Fehlerquelle mit einrechnet. Die verwendeten Tests setzen mindestens intervallskalierte Daten voraus, was die erhobenen Daten problemlos erfüllen. Die Höhe des Signifikanzniveaus wird angegeben, allerdings fehlt die Begründung.</p> <p>Ethik: Die Autorinnen weisen darauf hin, dass sie ethische Fragen diskutiert haben. Um welche Fragen es sich handelt ist in der Studie nicht ersichtlich.</p>		<p>Athletinnen wieder in den Sport zurückkehren können. Im Abstract ist das Fazit wesentlich schlüssiger, allerdings dürfte hier keine Aussage über den späten Rehabilitationsprozess gemacht werden, da die untersuchte Stichprobe nichts darüber aussagt. Eine weitere Schwäche ist, dass das gewählte Design mit lediglich zwei Durchführungen evtl. zu wenig sein könnte, aufgrund des Lernprozesses (siehe kritische Würdigung: Methode). Die Ergebnisse sind in der Praxis anwendbar, allerdings nur für eine sportliche Population im jungen Erwachsenenalter. Es wäre sowohl möglich, als auch spannend, diese Studie in einem anderen Setting, z.B. mit einer etwas weniger sportlichen Population zu wiederholen.</p>
--	---	--	---

Studie 2: Gustavsson, A., Neeter, C., Thomeé, P., Grävare Silbernagel, K., Augustsson, J., Thomeé, R., & Karlsson, J. (2006). A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. *Knee Surgery, Sports Traumatology, Arthroscopy*, 14(8), S. 778–788.

Zusammenfassung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Problembeschreibung: In der Rehabilitation von VKB Patientinnen und Patienten kommt es zur grundsätzlichen Frage, wann wieder mit physisch anstrengenden Aktivitäten begonnen werden darf.</p> <p>Ziel: Das Ziel dieser Studie ist, eine Testbatterie aus Sprungtests zu entwickeln, anhand derer man zwischen der verletzten und unverletzten Seite, bei Patientinnen und Patienten mit einer VKB-Verletzung, unterscheiden kann. Dazu gehört eine hohe Test-Retest-Reliabilität, Sensitivität, Spezifität und Treffgenauigkeit (Accuracy).</p> <p>Bezugsrahmen: In Literaturen werden verschiedene Sprungtests beschrieben. Dazu gehören Sprungtests auf Distanz, Zeit oder solche, die sich auf die gesprungene Höhe beziehen. Zur Evaluation bezüglich des funktionellen Outcomes nach VKB Verletzungen wird oft nur ein Sprungtest angewendet, nämlich der Front Hop. Allerdings ist die berichtete Sensitivität des Front Hops relativ gering (38-52%), um funktionelle Einschränkungen</p>	<p>Design:</p> <ul style="list-style-type: none"> - Test-Retest – Design - Die Studie möchte auch die Test-retest-Reliabilität erfassen, deshalb dieses Design. - Die Studie untersucht die fünf Tests auf die Spezifität, Sensitivität und Genauigkeit (Accuracy) und somit auf Teilaspekte der Validität. <p>Population: Die Population wird nicht definiert. Man kann davon ausgehen, dass es sich um VKB Patienten und Patientinnen handelt, die entweder operativ oder konservativ behandelt wurden.</p> <p>Stichprobe:</p> <ul style="list-style-type: none"> - 15 gesunde Probanden und Probandinnen (9m/6w) - 30 mit VKB konservativ (18m/12w) - 35 mit VKB operativ (25m/10w) <p>Einschlusskriterien:</p> <p>VKB konservativ:</p> <ul style="list-style-type: none"> - 4-44 Monate posttraumatisch - positiver «Vordere Schublade»- und Lachmantest durchgeführt von erfahrener orthopädischer Ärzteschaft und vom Physiotherapeuten/von der Physiotherapeutin des Patienten/der Patientin bestätigt - vorhandene Knieverletzung und Gefühl von Giving Way <p>VKB operativ:</p> <ul style="list-style-type: none"> - 6 Monate nach VKB-Rekonstruktion <p>Ausschlusskriterien:</p> <p>gesunde Probanden und Probandinnen:</p> <ul style="list-style-type: none"> - Rücken-, Hüft-, Knie- oder Fussbeschwerden <p>VKB konservativ:</p> <ul style="list-style-type: none"> - Akute Schmerzen oder Schwellung - Vorgängige Operation an beiden Beinen <p>VKB operativ:</p> <ul style="list-style-type: none"> - Akute Schmerzen oder Schwellung - Vorgängige Operation an beiden Beinen 	<p>Welche Ergebnisse werden präsentiert?</p> <ol style="list-style-type: none"> 1. Test-Retest-Reliabilität der fünf Sprungtests 2. Vergleich der Ergebnisse der Sprungtests zwischen männlichen und weiblichen, gesunden Testpersonen 3. Vergleich der Ergebnisse der verletzten und unverletzten Seiten bei VKB-Probanden und -Probandinnen 4. Vergleich der gesunden Testpersonen mit den VKB-Probanden und -Probandinnen bezgl. Seitendifferenz 5. Sensitivität, Spezifität, Accuracy 6. Korrelation der fünf Hop Tests untereinander 7. Ausgearbeitete Testbatterie <p>Welches sind die zentralen Ergebnisse der Studie?</p> <p>Zu 1. Die Studie zeigt signifikante Unterschiede zwischen Test 1 und Test 2 der Kontrollgruppe bei</p> <ul style="list-style-type: none"> - Männern (Front Hop/Square Hop) - Allen Testpersonen gemeinsam (Front Hop/Square Hop/Side Hop) <p>Es gibt keine signifikanten Unterschiede zwischen Test 2 und Test 3 bei allen 5 Sprungtests. Der ICC zwischen Testgelegenheit 2 und 3 variiert zwischen 0.85 – 0.97. Der Methodological error zwischen 3-6%.</p> <p>Zu 2. Beim Vertical Jump, Front Hop und beim Drop Jump followed by a Double Hop for distance werden signifikante Unterschiede zwischen Männern und Frauen festgestellt (Testgelegenheit 3). Beim Square Hop und beim Side Hop werden keine signifikanten Unterschiede zwischen Männern und Frauen festgestellt (Testung 3). Vergleicht man die absoluten Werte des Unterschieds von Seite zu Seite bei Männern und Frauen miteinander, können keine signifikanten Differenzen</p>	<p>Werden signifikante und nicht signifikante Ergebnisse erklärt?/Wie interpretieren die Forschenden die Ergebnisse? Es wird beschrieben ob etwas signifikant ist oder nicht. Der Wert ist auch sichtbar. Teils wird jedoch nicht genau erläutert, um welchen Wert es sich exakt handelt. Viele «nicht signifikante Werte» werden nicht erwähnt. Nur die offensichtlichen. Die Forschenden beschreiben, dass die Haupterrungenschaft der Studie die Testbatterie, bestehend aus dem Vertical Jump, dem Front Hop und dem Side Hop Test, sei. Diese Testbatterie habe nämlich eine hohe Fähigkeit um zwischen der verletzten und der unverletzten Seite eines Patienten und einer Patientin unterscheiden zu können.</p> <p>Kann die Forschungsfrage aufgrund der Daten beantwortet werden? Bedingt. Das Ziel der Forschenden ist, eine Testbatterie zu entwickeln, anhand derer man Unterschiede zwischen dem verletzten und dem nicht verletzten Bein besser feststellen kann. Dies ist ihnen gelungen. Um jedoch die genaue Aussagekraft der Testbatterie zu bestätigen, benötigt es in Zukunft weitere Untersuchungen.</p> <p>Werden Limitationen diskutiert? Es wird beschrieben, wo noch mehr Forschungsbedarf vorhanden ist. Ansonsten werden keine Limitationen genannt.</p>

<p>aufzudecken, welche mit einer VKB Verletzung assoziiert sind. Verschiedene Autoren und Autorinnen beschreiben, dass die Sensitivität durch eine Kombination von verschiedenen Sprungtests erheblich gesteigert werden kann.</p> <p>Forschungsfrage: Hypothese: Bei einer Testbatterie, die verschiedene Sprungqualitäten untersucht (maximale Sprungleistung auf einem Bein, kombiniert mit einem Sprungtest der auf die Erschöpfung abzielt), ist die Wahrscheinlichkeit höher, eine Diskrepanz bei den Sprungtests aufzudecken, als wenn nur ein einzelner Sprungtest durchgeführt wird. Die Forschenden gehen davon aus, dass eine solche Testbatterie auch zu einer höheren Sensitivität führt.</p>	<p>Wie wurde die Stichprobe gezogen? Es wird nichts darüber geschrieben, wie die Versuchspersonen erreicht wurden. Somit wird von einem Non-Probability Sampling ausgegangen.</p> <p>Wie wird die Auswahl der Teilnehmenden beschrieben und begründet? Alle Patientinnen und Patienten wurden aus derselben Physiotherapieklinik rekrutiert und durchliefen Physiotherapie unter Supervision anhand eines kriterienbasierten Protokolls.</p> <p>Gibt es verschiedene Studiengruppen?</p> <ul style="list-style-type: none"> - Testgruppe: VKB kons./VKB operat. - Kontrollgruppe: Gesunde Testpersonen <p>Datenerhebung: Welche Art von Daten wurde erhoben?</p> <ul style="list-style-type: none"> - physiologische Messungen (Sprungtests) - schriftliche Befragung (Tegner Score) <p>Wie häufig wurden die Daten erhoben?</p> <ul style="list-style-type: none"> - Sprungtests: 3x an verschiedenen Tagen (Intervall 3-13 Tage/3-19 Tage) - Tegner Score: 2x am selben Tag (1x Status vor Verletzung, 1x aktueller Status) <p>Messverfahren & oder Interventionen: Welche Messinstrumente wurden verwendet?</p> <ul style="list-style-type: none"> - Tegner Score: Messinstrument zur Erhebung der aktuellen physischen Aktivität des Patienten/der Patientin und zur Erhebung der physischen Aktivität vor der Verletzung. Die Patientinnen und Patienten können auf einer Skala von 0-10 ihr Aktivitätslevel in verschiedenen Kategorien angeben. - MuscleLab, Ergotest Technology (Höhenmessung des Vertical Jumps mittels Infrarot) - Massband, Stoppuhr <p>Welche Sprungtests werden getestet?</p> <ul style="list-style-type: none"> - Vertical jump (cm) - Front Hop (cm) (Entspricht von der Durchführung her demjenigen unserer Bachelorarbeit, allerdings mussten die Testpersonen bei dieser Studie die Hände hinter dem Rücken verschränken.) - Drop jump followed by a Double Hop for distance (cm) 	<p>festgestellt werden. Aus diesem Grund wurden weitere Berechnungen eingestellt.</p> <p>Zu 3. Es kann kein signifikanter Unterschied beim Square Hop im Seitenvergleich bei konservativ behandelten Patienten festgestellt werden. Alle anderen Tests weisen sowohl in der konservativen als auch in der operativen Gruppe signifikante Unterschiede im Seitenvergleich auf.</p> <p>Zu 4. VKB-Probandinnen und -Probanden weisen signifikant grössere Seitenunterschiede bei den Tests auf im Gegensatz zu den gesunden Versuchspersonen (ausgenommen Square Hop → keinen signifikanten Unterschied).</p> <ul style="list-style-type: none"> - LSI Kontrollgruppe: zw. 67-100% - LSI VKB konserv.: zw. 23-57% - LSI VKB operativ: zw. 14-49% <p>Zu 5.</p> <ul style="list-style-type: none"> - Sensitivität kons.: 43% - 77% - Sensitivität operativ: 51-86% - Spezifität: 67% – 100% - Genauigkeit konservativ : 58% - 80% - Genauigkeit operativ: 56-86% <p>Zu 6.</p> <ul style="list-style-type: none"> - Korrelation Kontrollgruppe – Testgruppe zwischen den 3 Maximum-Sprungtests (Vertical Jump, Front Hop und Drop Jump followed by a Double Hop for distance): 0.70-0.94 - Korrelation Kontrollgruppe – Testgruppe zwischen den 2 Ausdauer-Sprungtests (Square Hop und Side Hop): 0.72-0.92 - Korrelation zwischen Maximum-Sprungtests und Ausdauer-Sprungtests bei Kontrollgruppe: 0.29-0.57 - Korrelation zwischen Maximum-Sprungtests und Ausdauer-Sprungtests bei Patientinnen und Patienten: 0.62-0.88 	<p>Werden die Ergebnisse mit ähnlichen Studien verglichen? Ja zum Beispiel mit Östenberg et al., Itoh et al., Noyes et al., Barber et al., Wilson et al. und Manske et al..</p> <p>Welche Implikationen für die Praxis, Theorien und zukünftige Forschung sind beschrieben? Testbatterien sind sinnvoll, da diese genauere Werte hervorrufen als nur ein einzeln durchgeführter Hop Test. Es sei beängstigend, dass nach 6Mt. Postop. resp. 11Mt. Posttraum. Nur 1 von 10 Pat. einen normalen LSI zustande brachte.</p>
---	--	---	---

	<ul style="list-style-type: none"> - Square Hop (Wdh. innerhalb 30sek) - Side Hop (Wdh. innerhalb 30sek) <p>Datenanalyse: Welches Datenniveau weisen die erhobenen Variablen auf?</p> <ul style="list-style-type: none"> - Cm = Proportionalskaliert - Wdh. Innerhalb von 30 Sek. = Proportionalskaliert - Tegner Score (Fragebogen) = Ordinalskaliert <p>Welche statistischen Verfahren wurden zur Datenanalyse verwendet?</p> <p>Ermittlung der Reliabilität:</p> <ul style="list-style-type: none"> - Intraklassenkorrelationskoeffizient (ICC) mit einem Konfidenzintervall (CI) von 95% - Methodological Error (Vermutlich ein Wert, der die systematische Differenz der ersten zur zweiten Testgelegenheit darstellt) - Wilcoxon's signed rank test (für abhängige Werte, also für die gleiche Testgruppe) - Mann-Whitney U test (U-Test) (für unabhängige Werte, also für unterschiedliche Testgruppen) - Spearman Rank (Beziehung zwischen abhängigen Variablen) <p>Ermittlung der Validität:</p> <ul style="list-style-type: none"> - Lower Limb Symmetrie Index (LSI) (betroffenes Bein geteilt durch das nicht betroffene Bein mal 100) - Sensitivität (= Anzahl der Patientinnen und Patienten, bei denen der LSI unter 90% ist, geteilt durch die gesamte Anzahl der Patienten) - Spezifität (Anzahl gesunder Personen, die einen normalen LSI haben geteilt durch die Anzahl aller gesunden Personen) - Exaktheit (Accuracy = Anzahl der Patientinnen und Patienten, bei denen der LSI unter 90% ist, plus die Anzahl der gesunden Personen, die einen normalen LSI haben geteilt durch die Anzahl aller Patientinnen und Patienten und gesunder Personen) <p>Wurde ein Signifikanzniveau festgelegt? JA: $P \leq 0.05$ = Signifikant</p> <p>Ethik: Es werden keine ethischen Fragen diskutiert. Die Studie wurde jedoch durch das Human Ethics Committee der Universität Göteborg/Schweden genehmigt.</p>	<p>Zu 7. Der Vertical Jump, Front Hop und Side Hop Test wurden für Testbatterie ausgewählt. Square Hop fällt aus Testbatterie raus weil:</p> <ul style="list-style-type: none"> - Er nicht in der Lage war, zwischen der Seitendifferenz von gesunden und verletzten Personen zu unterscheiden. - Er die tiefste Spezifität (67%) und tiefste Accuracy (56%) hatte. - Die Korrelation zwischen dem Square Hop und dem Side Hop (0.72-0.92) bei allen Testpersonen, auf einen engen Zusammenhang der beiden Ausdauer Hop Tests hinweist. <p>Der Drop Jump followed by a Double Hop for distance fällt aus der Testbatterie raus, weil:</p> <ul style="list-style-type: none"> - Er eine sehr hohe Korrelation mit dem Front Hop (>0.80) aufweist. - Die Accuracy ist ähnlich bis gleich mit dem Front Hop. <p>→ Konklusion: Single Leg und Drop Jump messen das Gleiche</p> <p>Werden die Ergebnisse verständlich präsentiert (Textform, Tabellen, Grafiken)? Mehrheitlich JA. Die Signifikanz ist jeweils gut angegeben. Die Tabellen wirken etwas unübersichtlich (v.a. Tab. 2). Auch ist teils schwer ersichtlich, um welche Testgruppe es sich handelt, die in der Tabelle aufgeführt ist. Tab. 5 ist beim Square Hop und Side Hop unpräzise was die Masseinheit betrifft (Jumps = 1 Sprung oder 1 Runde oder wie?)</p>	
--	---	--	--

Würdigung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Fragestellung: Die Fragestellung bearbeitet für die Berufspraxis wichtige Aspekte. Für unsere Bachelorarbeit ist die Fragestellung jedoch nur teilweise relevant. So sind wir an der Reliabilität des Front Hop, Square Hop und Side Hop Tests interessiert. Die Studie untersucht diese Tests auch auf die Sensitivität, Spezifität und Accuracy, also auf Teilaspekte der Validität. Allerdings untersucht die Studie, ob diese Tests eine VKB-Verletzung erkennen können. Dies interessiert uns für diese Bachelorarbeit nicht in hohem Masse.</p> <p>Ist die Forschungsfrage klar definiert? Ja, die Forschungsfrage, in diesem Falle die Hypothese ist klar definiert.</p> <p>Wird das Thema / Problem im Kontext dargestellt? Ja, es wird erläutert, aus welchem Grunde diese Überlegungen zustande gekommen sind.</p>	<p>Design: Es ist eine logische Verbindung zwischen Forschungsfrage und gewähltem Design gegeben. Die Forschenden versuchen die Gefahren der internen Validität durch folgende Schritte zu kontrollieren:</p> <ul style="list-style-type: none"> - Die Tests wurden immer durch den gleichen Prüfer durchgeführt. - Die Probandinnen und Probanden wurden jeweils alle verbal ermutigt. (es steht nichts darüber, ob alle gleich ermutigt wurden) - Die Schuhe der Probandinnen und Probanden waren standardisiert. - Die Probandinnen und Probanden durften am Tag davor keine strengen physischen Aktivitäten durchführen. <p>Zur Raumtemperatur, Zeit der Testdurchführungen (zu welcher Tageszeit) und zu der Stimmungslage der Patientinnen und Patienten wird nichts gesagt. Es ist zudem nicht klar, ob die Patienten und Patientinnen ihre Ergebnisse der Sprungtests kannten. Dies könnte zu erheblichen Verzerrungen der Resultate führen. Für die gesunden Personen wurde die Abfolge der Sprungtests willkürlich festgelegt. Es ist nicht ersichtlich warum dies auf jene Weise durchgeführt wurde.</p> <p>Stichprobe: Die Stichprobe scheint angebracht, da einige Varianten beachtet werden (gesund, konservativ, operativ). Allerdings ist die Zeitspanne, zwischen Operation/Trauma und Studienbeginn der Stichprobe mit 4-44 Monaten sehr gross. Es stellt sich die Frage, ob Personen, die sich 44 Monate postoperativ/posttraumatisch befinden, nicht bereits als ausgeheilt gelten. Die Stichprobe ist insofern repräsentativ, da nicht nur Sportlerinnen und Sportler genommen wurden, sondern auch weniger sportliche. So können die Ergebnisse auf eine breite Population übertragen werden. Die Stichprobengruppen pro «Kategorie» sind eher klein, v.a. die Kontrollgruppe. Die Stichprobengrösse wird nicht begründet. Über Drop-Outs wird nichts berichtet. Aufgrund der Resultate kann man jedoch davon ausgehen, dass es keine gab, da die Zahlen mit den Anfangswerten übereinstimmen.</p> <p>Datenerhebung: Die Datenerhebung für die Fragestellung ist nachvollziehbar. Die Datenerhebungen sind beinahe identisch. Einzig die verflossene Zeit zwischen Test 1 und Test 2, respektive Test 2 und Test 3, ist unterschiedlich. Der Zeitraum der Kontrollgruppe ist definiert (3-13d/3-19d). Ob dieser nun auch für die Testgruppe mit VKB-Patienten und -Patientinnen zählt ist unklar. Falls JA, wäre dies kritisch zu betrachten, denn theoretisch könnte es sein, dass der eine Patient/eine Patientin innerhalb von 7 Tagen alle 3 Tests durchgeführt und ein anderer Patient/eine andere Patientin innerhalb von 33 Tagen alle 3 Tests absolviert hätte. Diese Zeitspannen könnten massive Unterschiede hervorrufen, da der</p>	<p>Sind die Ergebnisse präzise? Die Ergebnisse sind teilweise präzise. Es wird wohl meistens klar erwähnt, um welche Werte es sich handelt, doch gibt es auch Stellen, an denen man erst durch genaues betrachten der Tabelle sieht, wie dieser Wert zustande kam, und dass er unter anderen Aspekten viel tiefer liegen würde (Bsp. Seite 5 → Results → Ende 1. Abschnitt).</p> <p>Entsprechen Tabellen den folgenden Kriterien:</p> <p>präzise und Vollständig? Die Tabellen erscheinen mehrheitlich vollständig und präzise. Titel sind vorhanden, es ist jedoch manchmal unklar, auf welche Stichprobengruppe sich die Werte beziehen (Bsp. Tab 2). Zeilen und Spalten sind definiert. Die Testergebnisse werden in den verschiedensten Kombinationen verglichen.</p> <p>Ergänzung zum Text? Die Tabellen bilden eine Ergänzung zum Text. Somit werden die Resultate verständlicher. Die wichtigsten Ergebnisse daraus sind im Text erwähnt.</p>	<p>Werden alle Resultate diskutiert? Es werden hauptsächlich jene Resultate diskutiert, welche auch in bereits durchgeführten Studien erwähnt und somit miteinander verglichen werden. Es werden nicht alle Ergebnisse aus den Tabellen diskutiert.</p> <p>Stimmt die Interpretation mit den Resultaten überein? Grundsätzlich JA. Ob die hohen ICC-Werte auch von der Heterogenität der Personen beeinflusst werden, ist nicht klar ersichtlich. Allerdings sind die ICC-Werte bei den Männern und den Frauen einzeln kleiner als bei beiden zusammen. Dies könnte durch die erwähnte Heterogenität bedingt sein.</p> <p>Werden die Resultate in Bezug auf die Fragestellung/Hypothesen, Konzepte und anderen Studien diskutiert und verglichen? Die Resultate werden in Bezug auf die Hypothese erläutert. Auch werden die Werte mit anderen Studien verglichen.</p> <p>Wird nach alternativen Erklärungen gesucht? Teilweise wird beschrieben, weshalb ein Wert von jenen, anderer Studien abweicht. Grundsätzlich wird der Fokus auf die positiven Werte gelegt.</p> <p>Ist diese Studie sinnvoll? Werden Stärken und Schwächen aufgewogen? Es werden hauptsächlich Stärken hervorgehoben, v.a. bei der Interpretation der Tabellen. Wie tiefe Werte zustande kommen und was diese bedeuten, wird nicht erwähnt. In der Auswahl der Tests für die Testbatterie wird dafür erklärt, weshalb zwei Tests nicht verwendet wurden.</p>

	<p>Heilungsprozess noch immer voranschreitet. Die Daten wurden jeweils von allen Teilnehmenden erhoben und fast immer tabellarisch aufgezeigt. Das exakte Profil eines jeden einzelnen ist nicht ersichtlich. Tab. 2 zeigt nur die Korrelationswerte. Es sind keine absoluten Zahlen vorhanden, um diese zu überprüfen.</p> <p>Messverfahren: Bei den Messinstrumenten handelt es sich um zuverlässige Instrumente, die genaue Werte liefern. Der Side Hop und der Square Hop wurden anhand eines Videos aufgezeichnet, um die genaue Anzahl korrekter Sprünge evaluieren zu können. Die Auswahl des Tegner Scores wird begründet. Dieser wird jedoch für die Studie nicht wirklich gebraucht. Die anderen Messinstrumente, wie Stoppuhr und Massband, sind aufgrund der standardisierten Ausführung der Hop Tests bereits gegeben.</p> <p>Datenanalyse: Die Durchführung der Tests ist klar definiert. Statistische Verfahren wurden mehrheitlich sinnvoll angewendet. Werden die Sprungtests später für Verlaufsmessungen eingesetzt, wäre bei der Reliabilität der SEM und der SDC-Wert noch aussagekräftig gewesen, da nur anhand von dem gesagt werden kann, ob wirklich eine Veränderung stattgefunden hat. In der Studie fehlt sowohl die Angabe, um welchen ICC es sich handelt, als auch eine Begründung warum dieser ICC berechnet wurde. Die Tests entsprechen den Datenniveaus. Allerdings stellt sich die Frage, warum sie einige Tests verwendet haben, die nur ordinalskalierte Daten voraussetzen, denn das Datenniveau der meisten Variablen war porportionalskaliert. Die einzige Begründung wäre der ordinalskalierte Tegner Score. Dieser wurde für die Berechnungen jedoch gar nicht verwendet. Die statistischen Angaben erlauben nicht immer eine genaue Beurteilung. Die p-Werte sind phasenweise nicht ganz nachvollziehbar. Man muss sich teilweise etwas «durchsuchen», da gewisse Tabellen unübersichtlich gestaltet sind. Das Signifikanzniveau ist im Text definiert, wird aber nicht begründet. Es wird nirgends begründet, was die Prozentzahlen überhaupt bedeuten, respektive ab welcher Prozentzahl (z.B. beim ICC) ein Wert gut oder schlecht oder sehr gut ist. Im Text steht ausschliesslich der Wert, und beispielsweise, dass dieser einer hohen Test-Retest-Reliabilität entspricht.</p> <p>Ethik: Es wurde beim Ethischen Komitee eine Genehmigung eingeholt und die Versuchspersonen willigten schriftlich ein. In welcher Beziehung die Teilnehmenden zu den Forschenden stehen, ist nicht ersichtlich. Bekannt ist nur, woher die Teilnehmenden rekrutiert wurden. Woher die Auswahl der gesunden Versuchspersonen stammt, weiss man als Leserin und Leser jedoch nicht. Welchen Hintergrund die Rater haben, ist unbekannt.</p>		<p>Die Studie ist grundsätzlich sinnvoll und mehrheitlich aufschlussreich.</p> <p>Wie und unter welchen Bedingungen sind die Ergebnisse in die Praxis umsetzbar? Die Ergebnisse sind grundsätzlich gut umsetzbar in der Praxis. Da die Testbatterie viel Zeit benötigt, könnte der zeitliche Aspekt ein Problem in der Praxis darstellen. Für ältere Patientinnen und Patienten ist die Testbatterie wahrscheinlich schwierig durchführbar. Doch ist die Wahrscheinlichkeit, dass eine ältere Person in eine derart sportliche Aktivität zurückkehrt, dass jene Tests nötig würden, gering.</p> <p>Wäre es möglich diese Studie in einem anderen klinischen Setting zu wiederholen? Die Studie kann überall dort wiederholt werden, wo ein MuscleLab oder Ergotest Technology vorhanden sind, da ansonsten der Vertical Jump nicht durchführbar ist.</p>
--	--	--	--

Studie 3: Haitz, K., Shultz, R., Hodgins, M., & Matheson, G. O. (2014). Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation. *Journal of Orthopaedic & Sports Physical Therapy*, 44(12), S. 947–954.

Zusammenfassung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Ziel: Ziel der Studie ist, die Augenscheinvalidität, die Test-Retest- und die Interrater-Reliabilität der Functional Lower Extremity Evaluation (FLEE) zu bestimmen. Zudem wollten die Forschenden den FLEE-Score für gesunde Personen dokumentieren, um in der Rehabilitation einen Vergleich zu Echtwerten zu haben.</p> <p>FLEE: Der FLEE ist eine Testbatterie mit 8 standardisierten funktionellen Tests mit 3 Komponenten (Kontrolle, Kraft, Ausdauer). Der FLEE wurde medizinischem, physiotherapeutischem und sporttherapeutischem Fachpersonal der Universität Stanford entwickelt. Der ganze Test dauert 45 Minuten und ist somit in etwa so lang wie eine Trainingseinheit. Die unterschiedlichen Tests des FLEE sind von einfach zu komplex aufgebaut. Wenn jemand bei einem der ersten Tests bereits Schmerzen verspürt, sollen die weiteren Tests nicht durchgeführt werden. Die Reliabilität und die Normalwerte dieser Testbatterie sind unbekannt.</p>	<p>Design:</p> <ul style="list-style-type: none"> - Test-Retest- Design - Interrater-Reliabilität - Augenscheinvalidität (face validity) <p>Es wird nicht begründet, weshalb diese Designs gewählt wurden. Einzig die Augenscheinvalidität wird damit begründet, dass Personen im sportmedizinischen Bereich typischerweise funktionelle Assessmenttestungen durchführen.</p> <p>Stichprobe: Population: Die Population wird nicht definiert. Vermutlich handelt es sich um sportliche Personen, die wieder in ihren Sport zurückkehren wollen.</p> <p>Stichprobe:</p> <ul style="list-style-type: none"> - Es nahmen 49 unverletzte Sportstudentinnen und Studenten der Stanford University an der Studie teil. - Nach der ersten Testgelegenheit gab es neun Drop-Outs (wegen Krankheit, Muskelkater eines Workouts und Terminkonflikten). Es beendeten demnach insgesamt 40 Personen die Studie. - Feldhockey, Kunstturnen, Rugby, Fussball, Leichtathletik, Triathlon, Volleyball, Wrestling - 22m/18w (nach Drop-Outs) <p>Wie wurde die Stichprobe gezogen: Es wird nichts darüber geschrieben, wie die Probanden und Probandinnen erreicht wurden. Somit wird von einem Non-Probability Sampling ausgegangen.</p> <p>Begründung Auswahl der Teilnehmenden: Die Auswahl wird nicht begründet, man weiss lediglich woher sie kommen.</p> <p>Datenerhebung: Datenerhebung der Augenscheinvalidität: Für die Erhebung der Augenscheinvalidität wurden zwei Fragen (siehe unten) an Sportphysiotherapeuten und -therapeutinnen in Kalifornien, an Sporttrainer/Sporttrainerinnen und an Sportmediziner und -medizinerinnen geschickt. Diese hatten im Durchschnitt jeweils 13 Jahre Berufserfahrung. Im Ganzen wurden 200 Personen angeschrieben. 73 Personen beantworteten die Fragen. Die Forschenden begründen die Auswahl dieser Personen damit, dass üblicherweise diese Personengruppen ein funktionelles Assessment entwickeln und somit gut qualifiziert sind, die Augenscheinvalidität zu bestimmen.</p>	<p>Welche Ergebnisse werden präsentiert?</p> <ul style="list-style-type: none"> - Ergebnisse der Umfrage zur Validität (Augenscheinvalidität) - Reliabilität des FLEEs ohne LEFT anhand des ICC - Seitenvergleich aller Tests, geschlechtergetrennt <p>Welches sind die zentralen Ergebnisse der Studie? Validität:</p> <ul style="list-style-type: none"> - Aussage, diese Testaufgaben sind sehr wichtig → 58-71% der Befragten - Aussage, Tests regelmässig zu nutzen → 26-45% aller Befragten <p>Reliabilität:</p> <ul style="list-style-type: none"> - Test-Retest: ICC 0.71-0.95 - Interrater: ICC 0.83-1.00 - SRD und SEM sind ebenfalls dokumentiert <p>Normale Werte des Seitenvergleichs: Durchschnittlich 98.6-114.4% LSI (LSI = nicht dominantes Bein durch dominantes Bein)</p>	<p>Werden signifikante und nicht signifikante Ergebnisse erklärt? Wie interpretieren die Forschenden die Ergebnisse? Bezüglich der Augenscheinvalidität sind die Ergebnisse für «Hohe Wichtigkeit der Tests» ca. 1.5-2x grösser als die «regelmässige Anwendung» dieser Tests. Die Forschenden meinen, ein Grund dafür könnte die noch nicht erfasste Reliabilität sowie fehlende Normwerte sein.</p> <p>Für die Reliabilität sagen die ICC-Werte aus, dass jeder der durchgeführten Tests eine akzeptable Wiederholbarkeit für die Forschung bei Sportverletzungen aufweist.</p> <p>Der Timed Lateral Step-Down hat tiefere ICC-Werte als die anderen Tests. Es wird vermutet, dass dies aufgrund subjektiver Bewertung geschehen sein könnte. Obwohl eine hohe Subjektivität im Timed Lateral Step-Down vorliegt, wird dieser als wichtig eingestuft, da er eine Aussage über die qualitative Ausführung macht, was eine Wiederverletzung verhindern könnte.</p> <p>Normal-Werte: Die Forschenden meinen, die Normal-Werte können auf Sportlerinnen und Sportler im Rehabilitationsprozess übertragen werden, als Leitplanke für das angestrebte Ziel</p> <p>Kann das Ziel der Studie aufgrund der Daten beantwortet werden? Das Ziel der Studie kann aufgrund der erhobenen Daten erreicht werden.</p>

<p>Argumente für Forschungsbedarf: Es ist schwierig, den Return-to-Play Status nach einer Verletzung zu bestimmen. Die Wiederverletzungsrate ist 4x höher als jene einer Erstverletzung, da oftmals die vollkommene Funktion noch nicht zurückgekehrt ist. Aus diesem Grund sollte Klinikpersonal standardisierte Tests einsetzen, um bestimmen zu können, wann die Patientinnen und Patienten wieder in den Sport zurückkehren dürfen. Es wird darauf verwiesen, dass Noyes et al. bereits eine Testbatterie von Sprungtests zusammengestellt haben, die jedoch eine schlechte Sensitivität zeigt. Deshalb stellt sich die Frage, ob sich eine Testbatterie mit anderen Tests besser dazu eignen würde.</p> <p>Forschungsfrage/Hypothese Es ist keine Forschungsfrage oder Hypothese vorhanden. Allerdings wird ein Ziel formuliert (siehe oben).</p>	<p>Frage 1: Die Personen konnten für jeden Test des FLEE wählen, wie wichtig dieser ist. Dabei konnten sie zwischen «gar nicht wichtig», «etwas wichtig», «schon eher wichtig», «wichtig» und «sehr wichtig» entscheiden.</p> <p>Frage 2: Die Personen wurden gefragt, wie oft sie die Tests in der Praxis anwenden. Dazu konnten sie zwischen 0-10%, 10-25%, 25-40%, 40-55%, 55-70%, 70-85% und 85-100% auswählen.</p> <p>Häufigkeit der Datenerhebung für die Ermittlung der Reliabilität: - 2x Test 1-7 des FLEE, 1 Woche lag dazwischen - Test 8 (LEFT) wurde aufgrund Zeitmangels, dem erwarteten Lerneffekt und der bereits bekannten Reliabilität nur 1x erhoben.</p> <p>Datenerhebung der Interrater-Reliabilität: Für die Ermittlung der Interrater-Reliabilität wurden drei Rater (ein Physiotherapeut, ein Sporttherapeut und ein Forschungsassistent) ausgesucht. Der Physiotherapeut hatte 11 Jahre Erfahrung als Sporttrainer und 9 Jahre Erfahrung als Physiotherapeut. Der Sporttherapeut arbeitete seit 8 Jahren in der Sportmedizin. Beide Rater arbeiteten Vollzeit und hatten bereits einige Monate bis einige Jahre Erfahrung mit dem FLEE. Der Forschungsassistent wurde 3 Monate lang in den FLEE eingeführt, um die Studienteilnehmer bewerten zu können. Der Forschungsassistent verwaltete und zählte die Werte der beiden Testgelegenheiten für jeden Sportler. Die anderen zwei Rater notierten die Werte für jeden Sportler bei beiden Testgelegenheiten. Der Forschungsassistent gab den Sportlern jeweils vor jeder Testgelegenheit Informationen zu den Tests. Die anderen zwei Rater gaben keine weiteren Instruktionen. Die Rater kannten die Werte der anderen Rater nicht.</p> <p>Messverfahren & oder Interventionen: Verwendete Messinstrumente:</p> <ul style="list-style-type: none"> - Stoppuhr (Timed Lateral Step-Down, Lateral Leap and Catch, Single Leg Timed Hop, Square Hop, LEFT) - Massband (Front Hop, Single Leg Triple Hop, Crossover Hop for distance) - Metronom (Lateral Step Down, Lateral Leap and Catch) <p>Getestete Interventionen:</p> <ul style="list-style-type: none"> - Lateral Step Down - Lateral Leap and Catch - Front Hop (Die Pat. mussten die Hände bei den Hüften einstützen, somit weicht er etwas von der Definition des Front Hops in unserer Bachelorarbeit ab.) - Single Leg Timed Hop - Single Leg Triple Hop - Crossover Hop for distance - Square Hop Test - LEFT 	<p>Werden die Ergebnisse verständlich präsentiert (Textform, Tabellen, Grafiken)? Die Resultate sind in Tabellen erfasst und werden in einem sehr kurz gehaltenen Text wiedergegeben. Die Tabellen sind verständlich dargestellt.</p>	<p>Werden Limitationen diskutiert? Die Forschenden diskutieren, dass nur 1 von 3 Testern den Versuchspersonen Instruktionen gab, obwohl alle 3 unabhängig voneinander Resultate abnahmen und Punkte vergaben. Im wirklichen Leben sei es jedoch nicht so und deswegen könnten die Ergebnisse verzerrt sein. Eine zweite Limitation sei, dass die Normal-Daten von Sport zu Sport sehr variieren können und dass die Stichprobe zu klein war. Aus diesem Grund könne man die Normdaten nicht auf jeden übertragen. Die Forschenden weisen auch darauf hin, dass in dieser Studie die Reliabilität bei gesunden Personen erfasst wurde und dass diese deswegen nicht auf verletzte Personen übertragen werden darf.</p> <p>Werden die Ergebnisse mit ähnlichen Studien verglichen? Die Test-Retest Ergebnisse wurden mit ähnlichen Studien verglichen. Die Interrater-Reliabilität wurde bis anhin von keiner anderen Studie ausgewertet.</p> <p>Welche Implikationen für die Praxis, Theorien und zukünftige Forschung sind beschrieben? Als nächstes müsste die Interrater-Reliabilität an verletzten Sportlerinnen und Sportlern durchgeführt und evaluiert werden. Diese Testpersonen sollten sich in verschiedenen Rehabilitationszeitphasen befinden. Die Test-Retest-Reliabilität sei eher schwierig an verletzten Probandinnen und Probanden durchzuführen, da eine funktionelle Verbesserung sehr schnell voranschreiten kann. Die vorliegende Studie ist der erste Schritt in der Erforschung des FLEEs.</p>
--	--	--	---

	<p>Datenanalyse:</p> <p>Datenniveau der erhobenen Variablen:</p> <ul style="list-style-type: none"> - Cm = Proportionalskaliert - Wdh. in 30 Sek = Proportionalskaliert <p>Berechnete statistische Werte:</p> <ul style="list-style-type: none"> - ICC mit einem 95% CI (Reliabilität Test 1-7) - ICC(3,1) (Test-Retest-Reliabilität) - ICC(2,1) (Interrater-Reliabilität) - SEM (Standard Error of Measurement) - SRD (Smallest Real Difference) - LSI (Symmetrie Index) - SD (Standard Deviation) <p>Signifikanzniveau:</p> <p>Augenscheinvalidität:</p> <p>Für die Augenscheinvalidität wurden bei Frage 1 die Antworten «wichtig» und «sehr wichtig» mit «hoher Wichtigkeit» gewertet.</p> <p>Bei der Frage 2 wurden Werte von 70-100% als sehr häufige Einsetzung der Tests gewertet.</p> <p>Reliabilität:</p> <p>Die Bedeutung des ICC Wertes wurde festgelegt. So entsprechen Werte von 0.9-1 ausgezeichneter Reliabilität, von 0.8-0.9 hoher Reliabilität und von 0.6-0.8 ausreichender Reliabilität.</p> <p>Ethik:</p> <p>Teilnehmende der Studie gaben ihre schriftliche Einwilligung, welche protokolliert wurde. Zudem wurden die Rechte der Teilnehmenden geschützt. Über das Einholen einer Bewilligung bei einer Ethikkommission wird nichts berichtet.</p>		
--	--	--	--

Würdigung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Fragestellung: Die Studie zeigt die Test-Retest und Interrater-Reliabilität für verschiedene Assessmenttests auf, unter anderem für den Front Hop und den Square Hop Test, welche in der BA bearbeitet werden. Jedoch nicht für Reha-Patienten und -Patientinnen, sondern für unverletzte Versuchspersonen (Sportstudenten und -studentinnen). Zudem wird die Augenscheinvalidität der Tests untersucht. Leider handelt es sich dabei fokussiert um die Fragestellung, ob diese Tests wichtig sind und in der Praxis angewendet werden. Für uns wäre es spannender gewesen, wenn die Tests daraufhin geprüft worden wären, ob sie wirklich das messen, was behauptet wird.</p> <p>Ist die Forschungsfrage bzw. das Ziel klar definiert? Die Studie hat weder eine Fragestellung noch eine Hypothese aufgestellt. Das Ziel ist jedoch klar definiert.</p> <p>Wird das Thema / Problem im Kontext dargestellt? Das Problem ist klar definiert, es wird ebenfalls auf den Hintergrund dieser Studie, sowie bereits durchgeführter Studien verwiesen.</p>	<p>Design: Für das Ziel der Studie ist das gewählte Design angebracht.</p> <p>Gefahren der internen und externen Validität: Bei der Interrater-Reliabilität, bei welcher der Fehler vom Rater herkommt, versuchen die Forschenden die Gefahren der internen Validität durch folgende Schritte zu kontrollieren:</p> <ul style="list-style-type: none"> - Die drei Tester wussten die Ergebnisse der anderen Tester nicht. - Die Versuchspersonen wurden jeweils nur vom Forschungsassistenten instruiert. (Kann aber gleichzeitig zu einer Verzerrung führen, da dies den Autorinnen dieser Arbeit praxisfern erscheint.) - Alle Rater hatten schon viel Erfahrung mit dem FLEE. <p>Bezüglich der Test-Retest-Reliabilität ist in der Studie kaum ersichtlich, welche Schritte unternommen wurden, um die interne Validität zu sichern. So sieht man keine Angaben bezüglich Raumtemperatur, Tageszeit der Testdurchführungen und der Stimmungslage der Patienten und Patientinnen. Es ist zudem nicht klar, ob die Patientinnen und Patienten ihre Ergebnisse der Sprungtests kannten. Dies könnte zu erheblichen Verzerrungen der Ergebnisse führen. In der Studie von Reid et al. (2007) wird zudem erwähnt, dass nach der ersten Testdurchführung ein entscheidender motorischer Lernprozess stattfindet. Aus diesem Grund wäre ein Design mit einer vorhergehenden Übungseinheit sicherlich angebracht gewesen.</p> <p>Die externe Validität betrifft die Population, auf die die Ergebnisse übertragen werden können. Dazu kann man aus dem Text herauslesen, dass die Resultate nicht auf verletzte Patienten zu übertragen sind, da die Studie nur mit gesunden Personen durchgeführt wurde.</p> <p>Stichprobe: Die Stichprobenziehung ist angebracht, wenn sie sich auf eine sehr sportliche Population bezieht. Die Population, auf die die Ergebnisse übertragen werden können, ist jedoch nicht erwähnt. Es handelt sich bei der Stichprobe um eine relativ homogene Population.</p> <p>Stichprobengrösse: Die Stichprobengrösse ist mit 40 (Drop-Outs bereits abgezogen) eher im unteren Bereich, aber okay. Es wurde keine Sample Size Calculation gemacht. Die gewählte Grösse wird nicht begründet. Die Drop-Out-Rate ist mit 20% hoch. Die Forschenden geben die Gründe der Drop-Outs an. Allerdings ist unklar, ob die Drop-Out-Rate die Ergebnisse stark beeinflusst oder nicht.</p> <p>Datenerhebung: Die Datenerhebung ist bezogen auf das Ziel nachvollziehbar. Der Ablauf der</p>	<p>Sind die Ergebnisse präzise? Die Ergebnisse werden nur in einem kurzen Text und sehr oberflächlich präsentiert. Die Studie dokumentiert lediglich wenige Resultate aus den Tabellen heraus. Es wird im Fliesstext nicht beschrieben, wie die einzelnen SEM und SRD Werte einzuschätzen sind.</p> <p>Entsprechen die Tabellen den folgenden Kriterien:</p> <p>Präzise und vollständig? Die Tabellen sind alle mit einem Titel versehen.</p> <p>Ergänzung zum Text? Die Tabellen sind zum besseren Verständnis und zur Übersicht hilfreich. Sie ergänzen den Text in dem Sinne, dass einige Werte darauf vermerkt sind, welche der Text nicht dokumentiert.</p>	<p>Werden alle Resultate diskutiert? Es werden alle Resultate, die im Text erwähnt wurden diskutiert.</p> <p>Stimmt die Interpretation mit den Resultaten überein? Augenscheinvalidität: Es wird lediglich eine einzige Vermutung bezüglich der grossen Differenz zwischen Wichtigkeit und Anwendungshäufigkeit geäussert. Erachtens der Autorinnen dieser Bachelorarbeit gäbe es noch andere Aspekte, die eine Rolle spielen könnten.</p> <p>Reliabilität: Die Resultate stimmen mit der Interpretation teilweise überein. Rein vom ICC-Wert her ist alles korrekt. Es wird von einem LEICHT erhöhten SEM-Wert des Triple Hop und Crossover Hop gegenüber anderer Studien gesprochen. Die Werte der Vergleichsstudien liegen bei: Tripple Hop: 11.17-23.18 cm vs. 25.1 cm Crossover Hop: 15.95-21.16 cm vs. 26.1 cm Dies scheint ein sehr grosser und nicht nur leichter Unterschied zu sein. Bei den Normalwerten stimmt die Interpretation mit den Resultaten überein.</p> <p>Werden die Resultate in Bezug auf die Fragestellung/Hypothesen, Konzepte und anderen Studien diskutiert und verglichen? Fast alle Resultate werden mit anderen Studien verglichen und meistens kurz in einem Satz diskutiert.</p> <p>Wird nach alternativen Erklärungen gesucht? Es wird nach alternativen Erklärungen gesucht. Allerdings wird nie nach mehreren Erklärungen gesucht. Bei Abweichungen liegt immer nur eine Variante vor, welche diese beschreibt.</p>

	<p>Datenerhebung wird genau beschrieben. Die Methoden der Datenerhebung sind bei allen Teilnehmenden gleich (Jeder Teilnehmende hat jeden Rater gleich oft). Alle Teilnehmenden werden für die Tests identisch instruiert. Dies bedeutet: Instruktion durch lediglich eine Person, sowie gleiche Zeitabstände zwischen den beiden Testtagen für alle. Man kann davon ausgehen, dass von allen 40 Teilnehmenden, alle Daten erhoben wurden, explizit erwähnt ist es jedoch nirgends.</p> <p>Messverfahren: Bei den Messinstrumenten mit welchen die Performance-Tests gemessen wurden, handelt es sich um zuverlässige Instrumente (Stoppuhr und Massband), die genaue Werte liefern. Beim Timed Lateral Step-Down wurde mittels subjektiver Parameter erhoben. Dort könnte man sich über die Reliabilität und Validität streiten. Dieser Test ist jedoch für unsere BA nicht relevant.</p> <p>Die Tests werden mehrheitlich gut definiert und beschrieben. Lediglich der LEFT wird ungenau und der Lateral Leap and Catch komplex erläutert. Es ist nicht eindeutig ersichtlich, was nun als Fehler gewertet wurde (nicht überquerte Linien oder fehlende Qualität oder beides?).</p> <p>Mögliche Verzerrungen/Einflüsse auf die Interventionen: Einflüsse wie Ermüdung und Lernfähigkeit werden erwähnt. Es wird jedoch nicht erwähnt, dass der Square Hop eine hohe Fehlerquelle haben könnte, da der Proband/die Probandin die Runden selbst laut zählen muss. Aus eigener Erfahrung ist es schwierig, dass beim gleichzeitigen Zählen der Test effizient und korrekt durchgeführt werden kann (schwierige Multitasking-Aufgabe).</p> <p>Datenanalyse: Die Verfahren werden beschrieben und die Ergebnisse durch Abbildungen veranschaulicht.</p> <p>Entsprechen die verwendeten statistischen Tests den Datenniveaus: Ja, die verwendeten statistischen Tests entsprechen dem Datenniveau. So werden für die Berechnung des ICC Intervallskalierte Daten vorausgesetzt, was in dieser Studie der Fall ist. Dass für die Interrater-Reliabilität der ICC(2.1) berechnet wurde, scheint sinnvoll, denn dieser bezieht den Rater mit ein. Warum für die Test-Retest-Reliabilität der ICC(3.1) und nicht der ICC(1.1) berechnet wurde, ist nicht ersichtlich. Es fehlt jegliche Begründung für die Verwendung der verschiedenen ICCs.</p> <p>Erlauben die statistischen Angaben eine Beurteilung: Aus einigen Tabellen und auch aus dem Text geht nicht eindeutig hervor, ob sich die Zahlen auf das dominante, nichtdominante oder auf beide Beine beziehen. Wie die ICC-Ergebnisse zu beurteilen sind ist klar definiert.</p> <p>Ethik: Es werden keine ethischen Fragen diskutiert.</p>		<p>Ist diese Studie sinnvoll? Werden Stärken und Schwächen aufgewogen? Die Studie ist bezüglich Validität, was für unsere Bachelorarbeit Relevanz hätte, wenig sinnvoll. So untersuchen die gestellten Fragen im Fragebogen nicht direkt die Validität, sondern mehr, ob der Test als sinnvoll betrachtet wird. Bezüglich Reliabilität ist die Studie sinnvoll. Allerdings beziehen sich die Werte nur auf eine gesunde, sehr sportliche und junge Zielpopulation. Der Teil über Normalwerte kann sinnvoll sein, sofern diese auf eine sehr sportliche Population angewendet werden. Es werden Limitationen diskutiert und somit einige Schwächen genannt.</p> <p>Wie und unter welchen Bedingungen sind die Ergebnisse in die Praxis umsetzbar? Für die Praxis sind hauptsächlich die Normwerte wertvoll, da nun für Sportlerinnen und Sportler, welche sich in der Rehabilitation befinden, Werte verschiedenster funktioneller Tests vorliegen. Die Ergebnisse der Reliabilität zeigen sowohl für die Test-Retest-, als auch für die Interrater-Reliabilität für sportliche, unverletzte Personen gute Resultate. Die ganze Testbatterie dauert 45 Minuten. Dies scheint uns zu lange und kann teilweise in der Praxis kaum umgesetzt werden.</p> <p>Wäre es möglich, diese Studie in einem anderen klinischen Setting zu wiederholen? Die Studie könnte gut wiederholt werden. Spannend wäre es, mit einer sehr heterogenen Stichprobe oder mit unsportlichen oder älteren Personen.</p>
--	--	--	---

Studie 4: Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. American Physical Therapy Association, 87(3), S. 337–349.

Zusammenfassung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Hintergrund: Die postoperative Behandlung von VKB-Patienten und -Patientinnen nimmt in der Physiotherapie einen grossen Stellenwert ein. Immer wieder stellt sich die Frage nach geeigneten und standardisierten Tests, um die Wirksamkeit einer Intervention zu überprüfen. Es gibt einige Fragebögen, die dies untersuchen. Allerdings seien diese Fragebögen nicht auf funktionelle Messungen angepasst. Verschiedene Autorinnen und Autoren unterstreichen die Wichtigkeit von funktionellen Tests nach einer operativen VKB-Rekonstruktion. Sprungtests werden von Forschenden immer wieder als gute funktionelle Tests vorgeschlagen, da sie die neuromuskuläre Kontrolle, die Kraft und das Vertrauen in die Extremität widerspiegeln und ein Minimum an Ausrüstung und Zeit benötigen. Eine Serie von folgenden vier Sprungtests scheint sehr geeignet zu sein, da sie eine Vielzahl von Bewegungsprinzipien enthalten. Die vier Tests sind: Front Hop, 6m Timed Hop, Triple Hop for distance und Crossover Hop for distance. Schon andere Autorinnen und Autoren untersuchten die</p>	<p>Design Studiendesign: prospektiv (= vorausblickend) und beobachtend mit wiederholten Messungen Aufbau der Studie: Die Probanden und Probandinnen führten die vier Sprungtests durch und füllten bei vier verschiedenen Testgelegenheiten Fragebögen aus. Den Patientinnen und Patienten wurde das Ergebnis der Sprungtests nicht gesagt. Die Tests wurden jeweils von der gleichen Person durchgeführt. Die drei ersten Testzeitpunkte erfolgten in der 16. Woche nach der VKB-Rekonstruktion mit einem Minimum von 24 Stunden zwischen den einzelnen Testdurchgängen. Die erste Durchführung war dazu gedacht, dass die Patientinnen und Patienten die Sprungtests motorisch lernen konnten. Die zweite und dritte Durchführung diente zur Bestimmung der Reliabilität. Die vierte Durchführung, die 6 Wochen später erfolgte, war indiziert, um die longitudinale Validität bestimmen zu können.</p> <p>Stichprobe Versuchspersonen:</p> <ul style="list-style-type: none"> - Von insgesamt 117 Patienten und Patientinnen wurden 48 Personen anhand der unten aufgeführten Einschluss- und Ausschlusskriterien ausgesucht. Es gab während der Durchführung sechs Drop-Outs, also nahmen insgesamt 42 Testpersonen an der Studie teil. - 35 Versuchspersonen beendeten den Test für die Reliabilität. - 39 Versuchspersonen nahmen am Test zur longitudinalen Validität teil. <p>Einschlusskriterien:</p> <ul style="list-style-type: none"> - 15-45 Jahre - unilaterale VKB-Rekonstruktion (Ersatz durch Semitendinosus- und Gracilis-Sehne). - keine Verletzung und kein chirurgischer Eingriff in den letzten 2 Jahren beim gesunden Bein - Full Range of Motion beim operierten Bein im Seitenvergleich - Keine Schwellung und kein Erguss mehr vorhanden - Patientinnen und Patienten, die zusätzlich zur VKB-Rekonstruktion den Meniskus reparieren mussten, wurden in die Studie eingeschlossen, vorausgesetzt, dass sie dasselbe Nachbehandlungsschema bestreiten durften, wie bei einer VKB-Rekonstruktion. Das heisst, sofortige Vollbelastung und bei null Belastung kein Bewegungslimit. <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> - Zusätzlicher HKB-Riss - Zusätzlicher Seitenbandriss 	<p>Ergebnisse: → Welche Ergebnisse werden präsentiert? → Welches sind die zentralen Ergebnisse der Studie? Absolute Werte ANOVA operiertes und nicht operiertes Bein: Für alle absoluten Sprungtestergebnisse (operiertes und nichtoperiertes Bein) zeigte die ANOVA einen signifikanten Haupteffekt bezüglich der Zeit ($P < 0.001$). Post-hoc-Test operiertes Bein: Bei allen Tests, die beim operierten Bein durchgeführt wurden, zeigte der Post-hoc-Vergleich, dass sich die absoluten Werte der ersten Testgelegenheit signifikant von der zweiten Testgelegenheit unterschieden ($P < 0.01$). Zwischen der zweiten und der dritten Testgelegenheit wurde jedoch kein signifikanter Unterschied bei den absoluten Werten festgestellt ($P > 0.89$). Mit Ausnahme des Timed Hop Tests ($P = 0.17$) wurden zwischen der zweiten und der vierten Testgelegenheit signifikante Unterschiede der absoluten Werte festgestellt ($P < 0.001$). Post-hoc-Test nicht operiertes Bein: Bei allen Tests des nicht operierten Beins zeigte der Post-hoc-Vergleich, dass sich die absoluten Sprungtestwerte bei</p>	<p>Diskussion und Interpretation der Ergebnisse: Limitationen: In der Diskussion dieser Studie werden keine Limitationen aufgezeigt. Allgemeine Erkenntnisse: Generell zeigen die Werte über die vier Testgelegenheiten verteilt, dass sowohl beim operierten, als auch beim nichtoperierten Bein ein motorischer Lernprozess zwischen der ersten und der zweiten Testgelegenheit stattfand, welcher bei der dritten Testgelegenheit abflachte. Bei der vierten Testgelegenheit waren die Werte des operierten Beines im Vergleich zum nichtoperierten Bein deutlich besser, was darauf hindeutet, dass sich der Funktionsstatus des operierten Beines im Laufe der sechswöchigen Periode verbesserte. Die Autorinnen und Autoren diskutieren, dass sowohl die Erfassung des LSI als auch der absoluten Werte sehr wichtig waren. Denn so zeigen sich die LSI-Werte bei den ersten drei Testgelegenheiten über alles hinweg als sehr stabil, was auf einen Lerneffekt in beiden Beinen hinweist. Zudem zeigten sich die LSI-Werte bei Frauen und Männern gleich. Das Erheben der absoluten Werte sei jedoch genauso wichtig. So waren die LSI-Werte der ersten und zweiten Testgelegenheit gleich, die absoluten Werte jedoch verschieden. Hätte man nur den LSI berechnet, wäre dies verborgen geblieben. Reliabilität: Die Autoren und Autorinnen beschreiben, dass die in dieser Studie vorliegenden ICCs</p>

<p>Reliabilität dieser Sprungtests anhand des ICC. Allerdings liegen keine Ergebnisse zum Messfehler (SEM) und zur kleinsten erkennbaren Veränderung (SDC) der Sprungtests vor. Um die vier Sprungtests als standardisierte, leistungsorientierte Testserie einsetzen zu können, müssen noch mehr Informationen über die Messeigenschaften ermittelt werden. Insbesondere zur Validität und Reliabilität.</p> <p>Ziel: Die Reliabilität und longitudinale Konstruktvalidität von vier Sprungtests, bei Patientinnen und Patienten während der Rehabilitation nach einer VKB-Rekonstruktion, zu erforschen.</p>	<ul style="list-style-type: none"> - Zusätzliche Verletzungen an Rücken, Hüfte oder Fuss - Fortgeschrittene degenerative Veränderungen - Personen, die nicht Englisch sprechen, schreiben und verstehen können <p>Ziehung der Stichprobe: Alle Patientinnen und Patienten die an der Studie teilnahmen wurden in der Fowler Kennedy Sport Medicine Clinic operiert. Es wird jedoch keine Aussage dazu gemacht, wie die Patienten und Patientinnen ausgewählt/ kontaktiert wurden. Es ist von einem Non-Probability Sampling auszugehen.</p> <p>Ermittlung der Stichprobengrösse: Es wurde eine Sample Size Calculation durchgeführt. Diese basierte auf der Einschätzung des Reliabilitäts-Koeffizienten für den durchschnittlichen LSI-Wert, bei einem unteren Konfidenzintervall (CI) von 0.1, einem erwarteten ICC von mindestens 0.85, und einem einseitigen CI basierend auf $1 - \alpha$ ($\alpha=0.05$). Anhand dieser Einschätzung liegt die ermittelte Stichprobengrösse bei 36 Testpersonen. Anhand der Annahme, dass verteilt über die 4 Erhebungspunkte 25% der Testpersonen aussteigen werden, wurden 50 Versuchspersonen benötigt.</p> <table border="1" data-bbox="481 671 1263 986"> <tr> <td data-bbox="481 671 875 815"> <p>Sprungtests: Getestet wurden folgende vier Sprungtests: Front Hop, 6m Timed Hop, Triple Hop for distance und Crossover Hop for distance.</p> </td> <td data-bbox="875 671 1263 815"> <p>Fragebögen: In der Studie wurden zwei Fragebögen («Lower Extremity Functional Scale» = «LEFS» und «Global Rating of Change») verwendet.</p> </td> </tr> <tr> <td data-bbox="481 815 875 986"> <p>Art der Datenerhebung (Sprungtests): Für die vier Sprungtests wurde die absolute Entfernung (cm) oder die absolute Zeit (Sek.) aufgezeichnet, welche im Seitenvergleich (LSI) beurteilt wurden.</p> </td> <td data-bbox="875 815 1263 986"> <p>Art der Datenerhebung (Fragebögen): Die Patientinnen und Patienten konnten in den Fragebögen verschiedene Punkte vergeben (siehe unten).</p> </td> </tr> </table> <p>Wie häufig wurden die Daten erhoben? Die Daten der Sprungtests und des «LEFS» Fragebogens wurden bei allen vier verschiedenen Testgelegenheiten erhoben. Der «Global Rating of Change» Fragebogen wurde nur bei der letzten Testgelegenheit erhoben.</p> <p>Messverfahren & Intervention Beschreibung der Intervention (Sprungtests): Auf eine genaue Beschreibung der durchgeführten Sprungtests wird an dieser Stelle verzichtet, da für diese Bachelorarbeit nur der Front Hop relevant ist. Dieser wurde in dieser Studie gleichermassen durchgeführt wie in der Definition dieser Bachelorarbeit.</p> <p>Durchführung der Sprungtests:</p> <ul style="list-style-type: none"> - Jeweils eine Runde zum Üben, zwei Durchläufe die gemessen wurden. - Es wurde kein zusätzliches Warm-Up gemacht. - Der Test wurde mit dem nicht-operierten Bein gestartet. - Um die Ermüdung zu reduzieren durften die Versuchspersonen 	<p>Sprungtests: Getestet wurden folgende vier Sprungtests: Front Hop, 6m Timed Hop, Triple Hop for distance und Crossover Hop for distance.</p>	<p>Fragebögen: In der Studie wurden zwei Fragebögen («Lower Extremity Functional Scale» = «LEFS» und «Global Rating of Change») verwendet.</p>	<p>Art der Datenerhebung (Sprungtests): Für die vier Sprungtests wurde die absolute Entfernung (cm) oder die absolute Zeit (Sek.) aufgezeichnet, welche im Seitenvergleich (LSI) beurteilt wurden.</p>	<p>Art der Datenerhebung (Fragebögen): Die Patientinnen und Patienten konnten in den Fragebögen verschiedene Punkte vergeben (siehe unten).</p>	<p>der ersten Testgelegenheit signifikant von den Werten der zweiten Testgelegenheit unterschieden ($P < 0.05$). Es gab keinen signifikanten Unterschied in den absoluten Werten der zweiten und dritten Testgelegenheit ($P>0.1$). Im Gegensatz zum operierten Bein gab es keine signifikanten Unterschiede zwischen den absoluten Werten der zweiten und vierten Testgelegenheit ($P>0.1$), mit Ausnahme des Crossover-Hop-Tests ($P=0.035$).</p> <p>LSI-Werte ANOVA: Bei den LSI-Werten (Limb Symmetry Index Werten) zeigte die ANOVA ebenfalls einen Haupteffekt bezüglich der Zeit ($P<0.001$), sowohl für jeden Sprungtest einzeln als auch für die Kombination der Tests (Overall LSI). Post-hoc-Test: Bei allen Tests zeigte der Post-hoc-Vergleich, dass sich der LSI der letzten Testgelegenheit signifikant von allen anderen Testgelegenheiten unterschied ($P<0.005$). Er zeigte jedoch keine signifikanten Unterschiede zwischen der ersten, zweiten und dritten Testgelegenheit ($P>0.40$).</p> <p>Zusammenfassend: Im Allgemeinen zeigte der Vergleich der Sprungtests, dass zwischen der ersten und der zweiten Testgelegenheit, sowohl beim operierten, als auch beim nicht-operierten Bein, ein</p>	<p>auf eine hervorragende relative Reliabilität der Sprungtests hinweisen. Diese Ergebnisse korrelieren ausserdem mit verschiedenen Studien, die bislang über die Reliabilität dieser Sprungtests gemacht wurden.</p> <p>SEM und SDC: In früheren Studien wurden keine Aussagen zum SEM gefunden. Die Ergebnisse dieser Studie fördern die Anwendung von Hop Tests mittels bereitstellen der SEM- und SDC-Werte insofern, als dass Fachpersonen erkennen können, wie viel Vertrauen sie in das Assessment eines individuellen Hop Tests bezüglich LSI setzen können. So kann der LSI lediglich durch den Messfehler um +/- 4.99% variieren. Bei wiederholten Messungen sogar um +/- 7.05%.</p> <p>Validität: Die Ergebnisse stimmen mit dem, was die Forschenden unter Validität verstehen überein. So zeigte sich bei allen Ergebnissen der Sprungtests über die Zeitspanne der sechs Wochen eine grössere Veränderung im operierten Bein als im nicht operierten Bein. Die konvergente Validität, also die Korrelation des LSI und der Veränderung des «LEFS», zeigte beim Single Leg Hop Test, beim Crossover Hop Test und bei der Kombination aller Sprungtest eine leicht konvergente Tendenz. Interessanterweise zeigte nur die Kombination aller Sprungtests eine Korrelation mit dem «global rating of change». Die Autoren spekulieren, dass dies so ist, weil eine Veränderung in der Kombination aller Tests eine Veränderung der motorischen Leistung globaler widerspiegelt, als ein Test alleine.</p> <p>Anwendung in der Praxis: Die Autoren weisen darauf hin, dass sie die</p>
<p>Sprungtests: Getestet wurden folgende vier Sprungtests: Front Hop, 6m Timed Hop, Triple Hop for distance und Crossover Hop for distance.</p>	<p>Fragebögen: In der Studie wurden zwei Fragebögen («Lower Extremity Functional Scale» = «LEFS» und «Global Rating of Change») verwendet.</p>						
<p>Art der Datenerhebung (Sprungtests): Für die vier Sprungtests wurde die absolute Entfernung (cm) oder die absolute Zeit (Sek.) aufgezeichnet, welche im Seitenvergleich (LSI) beurteilt wurden.</p>	<p>Art der Datenerhebung (Fragebögen): Die Patientinnen und Patienten konnten in den Fragebögen verschiedene Punkte vergeben (siehe unten).</p>						

	<p>zwischen jedem Durchgang max. 2 min Pause einlegen, falls dies gewünscht wurde.</p> <ul style="list-style-type: none"> - Die Arme durften beim Sprung eingesetzt werden. - Die Versuchspersonen trugen jenes Schuhwerk, welches sie normalerweise in der Rehabilitation tragen. - Die Landung musste mindestens 2 Sek. gehalten werden können. <p>Der Test galt als nicht bestanden, wenn:</p> <ul style="list-style-type: none"> - Arme oder das andere Bein den Boden berührten - Die Balance nicht gehalten werden konnte - Ein nachfolgender «Hüpfen» erfolgte <p>Wenn der Test nicht bestanden war, wurde er wiederholt. Häufig waren ein bis zwei Extrasprünge notwendig.</p> <p>Beschreibung der Fragebögen (Messinstrument): «Lower Extremity Functional Scale (LEFS)»: Fragebogen mit 20 Fragen, der zwischen 0 und 80 Punkten variiert. Je höher die Ergebnisse, desto höher der funktionelle Status des Patienten. Frühere Studien haben die Eigenschaften des LEFS untersucht. So beträgt der Standardmessfehler (SEM) 3.4-3.9 LEFS Punkte, 90% CI für eine gegebene Punktzahl (+/- 6 Punkte), die kleinste erkennbare Veränderung auf dem 90% Konfidenzintervall ist 9 Punkte und die kleinste klinische Differenz ist ebenfalls 9 Punkte. Bei der letzten Durchführung erhielten die Probanden auch noch einen «Global Rating of Change» Fragebogen. Dieser beinhaltet die Veränderung in den letzten 6 Wochen. Die Probanden konnten sagen, ob sie das Gefühl haben, besser oder schlechter geworden zu sein oder ob sie gleich geblieben sind. Die Veränderung konnten sie auf einer Skala von -7 bis 7 ankreuzen.</p> <p>Datenanalyse: Datenniveau der erhobenen Variablen: Sprungtests: <ul style="list-style-type: none"> - cm = proportionalskaliert - Sek. = proportionalskaliert Fragebögen: <ul style="list-style-type: none"> - Punktzahl «LEFS» = ordinalskaliert, wird aber intervallskaliert behandelt - Punktzahl «Global Rating of Change» = ordinalskaliert, wird aber intervallskaliert behandelt Datenanalyse: Alle Ergebnisse der Sprungtests wurden als absolute Entfernung (cm) oder Zeit (Sek.) aufgezeichnet. Es wurde jeweils der Mittelwert der zwei aufgezeichneten Versuche gebildet, aus dem der LSI berechnet wurde. Der LSI wurde für jeden Sprungtest einzeln, aber auch über alle Kombinationen, ausgerechnet. Die Ergebnisse der vier Sprungtests wurden jeweils anhand einer ANOVA (Varianzanalyse) verglichen. Es wurde jeweils eine separate ANOVA für das operierte und das nichtoperierte Bein durchgeführt. Einem signifikanten</p>	<p>erhebliches, motorisches Lernen stattfand, welches bei der dritten Test Gelegenheit abflachte. Die signifikante Steigerung der Sprungtestergebnisse des operierten Beines im Vergleich zum nicht operierten, zwischen dem dritten und vierten Testtag, deutet darauf hin, dass sich die Sprungleistung im Verlauf der sechs Wochen verbessert hat.</p> <p>Reliabilität: Die ICCs schwankten zwischen 0.82 bis 0.93 und können somit als hervorragende relative Reliabilität bezeichnet werden. Der Single Hop Test und der Overall LSI zeigten die höchste relative Reliabilität. Der SEM (Standard error of measurement) war beim Single Hop Test und beim Overall LSI am niedrigsten, was mit der höchsten relativen Reliabilität korreliert. In der Tabelle werden zudem Fehler des Individuellen LSI und die kleinste erkennbare Veränderung aufgezeigt. Die Besprechung erfolgt in der Diskussion</p> <p>Longitudinale Validität: LSI-Veränderungen: 6.5% (95%CI=4.5-8.5) für den Single Leg Hop Test 7.9% (95%CI=5.3-10.5) für den 6m Timed Hop Test 5.3% (95%CI=2.8-7.8) für den Triple Hop Test 4.8% (95%CI=2.2-7.4) für den Crossover Hop Test 6.1% (95%CI = 4.2-8.0) für die Overall Combination der</p>	<p>Reihenfolge der vier Sprungtests beibehalten haben und dass sie davon ausgehen, dass die Tests jeweils immer schwieriger werden. Sie meinen, beim Single Leg Hop Test sei die Reliabilität vermutlich gleich, wenn er einzeln durchgeführt wird, anstatt in der Batterie und vergleichen dies auch mit Ergebnissen aus einer anderen Studie. Sie raten jedoch davon ab, die anderen drei Tests separat durchzuführen, da sie von einem motorischen Lernzuwachs in den ersten Tests ausgehen. Für die Praxis raten sie, sich genügend Zeit einzuplanen, um mit den Patienten und Patientinnen eine Proberunde durchzuführen, da das motorische Erlernen der Tests eingerechnet werden sollte.</p> <p>Schlussfolgerung: Laut den Autorinnen und Autoren zeigen die Ergebnisse, dass die beschriebenen Hop Tests ein reliables und valides Ergebnis für Patienten und Patientinnen mit einer VKB-Rekonstruktion aufzeigen können. Das wiederum unterstützt und ermöglicht die Interpretation, dass die Sprungtests auch in der Forschung und in der Praxis eingesetzt werden sollten.</p> <p>Beantwortung der Forschungsfrage: Das Forschungsziel wird zu einem grossen Teil erreicht. So weisen die Ergebnisse auf eine gute Reliabilität hin. Auch die Konstruktvalidität, welche die Autoren und Autorinnen beschreiben, wird teilweise bestätigt. Allerdings stellt sich hier die Frage, auf was sich die Validität bezieht und ob dadurch eine wesentliche Aussage gemacht werden kann und warum sie nur bei zwei Tests eine positive Korrelation erzielte.</p>
--	--	---	--

	<p>Haupteffekt folgend, wurde der Scheffe Post-hoc-Test durchgeführt, um die Ergebnisse aller Tests zu vergleichen.</p> <p>Ermittlung der Reliabilität: Die Reliabilität wurde aus den Ergebnissen der Sprungtests von der zweiten und dritten Testgelegenheit berechnet. Die Reliabilität wurde anhand des ICC(2.1) und des SEM geschätzt. Die kleinste erkennbare Veränderung (SDC) wurde ermittelt, indem der SEM mit der Quadratwurzel von 2 und 1.64 (für 90% CI) multipliziert wurde. Die Autoren und Autorinnen weisen darauf hin, dass sie zwei verschiedene Konfidenzintervalle (IC) verwenden. Bei Punktschätzungen wurden 95% CI, bei Interpretationen der Punktzahl einer Person dagegen 90% CI gerechnet. Dies tun sie, weil sie der Meinung sind, dass klinische Interpretationen auf ein einzelnes Subjekt liberaler interpretiert werden sollen, als Interpretationen, die sich auf die gesamte Stichprobe abstützen.</p> <p>Ermittlung der Konstruktvalidität: Die Überprüfung auf die Konstruktvalidität basiert auf zwei Theorien:</p> <ol style="list-style-type: none"> 1. Veränderungen bei den Sprungtests sollten beim operierten Bein signifikant grösser sein, als beim nichtoperierten. 2. Für die Ermittlung der konvergenten Validität (Untergruppe der Konstruktvalidität) wurde die Theorie aufgestellt, dass Veränderungen im LSI mit Veränderungen bei den Fragebögen korrelieren sollten. <p>Die longitudinale Validität wurde anhand der Differenz zwischen der Testgelegenheit 4 und dem Mittelwert der Testgelegenheiten 2 und 3 berechnet. Die Validität bekannter Gruppen wurde anhand eines gepaarten t-Tests berechnet. Die konvergente Validität wurde einerseits anhand der Korrelation zwischen den Veränderungen der LSI-Werte und des LEFS, als auch zwischen dem LSI-Wert und dem Global Rating of Change Fragebogen ermittelt. Dazu wurde der Pearson-Korrelations-Koeffizient (r) und ein niedriges einseitiges 95% Konfidenzintervall berechnet. Da zuvor berichtete Korrelationen zwischen funktionellen Tests und Fragebögen typischerweise zwischen 0-0,6 liegen, wurde entschieden die Beweiskraft für die longitudinale Validität wie folgt festzulegen: gut = $r > 0.5$, mittel = $r > 0.36$, niedrig = $r > 0.2-0.35$ und keinen Beweis $r < 0.2$.</p> <p>Ethik Von den Forschenden werden keine ethischen Fragen thematisiert. Die Studienteilnehmenden haben ihre schriftliche Einwilligung gegeben.</p>	<p>Sprungtests</p> <p>Die Änderungen der absoluten Punktzahl der Sprungtests beim operierten Bein sind statistisch grösser, als die Änderungen beim nicht operierten:</p> <p>Single Hop Test (paired t = 6.4, $P < 0.001$) 6m Timed Hop Test (paired t = 4.5, $P < 0.001$) Triple Hop Test (paired t = 3.3, $P = 0.002$) Crossover Hop Test (paired t = 3.1, $P = 0.004$)</p> <p>Korrelationen (r) zwischen den funktionellen Tests und den Fragebögen schwanken zwischen 0.26 bis 0.58. Der «Global Rating of Change» korreliert meistens mit dem Overall LSI.</p> <p>Die Ergebnisse sind in Textform und anhand von ergänzenden Tabellen präsentiert. Allerdings sind die Tabellen nicht selbsterklärend und es bedurfte einiger Zeit, bis sie richtig verstanden und interpretiert werden konnten.</p>	
--	--	--	--

Würdigung der Studie:

Einleitung	Methode	Ergebnisse	Diskussion
<p>Die Relevanz der Fragestellung dieser Bachelorarbeit, sowie für die Berufspraxis, ist nach Ansicht der Autorinnen dieser Arbeit gegeben. Allerdings ist sie nicht in der Lage, die Frage nach der Validität des Front Hop Tests vollständig zu beantworten.</p> <p>Die Studie beinhaltet weder eine konkrete Fragestellung noch eine Hypothese. Lediglich der Hintergrund und das Ziel der Studie ist klar definiert. Das Thema und die Problemstellung werden logisch hergeleitet und anhand verschiedener Literatur dargestellt.</p>	<p>Design: Die Verbindung zwischen der Forschungsfrage und dem gewählten Design ist logisch und nachvollziehbar. Die interne Validität wird durch folgende Schritte kontrolliert:</p> <ul style="list-style-type: none"> - Die Versuchspersonen erfuhren die Ergebnisse ihrer Tests nicht. - Die Testdurchführung war bei allen Erhebungen gleich und wurde durch die gleiche prüfende Person durchgeführt. - Um eine Ermüdung zu verhindern, durften die Patientinnen und Patienten, sofern gewünscht, zwischen den verschiedenen Tests, respektive zwischen den einzelnen Testdurchführungen, eine 2-minütige Pause einlegen. Laut den Autoren und Autorinnen der Studie wurden typischerweise weniger als 30 Sekunden Pause benötigt. Hierbei kann man sich fragen, ob es sinnvoll gewesen wäre, eine obligatorische Pause von 2min einzulegen, damit jeder Proband und jede Probandin auch tatsächlich «erholt» war. Die Einschätzung von Ermüdung ist sehr subjektiv. <p>Zur Raumtemperatur, Tageszeit der Testdurchführungen und zur Stimmungslage der Patienten wird nichts erwähnt.</p> <p>Stichprobe: In der Studie wird nicht erwähnt auf welchem Weg die Versuchspersonen kontaktiert wurden. Es ist von einem Non-Probability Sampling auszugehen. Die Stichprobe ist im Allgemeinen gut beschrieben. Es wird jedoch keine Aussage gemacht, auf welche Population sich die Stichprobe genau bezieht. Es kann angenommen werden, dass sich die Ergebnisse, unter Berücksichtigung anderer Einflussfaktoren, auf Personen zwischen 15-45 Jahren, mit einer unilateralen VKB-Rekonstruktion (Ersatz durch Semitendinosus- und Gracilis-Sehne) übertragen lassen können. Da die Stichprobenziehung nicht genau beschrieben wurde, könnte es jedoch sein, dass sich etwas weniger engagierte Personen nicht für die Studie gemeldet haben und dass sich somit die Ergebnisse nicht auf alle Personen mit einer unilateralen VKB-Rekonstruktion übertragen lassen. Die Stichprobengrösse wurde durch eine Sample Size Calculation ermittelt. Anhand dieser wurde eine Stichprobengrösse von 50 Patientinnen und Patienten ermittelt. Schlussendlich nahmen für die Ermittlung Reliabilität 35 und für die der Validität 39 Testpersonen teil. Das bedeutet, die ermittelte Grösse wurde nicht ganz eingehalten. Die Studie wurde durch einige Drop-Outs beeinflusst. Diese werden von den Autoren und Autorinnen angegeben und begründet.</p> <p>Datenerhebung: Die Datenerhebung ist für die Zielsetzung der Reliabilität und der Konstruktvalidität, nachvollziehbar.</p>	<p>Die Ergebnisse werden mehrheitlich präzise beschrieben. Es werden Tabellen als Ergänzung zum Text präsentiert. Allerdings sind vor allem die Ergebnisse der longitudinalen Validität für jemanden, der nicht Statistik studiert hat, schwer nachvollziehbar. Zudem wäre es spannend gewesen, die ICCs, den SEM und die SDCs für jedes Bein einzeln zu sehen. Aus Tabellen und Text ist nicht ersichtlich, von welchen Beinen (operiert oder nicht operiert) die Daten stammen.</p>	<p>Diskussion und Interpretation der Ergebnisse: Ansatzweise werden alle Ergebnisse diskutiert. Die Diskussion bezüglich der Validität und was dies nun für die Praxis aussagt, fällt leider etwas kurz aus. Die Autoren und Autorinnen beschreiben, dass die Studie beweist, dass es sich bei den Sprungtests um reliable und valide Messinstrumente handelt. Bezüglich der Validität kommen jedoch Zweifel auf. Sagt eine geringe Korrelation des LSI mit dem «LEFS» wirklich etwas über die Validität aus? Ist es nicht selbsterklärend, dass ein besserer LSI auch auf einen besseren funktionellen Status hinweist? Vor allem die Resultate der Reliabilität werden mit anderen Studien verglichen. Es fehlt jedoch die kritische Auseinandersetzung mit dem Konzept, bzw. den Hypothesen bezüglich der Validität. Leider wird in der Studie zu wenig nach alternativen Erklärungen gesucht.</p> <p>Schlussfolgerung Anwendung und Verwertung in der Pflegepraxis: Für die Ermittlung der Reliabilität, des Standardmessfehlers und der kleinsten erkennbaren Veränderung ist diese Studie sicherlich sinnvoll. Für die Ermittlung der Validität zeigt sie noch einige Schwachstellen auf (siehe oben). Die Stärken und Schwächen der Studie werden nicht diskutiert, was es zu bemängeln gibt. In der Praxis können die beschriebenen Sprungtests als ein reliables Messinstrument eingesetzt werden. Es kann auch davon ausgegangen werden, dass ein besserer LSI mit einem besseren funktionellen Status korreliert.</p>

	<p>Die Methoden der Datenerhebung wurden bei allen Teilnehmenden gleich durchgeführt. Die Daten wurden nicht bei allen Teilnehmenden erhoben, da es während der Studie einige Drop-Outs gab, diese werden jedoch von den Autorinnen und Autoren thematisiert. Einzig die freiwillige Pause zwischen den Testsequenzen war nicht standardisiert, da jeder Teilnehmer selbst die Dauer der Pause entscheiden durfte.</p> <p>Messverfahren: Die Autoren und Autorinnen weisen darauf hin, dass der «LEFS» Fragebogen ein bereits geprüfetes Messinstrument sei. Warum die Autoren und Autorinnen genau diesen Fragebogen verwenden wird in der Studie nicht diskutiert. Es wird auch nicht darauf eingegangen, warum angenommen wird, dass wenn der LSI und die Punktzahl des «LEFS» korrelieren, eine höhere Validität der Sprungtests angenommen werden kann. Relativ viele Fragen aus dem «LEFS» sagen unserer Meinung nach nichts über den Front Hop aus. So kann eine Person Mühe mit dem Front Hop Test bekunden, erreicht aber auf dem «LEFS» doch noch eine gute Punktzahl. Mögliche Einflüsse, beispielsweise eine schlechte psychische Verfassung der Testperson, welche sich negativ auf den «LEFS» oder die Sprungtests auswirken könnten, werden nicht diskutiert.</p> <p>Datenanalyse: Die Verfahren der Datenanalyse werden klar beschrieben. Die statistischen Verfahren werden sinnvoll und nachvollziehbar angewendet. Streng genommen entsprechen die angewendeten statistischen Tests jedoch nicht dem Datenniveau. So weisen die Fragebögen ordinalskalierte Daten auf, mit denen keine metrischen Verfahren wie z.B. eine Varianzanalyse durchgeführt werden dürften. In der Forschung werden allerdings häufig Fragebögen als intervallskalierte Daten verwendet, die im Gegensatz zu ordinalskalierten Daten metrische Verfahren erlauben. Zudem stellt sich die Frage, warum der ICC(2.1) berechnet wurde. In der Studie fehlt jegliche Begründung dafür. So wäre für die Test-Retest-Reliabilität die Berechnung des ICC(1.1) oder (3.1) angebracht gewesen, da diese den Rater nicht einbeziehen. Der ICC(2.1) ist der konservativste von allen, da er im Nenner noch den Rater als Fehlerquelle miteinrechnet. Die statistischen Angaben ermöglichen eine Beurteilung. Es werden klare, nachvollziehbare Angaben gemacht. Die Ermittlung des Standardmessfehlers und der kleinsten erkennbaren Veränderung sind sehr spannende, zusätzliche Parameter. Die Höhe des Signifikanzniveaus ist zwar nachvollziehbar, wird jedoch von den Autoren und Autorinnen nicht begründet. Diese erörtern lediglich, warum sie verschiedene Konfidenzintervalle verwendet haben.</p> <p>Ethik: Ethische Fragen und entsprechende Massnahmen wurden in der Studie nicht diskutiert. Es ist zum Beispiel nicht klar, in welcher Beziehung die Forschenden und Teilnehmenden zueinander stehen.</p>		<p>Würde diese Studie wiederholt, wäre es sinnvoll sich noch mehr Gedanken zur Validität zu machen und diese mit einem anderen Ansatz zu überprüfen.</p>
--	--	--	--