



Stellenwert der Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL) in der sozialmedizinischen Leistungsbeurteilung



Laufzeit:

01.07.2004 – 30.06.2006

Förderung:

Gefördert durch den Verein für Rehabilitationsforschung e.V., Norderney

Unterstützt durch den EFL-Bundesverband

Assoziiert mit dem NRW-Forschungsverbund Rehabilitationswissenschaften

Durchführung:

Institut für Rehabilitationsforschung e.V., Norderney; Abteilung Bad Rothenfelde

Auf der Stöwwe 11

49214 Bad Rothenfelde

Abteilungsleiter: Prof. Dr. Bernhard Greitemann, Dipl.-Oec.

Projektleiterin: Claudia Niemeyer (während der Projektlaufzeit noch Büschel), Dipl.-Psych.

Dokumentationsassistentin: Monika Schaidhammer, M.A.

Kooperationspartner:

Klinik Münsterland, Bad Rothenfelde

(Ärztliche Leitung: Prof. Dr. Bernhard Greitemann)

Brunswiek-Klinik, Klinik Friedrichshöhe, Bad Pyrmont

(Ärztliche Leitung: Dr. Detlev Kasprowski)

Kontakt:

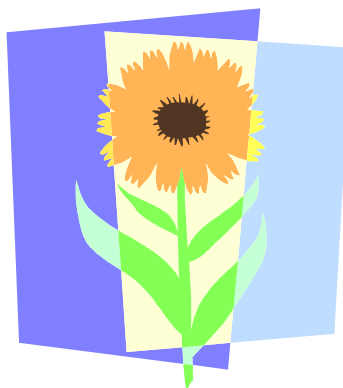
Dipl.-Psych. Claudia Niemeyer

Tel.: 05424/220-472

Fax: 05424/220-345

c.niemeyer.ifr@klinik-muensterland.de

Danksagung



Wir danken
dem Verein zur Förderung der Rehabilitationsforschung e.V., Norderney
für die Finanzierung des Projektes,
dem EFL-Bundesverband für seine Unterstützung
und allen beteiligten Mitarbeitern und Patienten
für ihre engagierte Mitarbeit!

Inhalt

Danksagung

Inhalt

Verzeichnis der Tabellen und Abbildungen

1. Einleitung	1
2. Die Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL)	3
2.1 Entwicklung	3
2.2 Einsatzbereiche	5
2.3 Konzeption, Aufbau und Durchführung.....	8
Konzeption	8
Aufbau und Durchführung	9
Indikationen und Kontraindikationen	11
3. Stand der Forschung: Testgüte der EFL	13
3.1 Sicherheit des Probanden in der EFL.....	14
3.2 Unabhängigkeit des EFL-Ergebnisses vom Untersucher – Objektivität.....	15
3.3 Zuverlässigkeit des EFL-Ergebnisses – Reliabilität	16
3.4 Aussagekraft des EFL-Ergebnisses – Validität.....	20
Face validity	20
Inhaltsvalidität	21
Konstruktvalidität.....	23
Kriterienorientierte Validität	27
3.5 Praktikabilität	29
3.6 Nützlichkeit	31
3.7 Zusammenfassung: Testgüte der EFL nach dem augenblicklichen Stand der Literatur	35

4. Ziele und Fragestellungen der Studie	39
4.1 Nutzen für die ärztliche Einschätzung.....	39
4.2 Nutzen für den Patienten.....	39
4.3 Intangible Kosten: Beschwerdezunahme durch EFL?	40
5. Methode	41
5.1 Design	41
5.2 Instrumente.....	41
EFL-übliche Instrumente	42
Weitere Bögen für Patienten	42
Bögen für Ärzte	43
Bogen für EFL-Therapeuten.....	43
5.3 Ablauf	44
5.4 Probanden	45
6. Datenbasis	46
6.1 Fehlende Daten	46
6.2 Soziodemographische Daten.....	47
6.3 Ausbildung und Beruf	48
6.4 Gesundheit	52
6.5 Leistungsbereitschaft.....	57
7. Frage 1: Inwiefern entsprechen die sozialmedizinischen Beurteilungen der behandelnden Ärzte den Ergebnissen der EFL? In welchem Maße und an welchen Punkten unterscheiden sie sich? Wovon hängen mögliche Unterschiede ab?	59
7.1 Design und Instrumente.....	59
7.2 Ergebnisse.....	60
Allgemeine Leistungsfähigkeit nach REFA.....	60
Subtests der EFL – Einzelfähigkeiten.....	62
7.3 Diskussion	65
Erklärungsversuch 1: Im Gegensatz zu den Ärzten berücksichtigt die EFL keine weiteren Belastungen.	66

Erklärungsversuch 2: Im Gegensatz zu den Ärzten berücksichtigt die EFL keine Entwicklungen und Prognosen	70
Erklärungsversuch 3: Patientenbezogene Prädiktoren.....	73
Soziodemographische Daten	73
Berufliche Situation	74
Eigene Einschätzung der beruflichen Fähigkeiten und berufliche Perspektive	74
Gesundheit.....	75
EFL-Erwartungen und -Befürchtungen	79
Leistungsniveau der Probanden	80
Erklärungsversuch 4: Die Ärzte lassen sich durch die Persönlichkeit des Patienten beeinflussen, die EFL nicht	83
Erklärungsversuch 5: Die Extrapolation der Testergebnisse auf eine Dauerbelastung ist fragwürdig.....	83
Exkurs: Ein Proband – zwei Ärzte	84
7.4 Zusammenfassung zu Frage 1	86
7.5 Grenzen der Studie für die Beantwortung dieser Frage.....	88
7.6 Fazit zu Frage 1	89
8. Frage 2: Trägt die EFL dazu bei, dass Patienten ein realistischeres Bild von ihrer eigenen funktionellen Leistungsfähigkeit bekommen? Wenn ja, hat das Auswirkungen auf ihre beruflichen Zielsetzungen?	90
8.1 Design und Instrumente.....	90
8.2 Ergebnisse.....	91
8.3 Diskussion	98
8.4 Zusammenfassung zu Frage 2	99
8.5 Grenzen der Studie für die Beantwortung dieser Frage.....	100
8.6 Fazit zu Frage 2.....	100
9. Frage 3: Führt die EFL zu einer erheblichen Zunahme der Beschwerden?	101
9.1 Probanden, die den EFL-Test aufgrund einer erheblichen Zunahme von Beschwerden abgebrochen haben	
Vp 0102 (Männlich, 40 Jahre)	101

Vp 1114 (Männlich, 44 Jahre)	102
Vp 1120 (Weiblich, 38 Jahre)	103
Bewertung der medizinischen Drop-Outs	103
9.2 Quantitative Analyse der verbleibenden 73 Probanden - Design und Instrumente	106
9.3 Ergebnisse.....	107
Schmerzentwicklung während der EFL	107
Prädiktoren: Soziodemographische Variablen	109
Prädiktoren: Beruf	109
Prädiktoren: Gesundheit.....	110
Prädiktoren: Schmerz.....	112
Prädiktoren: Leistungsbereitschaft	113
Prädiktoren: Test-Erwartungen und -Befürchtungen	114
Prädiktoren: Leistungsniveau	114
9.4 Diskussion	115
9.5 Zusammenfassung zu Frage 3	117
9.6 Grenzen der Studie für die Beantwortung dieser Frage.....	118
9.7 Fazit zu Frage 3.....	118
10. Allgemeine Diskussion der Studie	119
11. Zusammenfassung der Studie	121
12. Bedeutung der Ergebnisse für die Praxis – Fazit	124
13. Literatur	125
14. Eigene Veröffentlichungen und Präsentationen der Studie	130

Verzeichnis der Tabellen und Abbildungen

Tabellen:

Tabelle 1: Schlüsselemente arbeitsbezogener Anforderungen im D.O.T und der zu testenden Fähigkeiten in der EFL nach Isernhagen	4
Tabelle 2: Reihenfolge der EFL-Subtests an den beiden Testtagen	10
Tabelle 3: Überblick der Interessengemeinschaft Physiotherapie Rehabilitation (igptr) über verschiedene Studien zur Reliabilität der EFL	18
Tabelle 4: Codierungskategorien für funktionelle Leistungsfähigkeit	23
Tabelle 5: Vergleich der EFL-Resultate von Probanden aus verschiedenen Heimatländern	31
Tabelle 6: Geschlechterverteilung an den Erhebungsorten	47
Tabelle 7: Altersdurchschnitte in Abhängigkeit von Geschlecht und Ort	47
Tabelle 8: Zeiträume zwischen Krankheitsbeginn oder Unfall und EFL im Ortsvergleich	56
Tabelle 9: Codierung der Kategorien der allgemeinen Leistungsfähigkeit in EFL-Bericht und Arztbogen	61
Tabelle 10: Übereinstimmung von EFL und Arzt in den dichotom codierten Variablen	65
Tabelle 11: Zusammenhänge zwischen dem Wissen der Ärzte um Zusatzbelastungen der Patienten an ihren Arbeitsplätzen und der Abweichung ihrer Einschätzung vom EFL-Ergebnis	68
Tabelle 12: Zusammenhänge zwischen der Erfahrung der Ärzte und der Abweichung ihrer Einschätzung vom EFL-Ergebnis	72
Tabelle 13: Abweichungen der ärztlichen Einschätzungen vom EFL-Ergebnis im Ortsvergleich	73
Tabelle 14: Abweichungen der ärztlichen Einschätzungen vom EFL-Ergebnis in Abhängigkeit von verschiedenen Risikofaktoren der Patienten	76
Tabelle 15: Mittlere Differenzen zwischen den EFL-Ergebnissen und den ärztlichen Einschätzungen in Abhängigkeit von der Lokalisation des arbeitsbezogenen Problems.	77
Tabelle 16: Signifikante Zusammenhänge zwischen Angaben zu Schmerz in den letzten 7 Tagen und Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis	78
Tabelle 17: Signifikante Zusammenhänge zwischen Angaben zu Schmerz zu Beginn und Ende der EFL-Testtage und Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis	78
Tabelle 18: Kreuztabelle: Maximale Arbeitsschwere ohne Gesundheitsgefährdung vor und nach dem EFL-Test (Selbsteinschätzung der Probanden).....	97

Tabelle 19: Kreuztabelle: Vermuteter Zeitpunkt der Rückkehr ins Arbeitsleben vor und nach dem EFL-Test (Selbsteinschätzung der Probanden)	97
---	----

Abbildungen:

Abbildung 1: Teufelskreis der Beeinträchtigung beruflicher Teilhabe bei psychosomatischen Patienten	7
Abbildung 2: Ausschnitt aus einer EFL-Tabelle	10
Abbildung 3: EFL-Projekt: Design und Instrumente	41
Abbildung 4: Schulabschluss	48
Abbildung 5: Berufsausbildung	49
Abbildung 6: Berufliche Situation	50
Abbildung 7: Dauer der Arbeitsunfähigkeit der zum EFL-Test arbeitsunfähigen Probanden	51
Abbildung 8: Dauer der Arbeitsunfähigkeit vor dem EFL-Test in 6-Monatskategorien im Ortsvergleich	52
Abbildung 9: Risikofaktoren	53
Abbildung 10: Arbeitsbezogene Probleme im Ortsvergleich	54
Abbildung 11: Schlimmster Schmerz in den letzten 7 Tagen im Ortsvergleich	55
Abbildung 12: Schmerz in den letzten 7 Tagen im Ortsvergleich	55
Abbildung 13: Zuverlässigkeit der Leistungsbereitschaft nach Einschätzung der EFL-Therapeuten	57
Abbildung 14: AVEM: Zuordnung der Probanden zu Mustern	58
Abbildung 15: Frage 1: Unterschiede zwischen EFL-Ergebnis und ärztlicher Einschätzung – Instrumente	60
Abbildung 16: Allgemeine Leistungsfähigkeit – Abweichung des ärztlichen Urteils vom EFL-Ergebnis	61
Abbildung 17: Vergleich der EFL-Ergebnisse mit den ärztlichen Einschätzungen: Mittlere Anzahl Belastbarkeitskategorien, die das EFL-Ergebnis die ärztliche Einschätzung überstieg	63
Abbildung 18: Zusammenhänge zwischen EFL-Ergebnis und ärztlicher Übereinstimmung	64
Abbildung 19: Ärztliche Angaben über körperliche Zusatzbelastungen der Patienten an ihren Arbeitsplätzen	67
Abbildung 20: Ärztliche Angaben über psychische Zusatzbelastungen der Patienten an ihren Arbeitsplätzen	67
Abbildung 21: Vergleich der Angaben von Ärzten und Patienten über Zusatzbelastungen der Patienten an ihren Arbeitsplätzen	69
Abbildung 22: Mittlere Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis - Vergleich von Probanden mit niedrigen und hohen Leistungswerten in der EFL in den einzelnen Subtests	81

Abbildung 23: Mittlere Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis – Vergleich von Probanden mit niedrigen und hohen Leistungswerten laut ärztlicher Einschätzung in den einzelnen Subtests.....	82
Abbildung 24: Vergleich der Einschätzungen von EFL, Ober- und Stationsarzt bei zwei Patienten	85
Abbildung 25: Frage 2: Veränderung der Selbsteinschätzung der Probanden – Instrumente.....	91
Abbildung 26: Veränderung der eigenen Einschätzung der Leistungsfähigkeit der Probanden.....	92
Abbildung 27: Veränderungen im PACT.....	93
Abbildung 28: Selbsteinschätzung der Leistungsfähigkeit der Patienten im PACT im Vergleich zu ihrem EFL-Ergebnis.....	93
Abbildung 29: EFL-Testerkenntnisse der Probanden	94
Abbildung 30: Veränderung von Selbsteinschätzung und beruflichen Perspektiven der Probanden	95
Abbildung 31: Veränderungen der Selbsteinschätzung bezüglich der maximalen Schwere der Arbeit ohne Gesundheitsgefährdung und des vermuteten Rückkehrzeitpunktes in das Arbeitsleben	96
Abbildung 32: Bewegungskoordination - Vergleich der medizinischen Drop-Outs mit den Probanden, die die EFL zu Ende geführt haben	105
Abbildung 33: Kenntnisse über ergonomische Arbeitstechniken - Vergleich der medizinischen Drop-Outs mit den Probanden, die die EFL zu Ende geführt haben.....	105
Abbildung 34: EFL-Testerwartung: Nachweis der uneingeschränkten Leistungsfähigkeit - Vergleich der medizinischen Dropouts mit den Probanden, die die EFL zu Ende geführt haben	105
Abbildung 35: Frage 3: Beschwerdezunahme während EFL – Instrumente	107
Abbildung 36: Schmerzentwicklung während EFL.....	108
Abbildung 37: Profil eines Risikopatienten für erhebliche Schmerzzunahme während der EFL	117

1. Einleitung

Manchmal passiert es ganz plötzlich, manchmal ist es ein jahrelanger schleichender Prozess: Menschen, die mitten im Leben stehen, werden durch Unfall oder Krankheit so schwer geschädigt, dass sie nicht mehr in der Lage sind, ihren Beruf auszuüben und ihren eigenen Lebensunterhalt zu sichern. Die Deutsche Rentenversicherung beziffert die Zahl der Betroffenen auf etwa eine Million Menschen im Jahr (Deutsche Rentenversicherung, 2006).

Für diese Menschen hält die Deutsche Rentenversicherung ein breites Angebot hochwertiger rehabilitativer Maßnahmen bereit, die ihnen helfen können, wieder in ein normales Leben zurückzukehren. Den medizinischen ebenso wie den eher berufsorientierten Leistungen liegt dabei die International Classification of Functioning, Disability and Health (ICF) zugrunde, nach der die funktionelle Gesundheit eines Menschen das Ziel dieser Maßnahmen ist. Geschädigte Körperstrukturen sollen so behandelt werden, dass Funktionsstörungen überwunden oder kompensiert und Folgeprobleme in den Bereichen der Fähigkeiten, der Aktivitäten und der Teilhabe vermieden werden. Dabei spielt die Wiedereingliederung in den Arbeitsprozess nicht nur aus Kostengründen, sondern auch im Hinblick auf neuere Erkenntnisse über die Bedeutung der Erwerbsarbeit für die biopsychosoziale Gesundheit des Menschen eine entscheidende Rolle. Wie ernst diese Aufgabe zu nehmen ist, wird besonders deutlich in dem gesetzlichen Auftrag, den die Rentenversicherungen dazu übertragen bekommen haben.

Auf diesem Hintergrund erklärt sich der hohe Stellenwert, der gutachterlichen Tätigkeiten auch innerhalb der medizinischen Rehabilitation zugeschrieben wird – zu den zentralen Aufgaben der Reha-Ärzte gehört die sozialmedizinische Einschätzung eines jeden Rehabilitanden der Rentenversicherung. Diese setzt neben guten Kenntnissen der individuellen Belastungen des Rehabilitanden an seinem Arbeitsplatz eine zutreffende Bestimmung seiner funktionellen Fähigkeiten voraus.

Dabei wird der einfache Schluss von einer Diagnose auf die Erwerbsfähigkeit ebenso als unzulässig angezweifelt wie die in der Orthopädie üblichen klinischen Verfahren zur Messung körperlicher Parameter (Neutral-Null-Methode, Schober-/Ott-Indices, Finger-Boden-Abstand) oder die reine Auswertung bildgebender Verfahren (Abdel-Moty et al., 1993; Kraus, 1997; Die Kölnische Rück, 1999; SOMEKO, 2004). Selbstverständlich sind diese Verfahren als solche absolut berechtigt. So ist beispielsweise der Finger-Boden-Abstand ein hervorragendes Maß für die Flexionsfähigkeit der Wirbelsäule. Er sagt jedoch, ebenso wie die anderen klinischen Maße, kaum etwas über die Eignung eines Probanden aus, den Belastungen an einem bestimmten Arbeitsplatz standzuhalten. Dafür sind in der Regel erheblich komplexere Fähigkeiten nötig wie etwa Heben, Tragen, Stehen oder Sitzen, so dass isolierte Indices nur wenig zu einer treffenden sozialmedizinischen Prognose beitragen können. Die wesentliche Aufgabe des ärztlichen Gutachters ist es deshalb,

aus den Puzzleteilchen der klinischen Untersuchung, der Wertung diagnostischer Verfahren und der Verlaufsbeurteilung anhand seines Wissens über Arbeits- und Umgebungsbedingungen eine valide Einschätzung der Leistungsfähigkeit des Patienten zu erreichen. Wenn möglich, sollte diese durch objektivierbare Parameter untermauert werden.

Seit Jahren wird deshalb an der Entwicklung von Assessments gearbeitet, mit der die arbeitsbezogenen Fähigkeiten von Patienten ebenso strukturiert und standardisiert wie zuverlässig, aussagekräftig und risikoarm ermittelt werden können. State of the art sind zur Zeit die so genannten Functional Capacity Evaluations („FCEs“), die unter Zuhilfenahme spezieller Apparaturen eine objektive Messung der elementaren arbeitsbezogenen Fähigkeiten des Patienten erlauben sollen. Das in Deutschland am häufigsten eingesetzte FCE-System ist die Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL), die inzwischen in 101 Einrichtungen in Deutschland, unter anderem der Klinik Münsterland und dem Reha-Zentrum Bad Pyrmont, durchgeführt wird (EFL-Akademie, 2007).

Die EFL ist in den USA entwickelt worden, aber inzwischen auch in Europa hoch anerkannt. Themengebundene Veranstaltungen¹ und diverse Beiträge auf medizinischen und rehabilitationswissenschaftlichen Tagungen der letzten Jahre belegen, dass FCE-Verfahren als wichtige – zum Teil sogar als einzig gültige – Instrumente für Begutachtungen wahrgenommen werden. Mit Sorge beobachten wir dabei, dass der wissenschaftliche Anspruch und der wissenschaftliche Eindruck von Assessmentverfahren wie der EFL offenbar verführerisch wirken: In der Literatur wird immer wieder betont, dass FCEs aufgrund ihrer methodischen Grenzen in der sozialmedizinischen Einschätzung nur ein Baustein unter mehreren sein können (Kaiser et al., 2000; Gatty, 2002; Erbstöber et al., 2003). Trotzdem scheint ihnen in der Praxis häufig eine überragende Stellung zugestanden zu werden, da ihre Ergebnisse ohne ausreichenden Abgleich mit anderen Elementen der Begutachtung wie der klinischen Untersuchung und Beobachtung als Basis für Entscheidungen herangezogen werden.

Im Folgenden sollen in diesem Projektbericht auf der Basis der Auseinandersetzung mit der Literatur eigene Ergebnisse zum Stellenwert der EFL in der sozialmedizinischen Leistungsbeurteilung stationärer orthopädischer Reha-Patienten dargestellt und in die Kontexte von Forschung und Anwendung eingebettet werden.

¹ Zum Beispiel die gut besuchte Tagung des Instituts für Qualität in Prävention und Rehabilitation (IQPR) „Dialog zwischen Richtern und Sachverständigen zum Thema: Qualitätsstandards der Begutachtung – Möglichkeiten und Nutzen des Einsatzes von Profilvergleichsverfahren“ im Dezember 2004 in Kassel

2. Die Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL)

2.1 Entwicklung

Auf Anfrage einer Versicherung entwickelte die Physiotherapeutin Susan Isernhagen in den 1980er Jahren in den USA ihr System zur Evaluation der funktionellen Leistungsfähigkeit (EFL) (Kaiser et al., 2000). Ziel war es, detaillierte Aussagen über die ergonomisch sichere Belastbarkeit eines Menschen für häufige arbeitsbezogene physische Anforderungen treffen zu können. Die EFL sollte bei der Auswahl von Bewerbern für eine Stelle ebenso helfen wie bei der sozialmedizinischen Beurteilung im klinischen Alltag und vor Gericht. Bei der Festlegung der einzelnen zu testenden Fähigkeiten orientierte sich Isernhagen am Dictionary of Occupational Titles (D.O.T.) des U.S. Department of Labor, das 20 Schlüsselemente der physischen Arbeitsanforderungen beschreibt (Fishbain, 1999); (Tabelle 1).

Die EFL fand in den USA viel Zuspruch und gelangte von dort aus 1991 durch Oliveri und Hallmark-Itty nach Europa in die Schweiz. Auch dort konnte sie sich gut etablieren, und so beschloss die Schweizerische Arbeitsgemeinschaft Rehabilitation (SAR) 1995, sie als Standardassessment für die arbeitsbezogene Leistungserprobung in der Schweiz zu unterstützen und zu verbreiten. Die Anerkennung durch die SAR brachte die Bindung an festgelegte Standards mit sich, die eine hohe Akzeptanz durch Versicherungsträger, Rehabilitationsfachleute und Sozialversicherungsgerichte gewährleisten sollte.

Dieser hohe Qualitätsanspruch wurde von der 1999 gegründeten deutschen EFL-Akademie übernommen. Sie definierte Ausbildungs- und Qualitätskriterien neu und ließ diese in die ausschließlich durch sie durchgeführte Ausbildung einfließen. Ein weiterer wichtiger Baustein in der Qualitätssicherung ist neben dem Ausbildungsmonopol die Supervision aller EFL-Berichte, die als Qualitätsnachweis für eine erfolgreiche Zertifizierung dienen. Erst eine Zertifizierung berechtigt zur Durchführung des Verfahrens und wird vom ebenfalls 1999 gegründeten Bundesverband EFL an geschulte und erfahrene Anwender vergeben. Damit stellt der Bundesverband die Dachorganisation der inzwischen 101 EFL-Anwender in Deutschland (EFL-Akademie, 2007) dar (Kaiser et al., 2000).

Inzwischen wird in Deutschland kein anderes FCE-Verfahren in einer größeren Anzahl Einrichtungen eingesetzt. Es hat vor allem in Rehabilitationskliniken und -zentren Verbreitung gefunden, während ERGOS als zweites großes FCE-System eher in Berufsförderungswerken genutzt wird (Erbstößer et al., 2003).

Schlüsselemente arbeitsbezogener Anforderungen im D.O.T.	Zu testende Fähigkeiten in der EFL nach Isernhagen
Heben	Heben Boden- zu Taillenhöhe Heben Taillen- zu Kopfhöhe Heben horizontal
Tragen	Tragen rechte Hand Tragen linke Hand Tragen vorn
Schieben	Schieben/Drücken
Ziehen	Ziehen
	Handkraft rechts Handkraft links
Stehen	Längeres Stehen Stehen vorgeneigt
Gehen	Gehen
Sitzen	Längeres Sitzen Sitzen vorgeneigt
	Arbeit über Kopf
Klettern	
	Treppe steigen
	Leiter steigen
Balancieren	Gleichgewicht
Beugen	Wiederholte Kniebeugen Hockstellung
Knien	Knien
Krabbeln	
Kriechen	Kriechen
Bewegungsradius	Rotation im Sitzen rechts Rotation im Sitzen links Rotation im Stehen rechts Rotation im Stehen links
Handfertigkeit	Handkoordination rechts Handkoordination links
Fingerfertigkeit	
Tasten/Fühlen	
Sprechen	
Hören	
Sehen	

Tabelle 1: Schlüsselemente arbeitsbezogener Anforderungen im D.O.T und der zu testenden Fähigkeiten in der EFL nach Isernhagen²

² Diese Auflistung stellt die immer angegebene Anzahl von 29 zu testenden Einzelfähigkeiten korrekt dar. Bitte lassen Sie sich nicht irritieren, wenn Sie in Listen anderer Veröffentlichungen diese Anzahl nicht nachvollziehen können – oft werden Fähigkeiten zusammengefasst (z.B. wird nur Rumpfrotation ohne Angabe der Richtung angeführt) oder es werden mehrere Tests einzeln aufgeführt, die die gleiche Fähigkeit (z.B. Handkoordination) messen, so dass die Anzahl der Listenpunkte schwankt.

2.2 Einsatzbereiche

Kurz nach der Übernahme der Trägerschaft der EFL durch die SAR veröffentlichten Oliveri, Denier-Bont und Hallmark-Itty einen grundlegenden Artikel über die EFL in der Reihe „Suva Medizinische Mitteilungen“. Dort begründen sie den Einsatz des Testverfahrens so:

"Die Angaben des Klienten über dessen vermutete aktuelle Belastbarkeit bezüglich Arbeit sind subjektiv, besonders nach längerer Schonung und Arbeitsabsenz infolge Unfall oder Krankheit sind sie oft nicht realitätsbezogen und unzuverlässig. Klinische Daten wie beispielsweise Beweglichkeitsmaße von Gelenken und Wirbelsäule und Ergebnisse von Kraft- oder Ausdauer tests ergeben höchstens indirekte Hinweise. In einem ergonomischen Evaluationsverfahren mittels realistischen Arbeitssimulationstests kann hingegen die arbeitsbezogene Leistung direkt beurteilt werden. [...] Ziel dieser Untersuchung ist eine realitätsgerechte Beurteilung der arbeitsbezogenen Belastbarkeit bzw. eine detaillierte Erfassung der physischen Fähigkeiten und Defizite. Die EFL eignet sich zur Standortbestimmung bei Klienten mit chronischen Problemen des Bewegungssystems nach Verletzungen und Operationen oder als Folge von Krankheit."

(Oliveri et al., 1996, S. 15)

Ein Verfahren mit diesen Merkmalen kann vielfältig genutzt werden. Rivier und Seewer (2002) sehen Einsatzbereiche vor allem in der Verwendung der Daten im Rahmen von Gutachten und Gerichtsverfahren, als Orientierung in der funktionellen Rehabilitation und als Möglichkeit einer verbesserten Schätzung, ob physische Fähigkeiten, die zu einem bestimmten Zeitpunkt vorhanden sind, ausreichen, damit die wichtigsten Aufgaben am Arbeitsplatz erfüllt werden können („Job-match“).

Die Arbeitsgruppe um Kaiser lokalisierte in Übereinstimmung mit Oliveri (Oliveri, 1997, nach Kaiser et al., 2000) die EFL vor allem an der Schnittstelle zwischen medizinischer und beruflicher Reha. Als mögliche Ziele und Fragestellungen nennen die Autoren die detaillierte Erfassung der körperlichen Fähigkeiten und Defizite zur Planung einer beruflichen Rehabilitation, sowie zur realitätsgerechten Beurteilung der Arbeitsfähigkeit und der Arbeitsmöglichkeiten. Der Kern liegt damit für sie in der beruflichen, familiären und gesellschaftlichen Reintegration von Menschen nach Unfall oder Krankheit.

Die Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung des Verbandes Deutscher Rentenversicherungsträger (SOMEKO) legt in ihrem Abschlussbericht auftragsgemäß den Schwerpunkt auf die Bedeutung der EFL für die sozialmedizinische Einschätzung (SOMEKO, 2004). Sie möchte die EFL in erster Linie nutzen, um Antworten zu finden auf Fragen wie

- Kann jemand in den Arbeitsprozess zurückkehren?

- Welche Möglichkeiten für eine berufliche Umorientierung gibt es?
- Sind bestimmte Tätigkeiten zumutbar?
- Muss eine Rente gezahlt werden?

Häufig wird jedoch ein und derselbe FCE-Bericht, der für einen bestimmten Zweck in Auftrag gegeben wurde, von verschiedenen Anwendern für die Beantwortung unterschiedlicher Fragestellungen genutzt, wie Strong et al. (2004) betonen. Nach ihren Erkenntnissen verwenden erfahrene FCE-Anwender solche Assessments mit abnehmender Häufigkeit, um

- die Rückkehr eines Patienten in das Erwerbsleben zu planen. Seine physischen Möglichkeiten und Grenzen sollen ausgelotet werden, damit über die Notwendigkeit spezieller therapeutischer und unterstützender Maßnahmen entschieden werden kann.
- das Verhalten eines Arbeiters hinsichtlich Anstrengung, Einstellung und Compliance zu untersuchen.
- eine Mediation einzuleiten und alle Betroffenen an einen Tisch zu bringen, gegebenenfalls eine unabhängige dritte Meinung zu bekommen.
- eine Meinung oder Richtung zu validieren, zum Beispiel in behördlich gestützten Return-to-work-Prozessen oder um dem Arbeitgeber die Sicherheit zu geben, dass der Arbeiter bereit ist, wieder in den Arbeitsprozess zurückzukehren.
- über Ansprüche oder Lohnersatzleistungen zu entscheiden.
- Fortschritte in der Therapie oder im Return-to-work-Prozess zu messen.
- dem Arbeiter die Möglichkeit zu geben, seine eigenen Fähigkeiten und Grenzen kennen zu lernen, ihn wenn nötig zu ermutigen, mit seiner Behinderung klarzukommen und Perspektiven zu entwickeln.

Kontrovers diskutiert wird der Versuch, FCEs zur Bestimmung der Anstrengung und für Schiedsurteile zu verwenden. In Australien beispielsweise ist dieser Einsatz von FCEs verboten, in anderen Ländern ist er übliche Praxis (Strong et al., 2004).

Kraus (1997) führt aus, dass verschiedene Entwicklungen zu einem erhöhten Klärungsbedarf in diesen Bereichen und damit auch zu einer steigenden Nachfrage nach FCEs führen: Die fortschreitende Umgestaltung vieler Arbeitsprozesse hat eine Zunahme insbesondere solcher arbeitsbezogener Erkrankungen zur Folge, die durch häufige Wiederholungen verursacht werden. Außerdem werden die Arbeitnehmer im Schnitt zunehmend älter, und ältere Arbeitnehmer haben ein höheres Risiko, akut und chronisch zu erkranken. Auf der anderen Seite können aufgrund medizinischer Fortschritte auch Menschen mit funktionellen Einschränkungen immer länger und produktiver leben, so dass sich für sie in zunehmendem Maße die Frage stellt, wie

auch sie am Erwerbsleben teilnehmen können. Die Beschäftigung funktionell eingeschränkter Arbeitnehmer bedeutet jedoch für Versicherer, Case Manager und Betriebe ein höheres Risiko, so dass sie von den Ärzten präzise Aussagen über die Arbeitsfähigkeit ihrer Patienten einfordern.

Für die Kostenträger, insbesondere die Rentenversicherung, ist die Erhaltung beziehungsweise Wiederherstellung der Erwerbsfähigkeit ihrer Mitglieder ein sozialpolitisch erwünschtes Ziel. Berufliche (Re-)Integration trägt nicht nur zur finanziellen Entlastung der Rentenversicherung bei, sondern dient auch dem Wohl des Versicherten. Die Bedeutung der Arbeit für das Selbstverständnis und die gesellschaftliche Einbindung des Einzelnen wird im biopsychosozialen Modell der Schmerzentstehung expliziert und durch die Ergebnisse von Fugl-Meyer et al. gestützt. Sie berichten, dass Schmerzkranken nach erfolgreicher beruflicher Wiedereingliederung ein höheres soziales Wohlbefinden angeben als vorher (Fugl-Meyer et al., 1991 nach Beutel et al., 1998).

Nimmt man den psychosomatischen Aspekt chronischer Schmerzen, unter denen viele orthopädischen Patienten leiden, ernst, lohnt sich ein Blick über den disziplinären Zaun: Zwerenz et al. (2004) fanden heraus, dass psychosomatische Reha-Patienten ihre berufliche Belastungen als höher und ihre Ressourcen als geringer einschätzten als Matched Pairs aus der Allgemeinbevölkerung. Dieses Ergebnis erscheint umso bedeutsamer, wenn man die Darstellung von Beutel et al. (1998) hinzuzieht. Sie verweisen auf mehrere Studien, in denen nachgewiesen werden konnte, dass das Gelingen der beruflichen Wiedereingliederung einen wichtigen Prädiktor für die langfristige Besserung psychischer und psychosomatischer Störungsbilder darstellt. Andernfalls kommt nach Ansicht der Autoren ein verheerender Teufelskreis aus 1. Beeinträchtigung der Leistungs-, Arbeits- und Beziehungsfähigkeit; 2. beruflicher Enttäuschung oder Zurücksetzung, längerer Krankheit oder Arbeitslosigkeit und 3. Beeinträchtigung des Selbstvertrauens in die eigene Leistungsfähigkeit und der Leistungsmotivation in Gang (Abbildung 1).

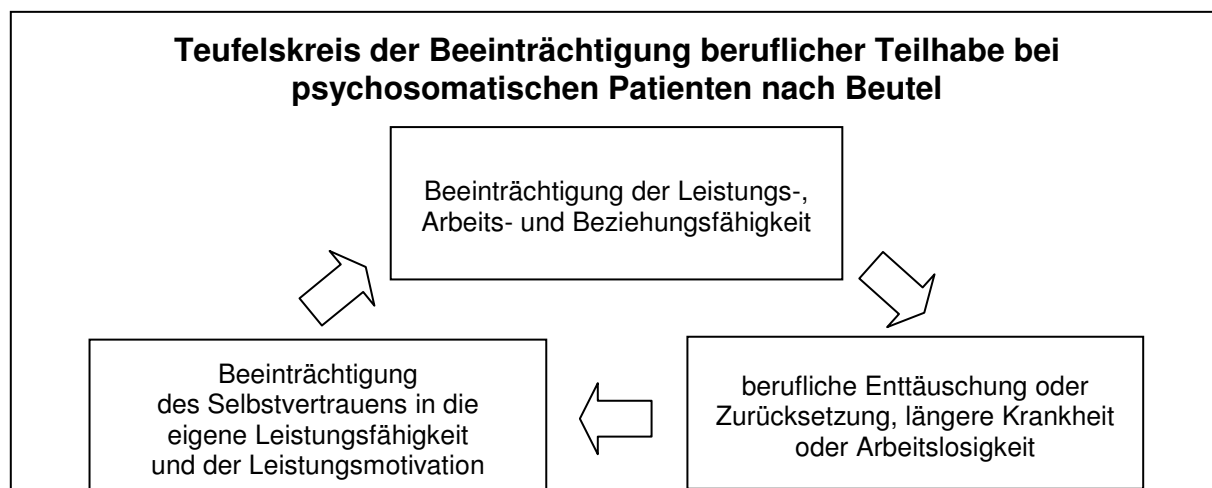


Abbildung 1: Teufelskreis der Beeinträchtigung beruflicher Teilhabe bei psychosomatischen Patienten nach Beutel et al. (1998); (Grafische Umsetzung durch Niemeyer)

Sowohl der finanzielle als auch der immaterielle Aufwand eines Versuchs, in das Arbeitsleben zurückzukehren, sind für die Betroffenen erheblich, die Kosten eines Scheiterns wären hoch. Umso bedeutungsvoller ist nicht nur für Arbeitgeber und Kostenträger, sondern auch für die Betroffenen eine valide Einschätzung ihrer beruflichen Leistungsfähigkeit, auf deren Basis sie adäquat planen können. Genau diesen Anspruch der Kostenträger und der Patienten selbst versprechen die FCE-Verfahren, so auch die EFL nach Isernhagen, zu erfüllen. Aber kann sie dieses Versprechen auch halten?

2.3 Konzeption, Aufbau und Durchführung

Konzeption

Mit ihrem Schwerpunkt auf der Messung der funktionellen Leistungsfähigkeit eines Menschen entsprechen die FCE-Systeme allgemein dem Ansatz der International Classification of Functioning, Disability and Health (ICF). Im Gegensatz zur International Classification of Diseases (ICD) stellt sie die Funktionsfähigkeit des Menschen, nicht seine diagnostizierten Erkrankungen in den Mittelpunkt:

„Der Begriff der Funktionsfähigkeit eines Menschen umfasst alle Aspekte der funktionalen Gesundheit.

Eine Person ist funktional gesund, wenn – vor dem Hintergrund ihrer Kontextfaktoren –

1. ihre körperlichen Funktionen (einschließlich des mentalen Bereichs) und Körperstrukturen denen eines gesunden Menschen entsprechen (Konzepte der Körperfunktionen und -strukturen),
2. sie all das tut oder tun kann, was von einem Menschen ohne Gesundheitsproblem (ICD) erwartet wird (Konzept der Aktivitäten),
3. sie ihr Dasein in allen Lebensbereichen, die ihr wichtig sind, in der Weise und dem Umfang entfalten kann, wie es von einem Menschen ohne gesundheitsbedingte Beeinträchtigung der Körperfunktionen oder –strukturen oder der Aktivitäten erwartet wird (Konzept der Partizipation [Teilhabe] an Lebensbereichen).“

(Deutsches Institut für Medizinische Dokumentation und Information, DIMDI;
WHO-Kooperationszentrum für das System Internationaler Klassifikationen World
Health Organization; 2005, S. 4)

Im Gegensatz zu psychophysischen Assessments der funktionellen Leistungsfähigkeit („Bis zum Geht-nicht-mehr“; der Patient bricht den Test ab) ist die EFL als kinesio-physisches Verfahren angelegt: Bei steigender Belastung achtet ein

speziell ausgebildeter Versuchsleiter, bei EFL-Tests in der Regel ein Physio-, Ergo- oder Sporttherapeut, auf charakteristische Veränderungen in Statik, Bewegungsabläufen und physiologischen Parametern wie Herzfrequenz, Atmung und Transpiration. Werden in diesen Indikatoren ergonomische Unsicherheiten oder andere Anzeichen für funktionelle Grenzen des Probanden erkennbar, beendet der Versuchsleiter den Test. Auf diese Weise sollen sowohl ein zu früher Abbruch als auch eine Überlastung vermieden werden. Im Einklang mit Befunden, die nur mäßige Zusammenhänge zwischen funktioneller Leistungsfähigkeit und empfundenem Schmerz nachgewiesen haben (Kopp et al., 1997/98), spielt das subjektive Empfinden des Probanden, so auch der von ihm geäußerte Schmerz, lediglich eine untergeordnete Rolle (Rivier & Seewer, 2002). Der Patient wird im Zweifelsfall motiviert, aber nicht gezwungen, den Test auch unter Schmerzen weiterzuführen. Die Grundvoraussetzung für die Durchführung einer EFL, die sich auch in den Indikationen beziehungsweise Kontraindikationen niederschlägt, ist ein medizinisch stabiler Zustand des Patienten (Kaiser et al., 2000).

Als besonders vorteilhaft an dieser Konzeption beurteilt die Interessengemeinschaft Physiotherapie Rehabilitation (igptr) (Website; keine Jahresangabe) neben einer größtmöglichen Sicherheit für den Patienten die Chance, die Konsistenz der Testergebnisse beurteilen und daraus Schlüsse auf Anstrengung und Kooperation des Patienten ziehen zu können.

Aufbau und Durchführung

Die EFL hat den Anspruch, Arbeitsbelastungen isoliert, aber realitätsgetreu zu simulieren. Ein wichtiger Faktor dafür ist die valide Erfassung der Dauerleistungsfähigkeit, die über eine verhältnismäßig lange Durchführungszeit von jeweils drei bis vier Stunden an zwei aufeinander folgenden Tagen gewährleistet werden soll.

Dabei ordnet eine durch die Schweizerische Arbeitsgemeinschaft Rehabilitation SAR (1997) im Manual streng festgelegte Testreihenfolge beiden Tagen die Prüfung bestimmter Einzelfähigkeiten zu; die als zentral angesehenen Fähigkeiten des Hebens und des Sitzens werden doppelt getestet (Tabelle 2). Auf diese Weise können Einbrüche aufgrund von Überlastung, aber auch Verbesserungen im Testergebnis erfasst und berücksichtigt werden.

Im Test wird das Leistungsmaximum des Patienten direkt ermittelt. Aus diesem wird dann per Formel die Dauerleistungsfähigkeit für den ganzen Arbeitstag extrapoliert wird; ein Globalindex wird nicht berechnet. Die Ergebnisse sowie besondere Vorkommnisse und Beobachtungen werden streng protokolliert und regelgeleitet in eine Kategorientabelle eingetragen, die als Fähigkeitenprofil genutzt werden kann. Ein Ausschnitt einer EFL-Tabelle ist in Abbildung 2 zu sehen, eine vollständige EFL-Tabelle finden Sie auf der beiliegenden CD).

Tag 1:	Tag 2:
<ul style="list-style-type: none"> • Anamnese, Sitzen • Klinische Untersuchung (diese wurde bei unseren Patienten bereits im Vorfeld durch den behandelnden Reha-Arzt durchgeführt) • Heben Boden- zu Taillenhöhe • Heben Taillen- zu Kopfhöhe • Heben horizontal • Ziehen • Stoßen • Tragen rechts • Tragen links • Tragen beidhändig • Arbeiten über Kopf • Vorgeneigte Haltung • Rumpfrotation rechts, links im Sitzen • Rumpfrotation rechts, links im Stehen 	<ul style="list-style-type: none"> • Heben Boden- zu Taillenhöhe • Heben Taillen- zu Kopfhöhe • Heben horizontal • Kriechen • Knien • Hocke • Wiederholte Kniebeugen • Handkoordination • Handkraft • Sitzen • Gehen • Treppe steigen • Leiter steigen • Gleichgewicht • Stehen

Tabelle 2: Reihenfolge der EFL-Subtests an den beiden Testtagen

Testdaten und Belastbarkeitswerte						
Evaluation der funktionellen Leistungsfähigkeit EFL						
Belastungsdauer pro 8 h	nie	gelegentlich	zeitweise	überwiegend	ständig	Belastungsgrenze
Belastungskategorie		max.	schwer	mittel	leicht	(Bemerkungen, Beobachtungen)
Lasten/Kraft (kg, kp =10N)						
Heben Boden zu Taillenhöhe		32,5	27,5	20	12,5	Anreißen beim Anheben, beim Runtergehen sehr schnell, Abbruch bzw. Unterbrechung der Bewegungen, Schmerz mimik, Hand im Kreuz, massiert Gesäßfalte, insgesamt wacklig
Heben Taillen- zu Kopfhöhe		15	12,5	10	5	Linkes Bein entlastet, Kiste auf oberes Brett mit viel Schwung und durch Rumpfrückneigung,
Heben horizontal		35	27,5	20	12,5	Arme voll abgespreizt, Schultern hoch, wackliger unsicherer Gang
Schieben/Drücken		40	30	20	10	
Ziehen		31	23,25	15,5	7,75	

Abbildung 2: Ausschnitt aus einer EFL-Tabelle

Dem Test vorgeschaltet werden ein Belastungs-EKG, eine Erhebung der Risikofaktoren und bei Probanden, die im Erwerbsleben stehen, eine Arbeitsplatzanamnese. Vor Beginn und direkt nach dem Ende des eigentlichen Tests füllen die Patienten den Fragebogen "Selbstbeurteilung der körperlichen Fähigkeiten – PACT"³ (Schweizerische Arbeitsgemeinschaft für Rehabilitation, 1996) aus. Auf diese Weise können neben den Momentaufnahmen auch Veränderungen errechnet werden. Abgerundet wird das Instrumentarium durch einen Schmerzfragebogen, in dem die Schmerzen in den letzten 7 Tagen sowie die aktuellen Schmerzen zu Beginn und zum Ende eines jeden Testtages auf einer numerischen Analogskala (0-10) festgehalten werden. Durch einen ausgebildeten Arzt interpretiert, soll die EFL die sozialmedizinische Einschätzung auf eine objektive, reliable und valide Basis stellen.

Indikationen und Kontraindikationen

In der Literatur gilt eine EFL dann als indiziert, wenn eine bestimmte Fragestellung beantwortet werden soll und der Patient für den Test ausreichende physische und psychische Ressourcen mitbringt.

Hart et al. (1993) halten eine EFL für nützlich, wenn

- der Behandlungsfortschritt stagniert,
- sich eine auffällige Diskrepanz zwischen subjektiven Beschwerden und objektiven Befunden zeigt,
- ein Patient Schwierigkeiten hat, in eine Erwerbsarbeit zurückzukehren, oder
- eine Bestimmung der funktionellen Leistungsfähigkeit für die Planung der beruflichen Zukunft oder einen medizinrechtlichen Vergleich erforderlich wird.

Die Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung (SOMEKO) (2004) fokussiert eher auf beruflich orientierte Rehabilitation und rät einen EFL-Test für folgende Zwecke an:

- Bestimmung des Restleistungsvermögens
- Feststellung der Arbeitsfähigkeit, z.B. bei längeren AU-Zeiten
- Beurteilung der Anforderungen eines bestimmten Berufes oder einer alternativen Tätigkeit
- Erstellung eines positiven und negativen Leistungsbildes
- Beschreibung eines individuellen Ausgangsniveaus der Leistungsfähigkeit zur Einleitung einer Arbeitserprobung und Planung weiterer Interventionen
- Screening bei der Auswahl eines geeigneten Arbeitsplatzes

³ Der Fragebogen „Selbstbeurteilung der körperlichen Fähigkeiten – PACT“ ist die Deutsche Version des 1989 erschienenen „Performance Assessment and Capacity Testing“ von Matheson & Matheson.

- Unterstützung der Arbeitsplatzgestaltung
- Unterstützung der Umsetzung auf einen leidensgerechten Arbeitsplatz

Grundsätzlich eignet sich die EFL für Patienten mit chronischen Problemen des Bewegungsapparates oder neurologischen Dysfunktionen, die muskuloskelettal und kardiovaskulär ausreichend belastbar sind. Körperliche und psychische Zustände, bei denen die Durchführung für den Patienten riskant wäre, beispielsweise weil sie sich noch in der Akutphase befinden, noch keine medizinische Stabilität erreicht haben oder „Red flags“ nach Waddell zeigen, sind Kontraindikationen. Auch mit Patienten, die Gehhilfen benötigen, mit dem Testleiter nur unzureichend kommunizieren können oder nicht ausreichend motiviert sind, sollte eine EFL nicht durchgeführt werden (Hart et al., 1993; Kaiser et al., 2000; Strong et al., 2004).

Kraus (1997) setzt sich mit medizinrechtlichen Fragen auseinander und empfiehlt besondere Vorsicht bei der Durchführung von FCEs und einen engen Austausch des Therapeuten mit dem überweisenden Arzt, damit der Therapeut im Falle von Komplikationen nicht im Nachhinein belangt werden kann. Besonderen Wert legt er darauf, dass der Proband vor dem Test eine Einverständniserklärung ("informed consent") unterschreibt, die die Risiken des Tests darstellt und seine Verpflichtung betont, bei Auftreten von Schmerzen oder anderen Beschwerden sofort das Personal zu unterrichten. Das enthebe das Personal nicht komplett der Gefahr, zur Verantwortung gezogen zu werden, gebe aber einen gewissen Schutz. Er geht jedoch davon aus, dass das Einhalten sinnvoller Vorsichtsmaßnahmen normalerweise einen sicheren und unkomplizierten Testablauf gewährleistet.

3. Stand der Forschung: Testgüte der EFL

Neben der EFL gibt es eine ganze Reihe von weiteren FCE-Systemen auf dem Markt, die mit verschiedenen Ansätzen und Schwerpunkten arbeiten und unterschiedlich gut evaluiert worden sind. King et al. (1998) haben in ihrem Artikel "A Critical Review of Functional Capacity Evaluations" festgestellt, dass es DEN besten Test nicht gibt, sondern der Tester immer die Aufgabe hat, den Test auszuwählen, der dem Probanden, der Situation und der Fragestellung am besten entspricht. In diesem Sinne reicht es für die Frage, ob jemand seine alte Arbeit wieder aufnehmen kann, völlig aus, die kritischen Anforderungen des konkreten Arbeitsplatzes zu prüfen, während die Eignung eines Probanden für den allgemeinen Arbeitsmarkt besser mit Hilfe eines stärker generischen Instrumentes zu ermitteln ist. Im Sinne von Transparenz und Standardisierung sollte man bei der Auswahl einer FCE expliziten Regeln folgen, für die King et al. die Checkliste des National Institute of Occupational Safety and Health (NIOSH) empfehlen. Sie rät, dabei zu fragen:

- Ist die Anwendung des Tests sicher?
- Liefert der Test reliable Ergebnisse?
- Ist der Test valide in Bezug auf die Anforderungen der Arbeit und kann er ein sicheres Belastungsniveau vorhersagen?
- Ist der Test valide hinsichtlich spezifischer arbeitsbezogener Fähigkeiten?
- Ist der Test praktisch in der Anwendung?
- Sagt der Test das Risiko zukünftiger Verletzung oder Erkrankung vorher?

(King et al., 1998; S. 857)

Auch Hart et al. (1993) orientierten sich in ihren „Guidelines for Functional Capacity Evaluation of People With Medical Conditions“ an diesen – und damit weitgehend den üblichen – Testgütekriterien, und erweiterten die Liste noch um den Punkt

- Nützlichkeit:
Grad, zu dem der Test die Bedürfnisse des Probanden, des Überweisers und des Kostenträgers erfüllt.

Sie betonen, dass diese Aspekte der Testgüte normalerweise voneinander abhängig sind und die Veränderung eines Aspektes auch Veränderungen in den anderen zur Folge hat. Aus diesem Grund müssen bei der Auswahl eines Tests immer entsprechende Prioritäten gesetzt und beachtet werden.

Im Folgenden möchten wir diskutieren, inwieweit die EFL nach dem aktuellen Stand der Literatur diesen Kriterien entspricht.

3.1 Sicherheit des Probanden in der EFL

Bei hoch belastenden Testverfahren wie den FCEs, bei denen Menschen an die Grenzen ihrer Leistungsfähigkeit geführt werden sollen, muss die Sicherheit der Probanden an erster Stelle stehen. Isernhagen und in der Folge viele EFL-Anwender und Forscher gehen davon aus, dass diese durch die kinesiophysische Konzeption der EFL gewährleistet ist, wenn die durchführenden Therapeuten eine entsprechende Ausbildung und Erfahrung mitbringen (Hart et al., 2003; Rivier & Seewer, 2002; Interessengemeinschaft Physiotherapie Rehabilitation (igptr), Website, keine Jahresangabe).

Voraussetzung für die Durchführung einer EFL sind medizinische Stabilität und die Abwesenheit von „Red flags“ nach Waddell. Schmerzen an sich stellen keine Kontraindikation dar, jedoch muss nach Isernhagen vor allem bei Patienten, die seit mehr als 4 Wochen inaktiv sind, mit einer Schmerzzunahme gerechnet werden (Hart et al., 1993). Dieses scheint sich im Klinikalltag zu bestätigen, wie verschiedenen Aussagen von Ärzten, Therapeuten und Patienten zu entnehmen ist.

Glücklicherweise scheinen bei verantwortungsvoll durchgeführten EFL-Tests ernsthafte Schäden kaum vorzukommen, zumindest haben wir in der Literatur keinerlei Hinweise darauf gefunden. Beachtet werden sollte jedoch, dass auch kurzzeitige erhebliche Verschlimmerungen der Beschwerden für die Patienten äußerst unangenehm sein oder bedrohlich wirken können. Eine solche Erfahrung kann sich negativ auf die Testmotivation, gegebenenfalls sogar auf die allgemeine Bereitschaft, sich anstrengenden Tätigkeiten zu stellen, niederschlagen. Eine Zunahme von Bewegungsängsten und eine Verstärkung der Überzeugung, Anstrengung führe zu Schmerzen und deshalb sei Schonung nötig, kann in diesen Fällen nicht ausgeschlossen werden⁴. Rivier und Seewer (2002) diskutieren in diesem Zusammenhang sogar die ethische Vertretbarkeit und fragen,

„...bis zu welchem Punkt darf man von jemandem, der Schmerzen empfindet, verlangen, sich maximal anzustrengen, auch wenn der Untersucher weiß, dass diese Anstrengung gefahrlos ist?“

(Rivier & Seewer, 2002; S. 39)

⁴ Reneman et al. (2003) haben zwar keine Korrelationen zwischen Kinesiophobie und funktioneller Leistungsfähigkeit gefunden – sie haben aber auch nicht die **Veränderung** der Kinesiophobie durch einen FCE-Test untersucht.

Deshalb ist unserer Ansicht nach gerade in kinesiophysischen Tests ein ausgesprochen sensibler Umgang mit dem Patienten wichtig. Grundlegend dafür sind auf Seiten des Therapeuten eine gute Ausbildung, viel Erfahrung, eine hohe Motivation und die Fähigkeit, eine gute und vertrauensvolle Beziehung zum Probanden aufzubauen. Nicht bei allen Testpersonen sind sämtliche Anzeichen des Erreichens der Leistungsgrenze gleichermaßen gut zu beurteilen⁵, so dass der gesamte Indikatorenkatalog ständig beachtet werden muss. Nur dann ist die Beurteilung einer Belastung als leicht, mittel oder schwer zuverlässig und wird über mehrere Beobachter hinweg in etwa gleich eingeschätzt. Auch kann man nicht davon ausgehen, dass jeder Patient aggraviert, wenn er selbst überraschend früh meint, seine Leistungsgrenze sei erreicht. Nur mit einem differenzierten Blick auf den einzelnen Probanden kann ein Höchstmaß an Sicherheit für ihn gewährleistet werden – und dieses ist ohne Zweifel unabdingbar.

3.2 Unabhängigkeit des EFL-Ergebnisses vom Untersucher – Objektivität

Nur mit Testverfahren, die unabhängig vom Untersucher immer auf die gleiche Weise durchgeführt werden, lassen sich allgemeingültige Ergebnisse erzielen.

Kraus (1997) kritisiert die reine klinische Untersuchung als ausgesprochen wenig objektiv in Bezug auf die Beurteilung funktioneller Fähigkeiten. Er hält FCE-Systeme darin grundsätzlich für überlegen:

"One of the most subjective decisions in medicine occurs when a physician completes a functional capacities form on the basis of the physical examination alone. It truly becomes an estimate of function and often a poor estimate. [...] A better method is to obtain a functional capacity *evaluation* that can test these capacities [...]"

(Kraus, 1997, S. 543-544)

Tatsächlich scheint die EFL ihren Anspruch auf Objektivität oder zumindest Intersubjektivität durch ihren hohen Grad an Standardisierung mit einem detaillierten Manual, einer exakten Festlegung von Testreihenfolge, Art der Durchführung und Kriterien für den Testabbruch einlösen zu können. Auf der anderen Seite berichten die Anwender selbst immer wieder von Unterschieden zwischen den durchführenden Therapeuten im Hinblick auf die „Härte“ gegenüber den Patienten und auf die Wahrnehmung, wann die ergonomisch sichere Grenze erreicht ist. Auch Die

⁵ Beispielsweise ist eine Rekrutierung von Hilfsmuskulatur bei adipösen Probanden erheblich schwerer zu beobachten als bei schlanken.

Kölnische Rück (1999) kritisiert, dass die Beobachtung des Patienten nur live ohne die Möglichkeit der Reproduktion erfolgt und stark von der Konzentration und der Beobachtungsgabe des Untersuchers abhängt. Fraglich ist ebenfalls, inwieweit ein gewisser Spielraum in der Durchführung auf Kosten der Standardisierung und damit möglicherweise auch der Objektivität sogar nötig ist, um Compliance und Gesundheit der einzelnen Patienten nicht zu gefährden.

3.3. Zuverlässigkeit des EFL-Ergebnisses – Reliabilität

Ein guter Test muss nicht nur objektiv, sondern auch zuverlässig sein und unter den gleichen Bedingungen immer wieder zu dem gleichen Ergebnis kommen. Nur dann können Unterschiede im Resultat auf Unterschiede in der Realität und nicht etwa auf Messfehler zurückgeführt werden.

Innes und Straker (1999a) setzen sich in ihrem Review mit der Reliabilität arbeitsbezogener Assessment-Verfahren, unter anderem der EFL und dem PACT, auseinander. Sie stellen die verschiedenen Formen der Reliabilität vor und betonen, dass für FCEs die Test-Retest-Reliabilität, auch Intra-Rater-Reliabilität genannt, und die Interrater-Reliabilität, also die Übereinstimmung der Urteile mehrerer Beurteiler, als die wichtigsten Reliabilitätsarten angesehen und am häufigsten ermittelt werden.

In ihrem Review haben die Autoren Literaturdatenbanken und verschiedene Sekundärquellen durchsucht und von 55 gefundenen Assessments 28 in die Analyse einbezogen. Sie schätzten zunächst die Qualität des Forschungsstandes hinsichtlich der Reliabilität der jeweiligen Assessments ein. Für die ausreichend untersuchten Verfahren folgte dann eine Beurteilung ihrer Zuverlässigkeit an sich. Für viele FCEs fanden sie eine schlechte, für die EFL eine eher mäßige Forschungssituation. Relativ gut abgesichert ist nach ihren Erkenntnissen lediglich die Retest-Reliabilität ihrer Hebekomponente. Auch dem PACT schreiben sie eine gute Retest-Reliabilität und zusätzlich eine sehr gute Split-half-Reliabilität zu. Die Autoren betonen, dass gerade bei FCEs ein guter Kompromiss zwischen Reliabilität und Validität getroffen werden muss – je näher eine FCE an der Arbeitswirklichkeit ist, umso valider ist sie. Dann allerdings kann sie nicht mehr so stark standardisiert sein und weist Einbußen bei der Reliabilität auf.

Gross & Battié (2002) sehen spezifische Varianzquellen, die die Reliabilität beeinträchtigen können, beim Testleiter, beim Probanden und in der Interaktion zwischen beiden: Kinesiophysische Verfahren erheben den Anspruch weitgehender Objektivität, doch dürfe man den Anteil subjektiver Wahrnehmung und Interpretation des Wahrgenommenen durch den Testleiter nicht unterschätzen – und sein Urteil

bestimmt letztlich das Ergebnis des Tests. Ebenso können Diskrepanzen in verschiedenen Testdurchgängen auf verschiedene Ausgangsbedingungen auf Seiten des Probanden, beispielsweise hinsichtlich seines Befindens, seiner Motivation und einer möglichen Schmerzintensität, zurückzuführen sein. Und letztlich könne auch die Interaktion zwischen Testleiter und Proband sehr unterschiedlich verlaufen und sowohl den Testleiter als auch die Performance des Probanden beeinflussen. Tatsächlich ermittelten Gross und Battié recht gute Reliabilitätswerte (vgl. Tabelle 3), wobei über alle Probanden hinweg im zweiten Test signifikant schwerere Gewichte gehoben und beidhändig getragen wurden als im ersten. Besonders große Unterschiede zwischen den Testergebnissen gingen mit erheblich differierenden Tagesformen der Probanden in Bezug auf Befinden, Motivation oder Schmerzintensität einher, so dass die Hypothesen der Arbeitsgruppe in Bezug auf einige Varianzquellen durch die Ergebnisse gestützt wurden.

Eine weitere Erklärung für eine solche Befundstruktur liefern Erbstößer et al. (2003) mit ihrem Hinweis auf die Reaktivität von FCEs: Möglicherweise werden bei mehrmaliger Testung die Ergebnisse durch Lerneffekte bei den Probanden oder einen Gewinn an Zutrauen zur eigenen Leistungsfähigkeit beeinflusst, so dass in der zweiten Messung höhere Leistungswerte ermittelt werden. Um zumindest einen Teil der Variabilität der Leistung – unabhängig von ihrer Ursache – erfassen zu können, scheint es auf dem Hintergrund dieser Befunde und Überlegungen absolut sinnvoll, die funktionelle Leistungsfähigkeit an mindestens zwei Tagen zu messen. Die Kölnische Rück (1999) regt sogar an, die EFL standardmäßig zu videographieren, so dass die biomechanischen Kriterien auch zu einem anderen Zeitpunkt oder durch einen anderen Rater beurteilt werden können, und die Abhängigkeit des Testergebnisses von der Beobachtungsgabe und der Konzentration des Raters reduziert wird.

Noch unklar ist es nach Gross & Battié (2002), welche Auswirkungen es in der normalen EFL-Praxis hat, dass immer wieder andere Testleiter die EFL durchführen. Ebenso betonen die Autoren die momentane Unsicherheit darüber, welchen Effekt Veränderung von Faktoren wie Disziplin oder Erfahrung der Therapeuten oder sogar des Testsettings auf die Reliabilität der EFL haben.

Gute Reliabilitätswerte scheint die Interessengemeinschaft Physiotherapie Rehabilitation (igptr) der EFL auf ihrer Homepage in ihrer kurzen Zusammenschau der Ergebnisse einzelner Studien zu bescheinigen (Tabelle 3).

	Kriterium	Intra-Rater-Reliabilität/ Test-Retest-Reliabilität	Interrater-Reliabilität
Oliveri et al. (2002)	Gehobene Last (leicht – mittelschwer – schwer) Heben Boden-Taillenhöhe; Heben Taillen-Überkopfhöhe; Heben horizontal	Weighted kappa $\kappa_w = .83$ Übereinstimmung: 87%	Weighted kappa $\kappa_w = .77$ Übereinstimmung: 80%
Isernhagen et al. (1999)	Gehobene Last (leicht – mittel – schwer) Gehobene Last (leicht – schwer) Gehobene Last (schwer) Gehobene Last (leicht) Gehobene Last (nicht leicht – nicht schwer)	Cohens Kappa $k = .68$ Cohens Kappa $\kappa = .81$ Übereinstimmung: 90% Übereinstimmung: 94%	
Gross & Battié (2002)	Maximale gehobene oder getragene Last Heben Boden-Taillenhöhe; Taillen- Kopfhöhe; Heben horizontal; Tragen links, rechts, vorn	Intra-Class-Coefficient: $ICC_{min} = .78$; $ICC_{max} = .94$	Intra-Class-Coefficient: $ICC_{min} = .95$; $ICC_{max} = .98$

Tabelle 3: Überblick der Interessengemeinschaft Physiotherapie Rehabilitation (igptr) über verschiedene Studien zur Reliabilität der EFL

Ergänzt werden könnte diese Tabelle durch die Studie von Reneman, Jaegers et al. (2002). Sie ließen 5 geschulte Rater Videoaufnahmen der Hebe- und Tragetests von 4 gesunden Probanden zwei Mal einschätzen und kamen so auf Inter-Rater-Reliabilitäten zwischen 87 und 96% sowie Retest-Reliabilitäten zwischen 93 und 97%. Leider haben die Autoren der igptr-Übersicht nur die Ergebnisse der Studien berichtet, ohne wie Innes und Straker Angaben zu deren Qualität zu machen.

Aus der Auflistung in Tabelle 3 wird deutlich, dass fast nur Reliabilitätsuntersuchungen von Hebetests veröffentlicht wurden. Möglicherweise

haben sich diese als besonders reliabel erwiesen, während Studien, die die Reliabilität anderer Komponenten untersuchten, einem Publication-Bias zum Opfer gefallen sind. Diese Annahme wird gestützt durch die Ergebnisse aus der Studie von Brouwer et al. (2003), die für nur 15 EFL-Subtests akzeptable Retest-Reliabilitäten auf der Basis von Werten für Cohens Kappa ($\kappa > .60$) und für nur 11 Tests akzeptable Werte auf der Basis der Intra-Class-Korrelationskoeffizienten ($ICC > .80$) ermittelten.

Ein anderer Grund für die starke Konzentration auf Hebetests könnte die herausragende Bedeutung sein, die ihnen zugeschrieben wird:

„Return to work decisions are often made or influenced by both the heaviest amount of weight that can be safely lifted and the weight that can be handled repetitiously. [...]

The accurate measurement of lifting abilities in these people is paramount in the determination of work related functional performance. [...] Since the vocational placement, safety and earning potentials of the person are affected by the reported lifting abilities, reliable measurement techniques are essential.“

(Isernhagen et al., 1999, S. 145)

Obwohl auch im DOT- und im REFA-System der Arbeitsschwere die Hebeleistung das einzige Kriterium für die Bestimmung des Leistungsniveaus ist, wird diese damit unserer Ansicht nach zu stark gewichtet. Sicherlich ist das Heben in den meisten vor allem körperlich schweren Berufen eine Schlüsselfähigkeit. Die für den einzelnen Patienten kritischen Arbeitsanforderungen können jedoch auch ganz andere sein. In solchen Fällen können auch Personen Probleme mit der Rückkehr in das Arbeitsleben haben, die nicht schwer heben müssen, aber in anderen wesentlichen Bereichen funktionell beeinträchtigt sind – beispielsweise ein Uhrmacher mit einer neurologischen Störung im Bereich der Hand und der Finger, eine Bäckereiverkäuferin, die nicht lange stehen, oder eine Bürokräft, die nicht lange sitzen kann. Aus diesem Grund ist für eine aussagekräftige FCE immer die Reliabilität der ganzen Testbatterie oder zumindest der kritischen Subtests entscheidend, nicht lediglich die der Hebetests.

Eine ähnliche Kritik wird auch von King et al. (1998) sowie von Innes und Straker (1999a) geäußert. Sie folgern, dass es zu wenig hochwertige Studien zu den anderen Subtests sowie der gesamten EFL gebe, als dass ihre Reliabilität (ebenso wie die der meisten anderen FCE-Systeme!) für klinische oder sogar juristische Zwecke als ausreichend nachgewiesen gelten könne (King et al., 1998; Innes & Straker, 1999a).

Gatty (2002) geht noch einen Schritt weiter und bezweifelt grundsätzlich, dass eine Bestätigung der Testergebnisse an mehreren aufeinander folgenden Tagen möglich ist – vor allem bei Patienten, die von Tag zu Tag mehr Beschwerden entwickeln.

3.4. Aussagekraft des EFL-Tests – Validität

Das vierte große Kriterium der Testgüte fokussiert auf die inhaltliche Korrektheit eines Tests: Misst er tatsächlich das, was er zu messen vorgibt? Liefert er die Informationen, die für eine bestimmte Entscheidung nötig sind?

So definiert, ist die Validität kein Kennzeichen eines Tests an sich, sondern muss immer kontextgebunden im Hinblick auf die Beantwortung einer bestimmten Fragestellung bewertet werden. Das erschwert ihre wissenschaftliche Untersuchung und ist sicherlich einer der Gründe für die extrem begrenzte Datenlage. Trotzdem ist die Frage nach der Validität weitreichend, da alle Formen der Validität für arbeitsbezogene Assessments von großer Bedeutung sind, wie Innes & Straker (1999b) betonen.

Bisher sind sich die Review-Autoren einig, dass die Validität der EFL noch nicht in ausreichendem Umfang erforscht wurde. Immerhin gehört sie jedoch zu den wenigen FCEs, für die in einzelnen Studien Komponenten der Validität untersucht worden sind. Allerdings werden auch diese Studien kritisiert – sowohl wegen der mäßigen Qualität der Publikationen (King et al., 1998; Innes & Straker, 1999b) als auch wegen der Vermutung der unzureichenden Unabhängigkeit der Forscher (Rivier & Seewer, 2002). Dabei bezieht sich die Kritik durchgängig auf alle Validitätsarten. Im Folgenden sollen ein Überblick über die bisherigen Befunde und Statements gegeben und diskutiert werden:

Face validity

Die Face validity ist eng verwandt mit der Inhaltsvalidität und gibt die offensichtliche Entsprechung eines Tests mit dem an, was er erfassen soll. Sie ist nicht im engeren Sinne messbar, sondern wird in der Regel durch einen Expertenkonsens eingeschätzt, wenn das Instrument komplett vorliegt. Eine große Rolle spielt sie in der Darlegung der Relevanz eines Tests gegenüber den Testleitern, den Klienten und den Empfängern der Ergebnisse wie im Falle der EFL vor allem Versicherungen, Case Managern und Arbeitgebern.

Die Face validity der EFL als grundsätzliches Messinstrument körperlicher Leistungsfähigkeit kann als relativ hoch angesehen werden. Hebt man jedoch auf ihren Einsatz als Vorhersageinstrument für eine sichere Rückkehr zur Arbeit ab, so kommt – vor allem von den Probanden selbst – immer wieder die Kritik, der Test sei zu alltagsfern. Neben der starken Fokussierung auf die Hebetests für die Bestimmung der allgemeinen Leistungsfähigkeit werden insbesondere die freundlichen Testbedingungen in einem warmen, trockenen Raum unter der Anleitung eines in der Regel freundlichen Therapeuten als unrealistisch angesehen.

Die direkte Übertragung der unter guten Bedingungen erzielten Testergebnisse auf die Leistungsfähigkeit unter erheblich weniger günstigen Bedingungen erscheint äußerst fraglich (vgl. King et al., 1998).

Weiterhin werden die im Vergleich zum Arbeitsalltag erheblich eingeschränkten Belastungszeiten und die starke Isolation der einzelnen Fähigkeiten bemängelt, wie der Kommentar eines unserer Probanden illustriert: „Hier soll ich tragen und dann eine Leiter hochklettern – aber ich bin Fensterbauer, in meinem Beruf muss ich mit den sperrigen schweren Fenstern auf die Leiter. Und das den ganzen Tag. Das ist doch etwas ganz anderes!“

Inhaltsvalidität

Die Inhaltsvalidität gibt den Grad an, zu dem die Testitems den Leistungsbereich abdecken, den der Test messen soll. Sie ist der Face validity sehr ähnlich und wird in der Regel auch durch Experten eingeschätzt. Die Inhaltsvalidität richtet sich jedoch eher an Experten und bezieht sich nicht nur auf den Test als Ganzen, sondern auch auf einzelne Teile. Sie zu etablieren sollte Bestandteil der Testentwicklung sein.

Oft wird die EFL von Experten aufgrund ihrer DOT-Basierung (vgl. Kapitel 2.1) als inhaltlich valide eingeschätzt. Innes und Straker bemängeln dieses Kriterium jedoch als unzureichend, da das DOT-System ihrer Auffassung nach keinesfalls als bestätigter Standard gelten kann (Innes & Straker, 1999b).

Ebenfalls kritisch betrachtet werden die bei der Face validity schon diskutierte Alltagsferne und der Ansatz, die arbeitsbezogene Leistungsfähigkeit nicht unter Realbedingungen zu messen, sondern auf der Basis der Ermittlung kurzfristiger Leistungsmaxima unter guten Bedingungen zu schätzen. So ist nach Aussage sowohl von Patienten als auch von Ärzten, Therapeuten und Wissenschaftlern (Abdel-Moty et al., 1993; Hart et al.; 1993; King et al., 1998; Gatty, 2002; Nellesen et al., 2002; Erbstößer et al., 2003) die Hochrechnung der Testleistung auf einen 8-Stunden-Tag bei 5 Arbeitstagen pro Woche in etwa 46 Wochen pro Jahr umstritten. Saunders et al. (1997) halten die gängige Praxis, eine allgemeine Formel zu diesem Zweck einzusetzen, sogar für ausgesprochen riskant und warnen vor der Gefahr von Überlastungsschäden. So sind beispielsweise viele Bandscheibenpatienten durchaus in der Lage, den Test „längeres Sitzen“ über 30 Minuten zu bestehen, benötigen dann aber eine längere Bewegungspause. Nach der EFL-Routine wird ihnen damit jedoch die Fähigkeit bescheinigt, „ständig“, das heißt in 90% der Vollarbeitszeit, sitzen zu können – womit sie vermutlich schnell Probleme bekommen werden.

Kraus (1997) führt weitere Faktoren an, die ein Arzt für eine Return-to-work-Prognose in Betracht ziehen muss, die jedoch typischerweise nicht durch eine FCE getestet werden. So ist beispielsweise die Fähigkeit, den Weg von zu Hause zur

Arbeit zu bewältigen, durchaus ein wichtiges Kriterium. Berücksichtigt werden müssen nach Kraus ebenso die Ausdauer des Patienten und die Zeit, die er zur Erholung benötigt. Braucht beispielsweise jemand aufgrund starker Einschränkungen im Alltag erheblich mehr Zeit für die tägliche Hygiene, müsse diese in die Schätzung der Zeitspanne, die er täglich arbeiten kann, eingerechnet werden. Trotz dieser Einschränkungen hält er FCEs für die besten momentan verfügbaren Werkzeuge auch für sozialmedizinische Einschätzungen. Kaiser et al. (2000) teilen diese Position. Sie befürworten in Grenzfällen eine mehrtägige Arbeitserprobung, halten die EFL aber für eine in den meisten Fällen recht gute Annäherung an die Wirklichkeit unter Berücksichtigung der Kosteneffizienz.

Ein weiteres Problem ist die Konfundierung von Belastungsintensität und Belastungsdauer in der EFL: Der Test setzt stets bei einer geringen Belastung an, die bis zum ergonomischen Maximum gesteigert wird. Das führt dazu, dass die Probanden mit höheren Leistungsmaxima vor Erreichen des Maximums schon erheblich länger belastet worden sind als die eher schwachen Probanden. So verläuft die Gesamtbelastung weniger linear als es auf den ersten Blick scheint. Die Arbeitsgruppe um Tittor versucht, dieses Problem zu lösen, indem sie ihren Körperlichen Aktivitätstest (KAT) nicht bei allen Probanden mit der gleichen Belastungsstufe beginnen lässt. Stattdessen wird durch einen vorgeschalteten PACT ermittelt, auf welcher Belastungsstufe sich der Proband selbst einschätzt und welche Hebeleistung er sich ohne Probleme zutraut. Diese Einschätzung liegt in der Regel unter dem in einem Test der funktionellen Leistungsfähigkeit ermittelten Leistungsmaximum, wird aber trotzdem noch einmal durch einen Physiotherapeuten anhand der Konstitution des Probanden auf Plausibilität überprüft. Der KAT beginnt dann auf der noch als leicht geltenden Belastungsstufe. Auf diese Weise durchlaufen die Probanden in der Regel nur wenige Belastungsstufen bis zur individuellen Leistungsgrenze (Lux, 2007; Mitteilung per Mail).

Ein weiteres spezielles Problem der EFL ist die Frage nach der Übertragbarkeit des Konzepts und der Messmethoden aus den USA auf deutsche Verhältnisse. Wie in Kapitel 1.1 beschrieben, basiert die EFL auf der Beschreibung einer großen Anzahl Berufe im Dictionary of Occupational Titles. Aber sehen die Berufe unter den Arbeitsbedingungen in den USA wirklich ähnlich aus wie in Deutschland oder sind die Anforderungsstrukturen in beiden Ländern sehr verschieden? Nellesen und Schüle (2000) merken an, dass zum einen erhebliche Unterschiede in der Gewinnung der Daten und zum anderen in der Interpretation der Ergebnisse bestehen könnten. Sie mahnen an, gemeinsame Abstimmungsarbeiten durchzuführen, um zu einer möglichst einheitlichen Vorgehensweise in der Rehabilitationsbeurteilung zu kommen. Diese Forderung wird umso verständlicher, wenn man sich nur den kleinen Bereich der Codierungskategorien anschaut. In

Tabelle 4 wird deutlich, dass wir momentan von einer einheitlichen Kategorisierung der Leistungsfähigkeit noch weit entfernt sind. Da dieses Problem jedoch bei weitem nicht allen Beteiligten bekannt, geschweige denn bewusst sein dürfte, sind Unschärfen und Missverständnisse in der Interpretation und der Umsetzung der Testergebnisse vorprogrammiert – und die Inhaltsvalidität der EFL leidet.

EFL-Tabelle:					
Gewicht:		Maximal	Schwer	Mittel	Leicht
Zulässige Häufigkeit:	Nie	Gelegentlich	Zeitweise	Überwiegend	Ständig
EFL-Manual (SAR, 1999): Objektive Beurteilungskriterien am Beispiel „Heben Boden- zu Taillenhöhe“					
Gewicht:		Maximal	Schwer	Mittel	Leicht
Zulässige Häufigkeit:		Selten 1-5%	Manchmal 6-33%	Oft 34-66%	Sehr oft 67-100%
EFL-Kursunterlagen (EFL-Akademie, 2001): Objektive Beobachtungskriterien am Beispiel HEBEN Taillenhöhe – Kopfhöhe					
Auch Kaiser et al. (2000): „Kategorien gemäß deutscher EFL-Richtlinien“					
Gewicht:		Maximal	Schwer	Mittelschwer	Leicht
Zulässige Häufigkeit:		Gelegentlich Bis zu 5%	Zeitweise Ca. 10%	Überwiegend 51-90%	Ständig >90%
Vortrag Greitemann, 2004: DOT-Kategorien					
Zulässige Häufigkeit:	Gelegentlich Bis 33%			Häufig Bis 66%	Ständig Bis 100%
Vortrag Greitemann, 2004: VDR-Kategorien					
Zulässige Häufigkeit:	Gelegentlich Bis 5%	Zeitweise 5-10%	Oft 10-50%	Häufig Bis 90%	Ständig Bis 100%

Tabelle 4: Codierungskategorien für funktionelle Leistungsfähigkeit

Konstruktvalidität

Unter Konstruktvalidität versteht man den Nachweis, dass ein Assessment die zugrunde liegenden Konstrukte auch tatsächlich abbilden kann.

Die basale Methode, diese Annahme zu untersuchen, ist ein Vergleich der konvergenten mit der diskriminanten Validität: Korreliert ein Test höher mit inhaltlich

ähnlichen Tests als mit inhaltlich verschiedenen, so spricht das für die Konstruktvalidität des Assessments. Korrelationen zwischen .3 und .6 gelten für konvergente Tests als „mittel bis gut“; Korrelationen über .6 als „gut“ (Innes & Straker, 1999b).

Ein konstruktvalides Instrument sollte zwischen Probandengruppen differenzieren können, deren Unterschiedlichkeit für das Konstrukt relevant ist. So wird von der EFL beispielsweise erwartet, dass sie Probanden, die ein bestimmtes Gewicht sicher heben können, von denen unterscheiden kann, die dazu nicht in der Lage sind.

Darüber hinaus kann man Hinweise auf die Konstruktvalidität finden, indem man aus dem theoretischen Konstrukt abgeleitete Hypothesen überprüft oder in verschiedenen Subgruppen Faktorenanalysen durchführt: Ergeben sich die erwarteten Ergebnisse oder lässt sich immer wieder die gleiche Faktorenstruktur nachweisen, kann dieses als Hinweis auf die Gültigkeit eines Konstruktes und die Adäquatheit seiner Messung verstanden werden.

Die Konstruktvalidität ist nach Innes & Straker (1999b) die breiteste Validitätsart. Sie kann durch Inhalts- und Kriteriumsvalidität gestützt werden und erweitert die Generalisierbarkeit der Ergebnisse über Personen und Situationen.

Innes & Straker (1999b) untersuchten die Validität von 28 FCE-Systemen hinsichtlich der Qualität des Forschungsstandes und des Ausmaßes der Validität. Sie berichten, dass die Konstruktvalidität eines Assessments nur selten evaluiert worden ist; für die EFL ergeben sich einige wenige mäßige bis gute Studien. Die Autoren zitieren eine Untersuchung von Farag (Farag, 1995 nach Innes & Straker, 1999b), nach der in der EFL systematisch geringere Leistungsgrenzen gemessen werden als in einem konkurrierenden psychophysischen Verfahren, ohne die Ergebnisse im Detail zu beschreiben. Innes konnte in einer auf die Hebetests beschränkte Reanalyse der Daten den Befund von Farag nicht bestätigen. Er fand keine Hinweise auf eine Fähigkeit der EFL, zwischen verletzten und nicht verletzten Arbeitern zu unterscheiden.

Wie in Kapitel 2.2 dargestellt, werden FCEs immer wieder dazu genutzt, anhand bestimmter Leistungsdaten die Anstrengungsbereitschaft der Probanden zu ermitteln, ohne dass die Validität dieser Indikatoren überprüft worden wäre. Lemstra et al. (2004) baten in einer verblindeten kontrollierten randomisierten Studie die Hälfte der Probanden, sich in einer FCE maximal anzustrengen. Die andere Hälfte wurde instruiert, sich nur zu 60% anzustrengen, den Testleiter aber ohne verbale Hinweise auf ihre Gruppe glauben zu machen, dass sie sich in der 100%-Gruppe befänden. Die Autoren kamen zu dem Ergebnis, dass nur 5 der 17 Indikatoren, die üblicherweise zur Ermittlung der Anstrengungsbereitschaft eingesetzt werden, in der Lage waren, zwischen Probanden der 60%-Gruppe und Probanden der 100%-

Gruppe zu unterscheiden. Die globale Spezifität lag bei 84,1%, die globale Sensitivität bei 65,2%. Weder Schmerzverhalten noch eine verzerrte Selbstwahrnehmung der eigenen funktionellen Fähigkeiten oder im Voraus erhobene klinische Maße verbesserten oder verschlechterten die korrekte Einordnung in die Gruppen. Die Autoren schlussfolgern, dass man vorsichtig damit sein sollte, Patienten maximalen oder submaximalen Einsatz in FCEs zu unterstellen – zumindest solange, bis Tests oder Testkombinationen mit besserer Sensitivität und Spezifität entwickelt und überprüft worden sind.

Häufiger als Untersuchungen der Fähigkeit der EFL, zwischen relevanten Gruppen zu unterscheiden, sind Studien, in denen Zusammenhänge von EFL-Ergebnissen mit externen Variablen überprüft werden. So verglichen Gross & Battié (2003) die EFL-Ergebnisse von 321 Probanden, die im Schnitt seit 307 Tagen unter arbeitsbezogenen Kreuzschmerzen litten und als medizinisch stabil angesehen werden konnten, mit deren Angaben in zwei unterschiedlichen Schmerzskalen (Pain Disability Index, PDI, und Visuelle Analogskala, VAS). Sie geben als Ergebnis hohe Korrelationen zwischen den Werten in den Schmerzassessments (.079) sowie moderate Zusammenhänge beider zum EFL-Ergebnis (-.44 bis .52 zwischen PDI und EFL-Ergebnis sowie .34 bis .45 zwischen VAS und EFL-Ergebnis) an⁶. Damit überstieg die konvergente tatsächlich die diskriminante Validität und stützte die Annahme einer guten Konstruktvalidität. Allerdings wurde mit diesem Design eher die konvergente Validität der Schmerzassessments als die der EFL gemessen, so dass damit auch nur eine Aussage über die Konstruktvalidität der Schmerzskalen, kaum aber über die der EFL getroffen werden kann. Sicher ist jedoch, dass die der EFL zugrunde liegende Annahme einer Unabhängigkeit von Leistungsfähigkeit und Schmerz nicht bestätigt werden konnte.

Doch nicht nur das Ausmaß des Schmerzes zeigt Zusammenhänge zur funktionellen Leistungsfähigkeit der Probanden in einem EFL-Test, auch die daraus resultierenden subjektiven schmerzbedingten Behinderungen spielen nach Reneman, Jorritsma et al. (2002) eine Rolle. Sie wiesen Korrelationen des EFL-Ergebnisses von -.20 (n.s.) zum Roland Disability Questionnaire, von -.50 ($p < .001^{**}$) zum Quebec Back Pain Disability und von -.52 ($p < .001^{**}$) zum Oswestry Disability Questionnaire nach.

Ebenso berichten Gross und Battié von einer Studie von Geisser et al. (2003) in der Zusammenhänge von FCE- (wenn auch nicht EFL-) Ergebnissen zu arbeitsbezogenen Wahrnehmungen wie Arbeitszufriedenheit und Unterstützung am Arbeitsplatz gefunden wurden. Darüber hinaus belegten Lackner et al. (1996) mit Korrelationen zwischen .39 ($p < .01^{**}$) und .67 ($p < .001^{**}$) den prädiktiven Wert von

⁶ Es erscheint unplausibel, dass die Korrelationen zwischen den EFL-Ergebnissen und den Schmerzbögen unterschiedliche Richtungen haben. Stattdessen wären eher einheitlich negative Zusammenhänge zu erwarten gewesen. Leider liegt uns nur das Abstract zu dieser Arbeit vor, so dass wir die dort angegebenen Werte nicht im vollständigen Artikel überprüfen können.

aufgabenbezogenen Selbstwirksamkeitsüberzeugungen in Bezug auf ihre Leistung in funktionellen Tests des Hebens, Tragens, Ziehens und Schiebens.

Befunde von Gross und Battié (2005) stützen die Vermutung, dass die EFL tatsächlich körperliche Leistungsfähigkeit misst: Sie fanden heraus, dass jüngere Probanden erwartungsgemäß mehr Gewicht vom Boden auf Taillenhöhe heben konnten als ältere. Ebenso konnten Männer mehr absolutes Gewicht heben als Frauen. Da diese Probanden jedoch in der Regel auch die körperlich schwereren Arbeiten verrichten, gab es keine Zusammenhänge mehr zwischen Alter oder Geschlecht auf der einen und der Anzahl der im Vergleich zu den Arbeitsanforderungen bestandenen Tests auf der anderen Seite.

Es gibt also Hinweise darauf, dass die EFL durchaus in der Lage ist, körperliche Leistungsfähigkeit zu erfassen, ihr Ergebnis aber auch von weiteren, in der Regel eher psychischen oder psychosozialen als körperlichen Faktoren abhängt.

Insgesamt kommen Innes und Straker (1999b) in ihrem Review zu dem Schluss, dass die Konstruktvalidität der EFL bisher nicht als ausreichend bestätigt eingestuft werden kann. Unserer Kenntnis nach gilt dieses bis heute. Gross und Battié ziehen daraus Konsequenzen für die Konzeption der EFL und damit auch für die Interpretation der Ergebnisse:

“Functional capacity evaluations should be considered behavioral tests influenced by multiple factors, including physical ability, beliefs, and perceptions.”

(Gross & Battié, 2005, S. 321)

Auch Rivier und Seewer (2002) bezweifeln angesichts der Befundlage, dass die EFL eine situationsübergreifende funktionelle Leistungsfähigkeit des Probanden abbildet, und begrenzen ihre Aussagefähigkeit auf

„eher die Leistungen, die er bereit war zu erbringen, und dies an einem bestimmten Tag und zu einer bestimmten Zeit“.

(Rivier & Seewer, 2002, S. 39)

Korrekterweise müsste man in diesem Fall statt von einer „*Functional Capacity Evaluation*“ von einer „*Functional Performance Evaluation*“ sprechen und konstatieren, dass die EFL ihrem Anspruch auf situationsübergreifende Aussagekraft nicht gerecht wird.

Kriterienorientierte Validität

Die kriterienorientierte Validität eines Tests gibt darüber Auskunft, wie gut die Testergebnisse mit einem relevanten äußeren Kriterium korrelieren. Zur Ermittlung der konkurrenten Validität wird der in Frage stehende Test ungefähr gleichzeitig mit einem anderen, bereits validierten Assessment durchgeführt, so dass beide eine Aussage zum gleichen Merkmal treffen. Dieses Verfahren wird häufig genutzt, um ein neues Messinstrument am so genannten Gold-Standard zu validieren.

Wird hingegen zuerst der Test durchgeführt und dann in einem angemessenen zeitlichen Abstand das Kriterium erhoben, so spricht man von prädiktiver Validität. Sie ist besonders wichtig für klinische Assessments und Entscheidungen. Im Falle der EFL geht es in der Regel um ihre Fähigkeit, eine sichere Rückkehr des Probanden zur Arbeit zu prognostizieren. Ein gutes EFL-Ergebnis sollte also mit einer schnellen und dauerhaften Wiederaufnahme einer Erwerbstätigkeit in Zusammenhang stehen⁷.

Innes und Straker (1999b) stellten in ihrem Review fest, dass es Belege für eine zumindest mittlere kriteriumsorientierte Validität für einige FCE-Systeme gibt, die EFL gehörte jedoch nicht dazu. Für die EFL berichten sie lediglich zwei Untersuchungen von Isernhagen et al., in denen zwar Return-to-work-Quoten von EFL-Probanden angegeben worden wären, jedoch keine statistische Auswertung erfolgt sei, auf deren Basis man eine wirkliche Aussage über die kriterienbezogene Validität der EFL treffen könne (Isernhagen, 1995; Isernhagen 1996; beide nach Innes & Straker, 1999b). Auch King et al. (1998) sowie Rivier und Seewer (2002) halten die prädiktive Validität der EFL für noch nicht ausreichend nachgewiesen und fordern weitere Forschung in diesem Bereich. Inzwischen haben Matheson et al. (2002) Ergebnisse veröffentlicht, die einen Zusammenhang zwischen der Hebeleistung von Probanden mit chronischem Rückenschmerz im EFL-Test und ihrer Wahrscheinlichkeit, innerhalb von 6 Monaten wieder in das Erwerbsleben zurückzukehren, belegen. Zu vergleichbaren Ergebnissen kamen auch Fishbain et al. (1999) für die „Dictionary of occupational titles residual functional capacity battery“, einem der EFL recht ähnlichen FCE-System. Allerdings wurden in beiden Studien weitere personen- und kontextbezogene Faktoren gefunden, die mindestens in ähnlicher Höhe wie die EFL-Hebeleistung mit der Rückkehr zur Arbeit korrelierten.

Gross et al. (2004) kritisieren sogar die Auswahl der Prädiktorvariablen in diesen Studien. In den bisherigen Studien wurde versucht, aus der absoluten Hebeleistung den Zeitpunkt und die Sicherheit einer Rückkehr zur Arbeit vorherzusagen. Nach der Arbeitsgruppe muss aber nicht jeder Proband für eine Rückkehr in seinen Beruf gleichermaßen schwer heben können – entscheidend ist vielmehr, dass seine Fähigkeiten für die Anforderungen an seinem individuellen Arbeitsplatz ausreichend sind.

⁷ Allerdings hängt diese natürlich nicht nur von den funktionellen Fähigkeiten des Patienten ab, sondern auch von der Verfügbarkeit einer leidensgerechten Tätigkeit!

Auf dem Hintergrund dieser Überlegungen überprüften die Wissenschaftler in ihrer Studie mit 246 Arbeitern mit chronischen Rückenschmerzen die Vorhersagekraft von drei Indikatoren einer guten EFL-Leistung für den Zeitpunkt und die Sicherheit der Wiederaufnahme ihrer letzten Tätigkeit. Diese Variablen waren:

- Die absolute Leistung im Test „Heben Boden- zu Taillenhöhe“
 - ⇒ Wie viel Kilogramm kann der Proband vom Boden zur Taillenhöhe maximal heben?
- Das Bestehen aller Arbeitsanforderungen
 - ⇒ Zeigt der Proband in allen EFL-Subtests Fähigkeiten, die mindestens den Anforderungen an seinem Arbeitsplatz entsprechen?
- Das Bestehen der einzelnen Arbeitsanforderung „Heben Boden- zu Taillenhöhe“
 - ⇒ Zeigt der Proband in dem einzelnen EFL-Subtest „Heben Boden- zu Taillenhöhe“ eine Fähigkeit, die mindestens der Anforderung an seinem Arbeitsplatz entspricht?

Im Ergebnis gingen alle drei Variablen mit einer schnelleren Wiederaufnahme der Arbeit einher, klärten nach Kontrolle aller potenziellen konfundierenden Variablen jedoch nur noch etwa weitere 10% der Varianz auf. Verblüffend war, dass Patienten, die mehr Tests bestanden hatten, innerhalb eines Jahres häufiger aufgrund von Rezidiven erneute Ansprüche an ihre Versicherung stellten als jene mit einer geringeren Anzahl bestandener Tests. Nur 4% der Probanden konnten laut EFL alle kritischen Arbeitsanforderungen meistern (streng genommen hätten nur diese Probanden an den Arbeitsplatz zurückkehren dürfen!), von ihnen erlitt keiner einen ernsthaften Rückfall. Als Prädiktoren für Rezidivierung gänzlich ungeeignet waren sowohl die absolute Hebeleistung als auch das Bestehen oder Nicht-Bestehen des Tests Heben Boden- zu Taillenhöhe.

Die Autoren schlussfolgern, dass der prädiktive Wert der EFL für die Vorhersage von Zeitpunkt und Sicherheit einer Rückkehr an den Arbeitsplatz eher gering bis nicht vorhanden ist und ziehen das Fazit:

"In the case of FCE, predictive validity has been considered a key psychometric property, as predicting ability to resume and sustain future work is a primary use. Our earlier findings and those of others indicate FCE adds modestly to the prediction of initial return to work beyond that provided by other prognostic factors. Our current results are the first related to prediction of sustained recovery and are contrary to FCE's theoretical basis regarding which clients are likely to avoid future recurrence. Case managers, health care professionals, and other users of FCE information will have to decide whether the limited validity and usefulness of the information obtained with FCE outweighs the administrative burden and costs associated with such testing."

(Gross & Battié, 2004, S. 923)

Eine mögliche Erklärung für die schwache Vorhersagekraft der EFL ist der Schnappschuss-Charakter des EFL-Ergebnisses. Selbst wenn medizinische Stabilität, also das Erreichen eines Krankheitsplateaus, zu den Indikationsvoraussetzungen gehört, ist eine Einpunktmessung, wie sie jede FCE darstellt, kritisch zu sehen. Sie berücksichtigt weder normale Schwankungen der Tagesform noch gerichtete längerfristige Veränderungen des Beschwerdebildes oder Prognosen (Die Kölnische Rück, 1999). Abdel-Moty et al. (1993) weisen auf die Unterscheidung von „limitations“ (das, was ein Patient nicht tun *kann*) und „restrictions“ (das, was ein Patient nicht tun *darf*) hin. Beide Einschränkungen müssen in einer verantwortungsvollen sozialmedizinischen Einschätzung angemessen berücksichtigt werden. Bedenkt man dieses ebenso wie die Beschränkung der EFL-Ergebnisse auf die Erfassung von Performance anstelle von Capacity, erstaunt die begrenzte prädiktive Leistung der EFL kaum noch.

Umso wichtiger ist der Hinweis von Strong et al. (2004), dass sich die Nutzer von EFL-Ergebnissen häufig nicht über diese Grenzen des Tests im Klaren sind. Beispielsweise beziehen sie sich noch lange nach dem Test auf dessen Ergebnisse, obwohl völlig unklar ist, über welchen Zeitraum diese als valide gelten können.

Auf dem Hintergrund unzureichender Werte für Reliabilität und prädiktiver Validität der EFL kritisiert Gatty (2002) das gebräuchliche Vorgehen, FCEs nur einmal durchzuführen und das Ergebnis ohne genügend flankierende Untersuchungen als Basis für Entscheidungen zu nehmen. Dieses ist tatsächlich ein Problem, zumal nicht nur die FCE-Leistung, sondern auch die arbeitsbezogene Leistungsfähigkeit des Patienten von den unterschiedlichsten körperlichen, psychosozialen und umgebungsbedingten Faktoren abhängt. Umso wichtiger ist es nach unseren Erkenntnissen für den behandelnden Arzt, von dem einzelnen Patienten zu erfahren, welchen zusätzlichen Belastungen er im Alltag ausgesetzt ist, aber auch, welche Ressourcen und Unterstützungsmöglichkeiten er zur Verfügung hat. Nur dann kann die sozialmedizinische Beurteilung gut auf die Realität des Patienten abgestimmt und eine langfristig valide Aussage getroffen werden.

3.5 Praktikabilität

Die Frage nach der Praktikabilität der EFL im Klinikalltag lässt sich mit einem Blick auf die dafür nötigen Voraussetzungen auf Seiten der Einrichtung und der potenziellen Probanden beantworten.

Auf Seiten der Einrichtung fallen vor allem die monetären Kosten und der zeitliche Aufwand für eine EFL ins Auge. Die hohen Qualitätsstandards bringen es mit sich, dass erst das Durchlaufen einer speziellen Ausbildung durch mindestens einen Arzt und einen Physio-, Ergo- oder Sporttherapeuten zum Erwerb von Lizenzen für die

Durchführung von EFLs berechtigt. Geeignetes Personal muss also zunächst identifiziert und anschließend qualifiziert werden. Der Erwerb der nötigen zwei Lizenzen und der erforderlichen Ausstattung schlägt mit ungefähr 20.000 Euro zu Buche. Diese Summe klingt nach einer erheblichen Investition, ist jedoch im Vergleich beispielsweise zu den etwa 10fachen Anschaffungskosten eines ERGOS-Systems eher gering (Greitemann, 2004).

Weitere laufende Kosten beschränken sich auf den Einsatz des Personals, wobei der Zeitaufwand mit jeweils etwa 3-4 Stunden Durchführungsdauer plus ca. 3 Stunden Aufwand für das Erstellen des Berichts inklusive Supervision durch den EFL-Arzt relativ hoch ist. Dieses schränkt die Praktikabilität sicherlich ein und führt in Einrichtungen, die nicht auf die Durchführung von EFLs spezialisiert sind, vermutlich zu einer eher sparsamen Verordnung, zumal der Test oft nicht speziell vergütet wird, wenn er im Rahmen einer Reha stattfindet (Erbstößer et al., 2003). Als Auftragsarbeit werden dem Auftraggeber ca. 850 Euro in Rechnung gestellt (Greitemann, 2004).

Auf Seiten der Patienten sind nur wenige Einschränkungen in der Anwendbarkeit festzumachen. Aufgrund ihrer breiten Konzeption kann die EFL im Prinzip in der orthopädischen und zum Teil der neurologischen Rehabilitation vielfältig eingesetzt werden, sowohl im stationären als auch im ambulanten Setting. Sie ist nicht auf Probanden bestimmter Krankheitsbilder oder beruflicher Ausrichtungen beschränkt (Rivier & Seewer, 2002), und auch der kulturelle Hintergrund spielt zunächst einmal keine Rolle.

Dieses zunehmend wichtige Charakteristikum konnte durch Oesch (2000) belegt werden. Er zeigte, dass Menschen mit Migrationshintergrund (Ex-Jugoslawien, Italien/Spanien/Portugal, Türkei) als Gesamtgruppe zwar signifikant weniger Last heben, geringere Handkraftwerte aufweisen und sich im PACT weniger zutrauen als Menschen aus der Schweiz, Österreich und dem Fürstentum Liechtenstein. Allerdings weisen sie auch häufiger inkonsistente Resultate auf, die als Zeichen für eine schlechtere psychosoziale Situation gewertet werden können. Schaut man sich nur die Ergebnisse der Probanden mit konsistenten EFL-Ergebnissen an, so unterscheiden sich die Menschen mit Migrationshintergrund in ihrer Hebeleistung nicht mehr signifikant von den „Einheimischen“. So ist laut Oesch mit der EFL eine individuelle Beurteilung der Leistungsfähigkeit unter Berücksichtigung der klinischen Präsentation und der Konsistenz der Befunde, jedoch unabhängig von der Nationalität möglich.

	CH/FL/A n=91	EX-J n=54	I/E/P n=34	Türkei n=12	CH/FL/A vs. Ex-J	Ex-J. vs. I/E/P
Inkonsistent (% Fälle)	23%	63%	62%	42%		
PACT	122	75	105	110	p<.0005**	p=.002**
Gehobene Last (kg)						
Alle	21	15	18	20	p=.001**	n.s.
Konsistente	22	19	21		n.s.	
Handkraft (kg)						
Alle	39	31	33	35	p=.004**	n.s.
Konsistente	40	39	39		n.s.	

Tabelle 5: Vergleich der EFL-Resultate von Probanden aus verschiedenen Heimatländern (nach Oesch, 2000)

Insgesamt präsentieren sich die soziodemographischen und beruflichen Basisdaten des Probanden nicht als bedeutsam für eine Indikationsstellung zur EFL. Entscheidend sind vielmehr medizinische Stabilität und eine ausreichende Belastbarkeit des Probanden wie auch ein hinreichendes Verständnis für den Test und ein entsprechendes Maß an Compliance (vgl. Kapitel 2.3).

3.6 Nützlichkeit

Nach Lienert & Raatz (1994) ist ein Test ist dann nützlich,

„wenn er ein Persönlichkeitsmerkmal oder eine Verhaltensweise misst oder vorhersagt, für dessen Untersuchung ein praktisches Bedürfnis besteht. Ein Test hat demgemäß eine hohe Nützlichkeit, wenn er in seiner Funktion durch keinen anderen Test vertreten werden kann, und er hat eine geringe Nützlichkeit, wenn er ein Persönlichkeitsmerkmal prüft, das mit einer Reihe anderer Tests ebenso gut untersucht werden kann.“

(Lienert & Raatz, 1994, S. 13).

Konkurrenzverfahren zu einem FCE-Test wären im Rahmen der orthopädischen Rehabilitation zum einen die üblichen diagnostischen Methoden, die als unzuverlässig angezweifelt werden (vgl. Einleitung), und zum anderen Arbeitserprobungen. Diese dauern jedoch im Vergleich zu Testverfahren deutlich länger und sind aufgrund des höheren Aufwandes nicht nur erheblich teurer, sondern lassen sich auch schlechter in eine zeitlich sehr begrenzte Rehabilitationsmaßnahme

integrieren. Ein weiterer Vorteil von FCEs gegenüber Arbeitserprobungen besteht sicherlich in der breiteren Anwendbarkeit – sie werden nicht nur für eine Return-to-work-Prognose eingesetzt, sondern sollen auch auf verschiedenste andere Fragen Antworten geben. Ein Beispiel ist eine Art Bestandsaufnahme der funktionellen Fähigkeiten zur Planung und gegebenenfalls Evaluation der rehabilitativen Maßnahmen; diese könnte mit anderen Mitteln vermutlich kaum in gleicher Qualität geschehen.

Um zu beurteilen, wie gut eine FCE den Bedürfnissen der Beteiligten entspricht, liegt es nahe, die Einschätzungen von ihren Auftraggebern, Anwendern und Probanden zu erheben. Genau dieses haben Strong et al. (2004) getan und 68 FCE-Probanden sowie 81 Nutzer der Ergebnisse⁸ unter anderem gefragt, wie sie den Nutzen von FCE-Berichten bewerten. Beide Seiten fanden FCEs ertragreich⁹. So antworteten 2 Monate nach der FCE 48,5% der Probanden und weitere 6 Monate später noch 42% der Probanden, dass die FCE sehr nützlich oder nützlich für sie gewesen sei. Als Gründe dafür gaben sie an, dass der Test ihnen Wissenswertes über ihre Fähigkeiten und Grenzen gezeigt sowie die Einleitung von Aktivität und Veränderung erleichtert habe. Weitere Gründe waren die Bereitstellung substanzieller Informationen für die Zahlung finanzieller Leistungen, Unterstützung bei Widerspruchsverfahren und Hilfe in der Krankheitsbewältigung. Gewisse Vorbehalte gegenüber der Methode und der Verwendung der Ergebnisse hatten jedoch fast alle, verbunden mit Wünschen und Verbesserungsvorschlägen an die Arbeitgeber und die FCE-Dienstleister. Schwerpunkte dabei waren neben einem verantwortungsbewussten Einsatz von FCEs mehr Transparenz im Hinblick auf die Ergebnisse und deren Verwendung.

Die Einschätzungen der Nutzer fielen noch positiver aus als die der Probanden: Von ihnen schätzten 89% die FCE als nützlich oder sehr nützlich ein. Sie verwendeten die Ergebnisse einzelner FCEs teilweise mehrfach zur Beantwortung unterschiedlichster Fragestellungen, ohne dass dieses vorher in der Überweisung an die durchführende Einrichtung gestanden hätte. Offiziell bezog sich etwa die Hälfte der FCEs auf ein Job-match. In 12 Fällen sollte ein globales Screening durchgeführt werden, bei 17 Probanden stand die Planung der beruflichen Zukunft im Mittelpunkt. Ein besonders großer Nutzen wurde den FCEs für die Planung der Rückkehr an den Arbeitsplatz zugeschrieben. Ebenso wurden sie als gute Hilfe erlebt, alle Beteiligten zu Verhandlungen über das weitere Vorgehen an einen Tisch zu holen. Der umstrittene Einsatzbereich der Abschätzung von Anstrengungsbereitschaft, Einstellung und Compliance wurde von 16 der 50 als erfahren eingestuften Nutzer der Ergebnisse als bedeutsam genannt. Strong et al. betonen, dass sich nur wenige Nutzer der begrenzten empirischen Belege der prädiktiven Validität isolierter FCE-Ergebnisse im

⁸ darunter vor allem Angehörige des Workplace Safety and Insurance Board Ontario und verschiedene Arbeitgeber

⁹ In Bezug auf die Nutzer erstaunt das nicht - sonst würden sie ja kaum FCEs in Auftrag geben.

Hinblick auf die Arbeitsleistung der Probanden bewusst waren. Umso erstaunlicher ist es, dass trotzdem viele die FCEs dann als besonders nützlich einstufen, wenn ihre Ergebnisse durch die Resultate anderer Tests ergänzt werden.

Weitere Hinweise auf die erlebte Nützlichkeit insbesondere von EFL und ERGOS bietet die Studie von Erbstößer et al. (2003). Sie leisteten eine Bestandsaufnahme der FCE-Systeme in Deutschland und vertieften erste Erkenntnisse mit Hilfe weiterer Befragungen der Einrichtungen, die bereits EFL oder ERGOS anwendeten, sich in der Schulungsphase befanden oder eine Anschaffung planten.

In der Rückschau gaben die EFL-Anwender an, sich für diese Investition entschieden zu haben, um nicht nur ihre Reha-Maßnahmen allgemein und ihre Diagnostik der arbeitsbezogenen Leistungsfähigkeit sowie der Körperfunktionen im Besonderen zu verbessern, sondern auch ihr Leistungsangebot zu erweitern. Weiterhin waren sowohl Qualitätssicherung als auch eine Optimierung der Erfolgskontrolle von Reha-Maßnahmen wichtige Argumente. Etwas weniger bedeutsam waren die Erweiterung der Kompetenz durch Forschung und Entwicklung sowie Marketing und Werbung. ERGOS- und EFL-Anwender haben sich gleichermaßen für ihr System entschieden, weil sie der Überzeugung waren, dass das System aussagefähig, objektiv und realitätsbezogen ist. Weiterhin ermögliche es den Vergleich von Leistungsprofil und Anforderungsprofil. Man können neben den arbeitsbezogenen funktionalen Fähigkeiten auch Faktoren wie Selbstlimitierung, Leistungsbereitschaft, Schmerzverhalten etc. durch Verhaltensbeobachtung detailliert erkennen. Weiterhin liefere das System wichtige Basisdaten zur weiteren beruflichen Maßnahmenplanung. Speziell EFL-Anwender betonten, dass die EFL einfach, flexibel und kostengünstig sei. Sie könne sowohl in der Planung als auch in der Durchführung der Therapie eingesetzt werden, ermögliche ein detailliertes Erkennen allgemeiner funktionaler Fähigkeiten und liefere wichtige Basisdaten zur weiteren medizinischen Behandlungsplanung. Viele EFL-Anwender und Schuler, die in dieser Studie befragt wurden, sehen die EFL als einen Baustein unter mehreren im Rahmen einer Beurteilung – knapp 50% von ihnen setzten weitere Test- oder Dokumentationssysteme ein, wobei ergänzende Tests der geistigen, psychischen oder sozialen Kompetenzen im Vordergrund standen.

Die Anwender beider Systeme nannten neben der Diagnostik die Planung, Durchführung und Evaluation der Therapie, die Prävention und die Veränderung der Selbsteinschätzung der Probanden hin zu einer realistischeren Vorstellung als bedeutsame Einsatzbereiche. Die häufigsten Fragestellungen betrafen die Eignung für den bisherigen Beruf, die Frage der Neuorientierung und den Vergleich von Anforderungs- und Leistungsprofil. Dabei wurde der EFL für Industriebereiche eine höhere Gültigkeit zugeschrieben als für (einige) Angestelltenberufe, ohne dass diese Annahme empirisch abgesichert worden wäre.

Auf der Basis des theoretischen Hintergrundes, des jeweiligen Standes der Forschung und zum Teil eigener Erfahrungen mit der EFL haben sich auch Kaiser et al. (2000) und Rivier & Seewer (2002) mit den Charakteristika der EFL auseinandergesetzt, die eine große Nähe zur Frage der Nützlichkeit aufweisen. In den Artikeln wird besonders hervorgehoben, dass die EFL

- standardisiert ist, so dass Ergebnisse besser reproduzierbar sind.
- eine Quantifizierung der Leistungsfähigkeit ermöglicht.
- eine umfassende, systematische Leistungsevaluation darstellt.
- arbeitsrelevante, realistische Anforderungen an den Probanden stellt.
- die Beobachtung von Bewegungsmustern und Verhalten ermöglicht.
- ambulant durchgeführt werden kann.
- mit dem lokalen medizinischen und versicherungsmäßigen System sowie den Arbeitsplätzen vor Ort verbunden ist.
- durch berufsspezifische Ergänzungstests individuell adaptiert werden kann.
- eine Schätzung der Kohärenz der Leistungen erlaubt.

Oesch (2000) weist zusätzlich darauf hin, dass die EFL

- eine individuelle Beurteilung der Leistungsfähigkeit unter Berücksichtigung der klinischen Präsentation und der Konsistenz der Befunde unabhängig von der Nationalität des Probanden ermöglicht.

Von vielen Autoren wird die EFL grundsätzlich als nützlich eingestuft, auch Rivier und Seewer ziehen insgesamt eine positive Bilanz:

"Eine EFL erlaubt es, ein Band zwischen zwei Welten zu spannen, die wenig Gemeinsames haben: Der Welt der Arbeit, die auf dem Grundsatz der Produktivität beruht, und der Welt der Medizin, deren vorrangiges Ziel die Behandlung einer funktionellen Störung oder körperlichen Schädigung ist. Eine EFL integriert die medizinische Komponente, wobei sie gleichzeitig dem getesteten Individuum einen hohen Sicherheitsgrad der Testung garantiert. Sie vermeidet aber die Fokussierung auf die Beeinträchtigungen (z.B. statische Störungen), welche bei der klinischen Untersuchung oder bei der Bildgebung festgestellt worden sind. [...]"

Die bescheidene Erfahrung der Autoren dieses Artikels (bis heute sind lediglich 31 eigene Evaluationen durchgeführt worden) ist bisher positiv, falls die Indikation einer solchen Evaluation direkt zwischen dem Auftraggeber und dem Arzt, der die Untersuchung supervisiert, diskutiert wird, und wenn der Auftraggeber ein detailliertes Profil der wichtigsten Aufgaben am Arbeitsplatz des zu untersuchenden Klienten liefert."

(Rivier & Seewer, 2002, S. 40)

Wie aus den referierten Beiträgen deutlich wird, basiert die Auseinandersetzung mit der Nützlichkeit des Verfahrens in erster Linie auf Erfahrungen und Überlegungen. Eine empirische Studie, die sich mit den Fragen beschäftigt, welche zusätzliche Information die EFL den Ärzten für ihre sozialmedizinische Einschätzung tatsächlich bringt, oder ob die Probanden ihre eigene Leistungsfähigkeit nach einer EFL realistischer einschätzen können, steht nach unserer Kenntnis bisher noch aus.

3.7 Zusammenfassung: Testgüte der EFL nach dem augenblicklichen Stand der Literatur

Wird die EFL verantwortungsbewusst durchgeführt, scheint die **Sicherheit** des Probanden während der EFL weitgehend gegeben zu sein. Berichte über ernsthafte Verletzungen haben wir in der Literatur nicht gefunden. Gerade bei Patienten, die bereits länger inaktiv sind, muss jedoch mit einer deutlichen Zunahme von Beschwerden gerechnet werden. Es ist nicht ausgeschlossen, dass diese zu Motivations- beziehungsweise Complianceproblemen führen und Bewegungsängste verstärken können.

Entscheidend für die Sicherheit der Probanden sind nach unserem Kenntnisstand eine solide Ausbildung und umfangreiche Erfahrung des Testleiters. Er sollte eine gute und vertrauensvolle Beziehung zum Probanden aufbauen können und den einzelnen Probanden mit differenziertem Blick beobachten. Gerade in kinesiophysischen Tests, in denen der Proband einen großen Teil der Kontrolle an den Versuchsleiter abgibt, erscheint uns ein sensibler Umgang mit dem Probanden unabdingbar.

Die **Objektivität** der EFL scheint aufgrund des hohen Standardisierungsgrades gewährleistet zu sein. Das Ergebnis ist aber stark von der Beobachtungsgabe und der Interpretation des Testleiters – und damit von erheblichen subjektiven Einschätzungen – abhängig. Maßgeblich für eine Maximierung der Objektivität der EFL ist die strenge Orientierung des Testleiters an tatsächlichen Beobachtungen bei einer Minimierung persönlicher Interpretation. Es wurde angeregt, zur Steigerung der Objektivität und damit auch der Reliabilität die Testdurchführung per Video aufzunehmen und auf diese Weise das Ergebnis von der aktuellen Konzentration und dem einmaligen Urteil des Testleiters unabhängiger zu machen.

Die Forschungssituation zur **Reliabilität** der EFL wie auch der meisten anderen FCE-Systeme wird als eher mäßig angesehen. Aufgrund der Notwendigkeit, den Test im Sinne seiner Validität möglichst gut an die Realität des Probanden anzupassen,

kann die EFL nicht maximal standardisiert werden und damit auch nicht maximal reliabel sein. Als Varianzquellen werden alle Beteiligten gesehen: der Proband, der zu den beiden Testdurchläufen gegebenenfalls verschiedene Ausgangsbedingungen beispielsweise hinsichtlich seines Befindens und seiner Motivation mitbringt („Tagesform“), der Testleiter, dessen subjektive Wahrnehmung und Interpretation des Gesehenen das Ergebnis maßgeblich bestimmt, sowie die Interaktion zwischen beiden. Trotz dieser vielfältigen Varianzquellen konnten in verschiedenen Studien relativ gute Interrater und Retest-Reliabilitäten nachgewiesen werden – die allerdings fast nur für Hebetests bestimmt worden sind.

Für eine gute Reliabilität der Ergebnisse ist es wie für eine hohe Objektivität wichtig, dass sich die Testleiter möglichst streng an das Manual halten und ihr Urteil auf echte Beobachtungen statt auf Interpretation stützen. Die Anregung, die Durchführung auf Video aufzunehmen, könnte dabei durchaus hilfreich sein. Ein wichtiges Charakteristikum der EFL gegenüber vielen anderen FCE-Systemen für die Reliabilität der Testergebnisse ist die Wiederholung der wichtigsten Tests am zweiten Tag und die Verrechnung der beiden erzielten Werte, die den Einfluss starker Tagesformschwankungen auf das Gesamtergebnis erheblich reduzieren. Dieser Aspekt sollte keinesfalls aufgegeben werden.

Auch die Datenlage zur **Validität** der EFL ist zurzeit noch sehr begrenzt, im Gegensatz zu anderen FCE-Systemen liegen aber immerhin einige Statements und Studien vor.

Die **Face validity** ist für die Absicht, körperliche Leistungsfähigkeit zu messen, hoch. Für das Ziel, eine Aussage über eine sichere Rückkehr zur Arbeit zu treffen, wird sie deutlich geringer eingeschätzt, denn dafür scheinen die Testbedingungen vielen Beteiligten zu alltagsfern. Gründe dafür liegen vor allem in der starken Standardisierung der EFL sowie vermutlich dem Versuch, den Aufwand und damit die Kosten in Grenzen zu halten.

Die **Inhaltsvalidität** der EFL wird oft aufgrund ihrer DOT-Basierung als hoch angesehen. Einige Autoren bezweifeln aber, dass dieses Kriterium angemessen und ausreichend ist. Zieht man die EFL als Messinstrument für die Fähigkeit eines Probanden, schnell und sicher eine Erwerbstätigkeit (wieder) aufzunehmen, heran, finden sich Einschränkungen der Inhaltsvalidität vor allem in der Alltagsferne der Testbedingungen, in der Extrapolation der Ergebnisse aus relativ kurzfristigen Maximalbelastungen auf die Belastungen einer vollzeitigen Tätigkeit sowie in der Konfundierung von Belastungsintensität und Belastungsdauer. Auch die Übertragbarkeit der aus den USA stammenden Konzeption des Tests auf deutsche Verhältnisse ist noch nicht überprüft, könnte jedoch erheblichen Einfluss auf die Inhaltsvalidität haben.

Die **Konstruktvalidität** der EFL wurde nur selten in Studien überprüft. Darin zeigt sich eine sehr begrenzte Fähigkeit der EFL, zwischen verschiedenen theoretisch

relevanten Gruppen zu unterscheiden. Allerdings haben sich durchaus erwartete Differenzen in den Korrelationen zu anderen mehr oder weniger verwandten Verfahren nachweisen lassen. Empirische Widersprüche zum theoretischen Ansatz ergeben sich aus Befunden, die Zusammenhänge zwischen dem EFL-Ergebnis und Schmerz sowie weiteren eher psychosozialen und motivationalen als körperbezogenen Variablen aufzeigen.

Als Konsequenz daraus muss nach unserer Überzeugung bei der Interpretation der Testergebnisse bedacht werden, dass die EFL nicht wie ursprünglich angenommen die reine körperliche Leistungsfähigkeit („Capacity“) misst. Stattdessen wird offenbar die Leistung erfasst, die ein Proband zu einer bestimmten Zeit unter bestimmten Bedingungen zu erbringen bereit ist („Performance“). Diese wird durch multiple Faktoren, vor allem durch körperliche Fähigkeiten, Überzeugungen und Wahrnehmungen beeinflusst.

Die **Kriterienorientierte Validität** der EFL orientiert sich in der Regel an dem Prüfstein der erfolgreichen Rückkehr ins Arbeitsleben. In einigen Studien sind signifikante Zusammenhänge zwischen guten EFL-Leistungen und einer schnelleren – allerdings nicht unbedingt sichereren – Rückkehr in das Erwerbsleben nachgewiesen worden. Diese klären unter Kontrolle konfundierender Variablen jedoch nur noch wenig zusätzliche Varianz auf, so dass fraglich ist, ob sich der relativ hohe Aufwand einer EFL für eine solche Prognose lohnt. Möglicherweise begründen sich diese relativ schwachen Werte der EFL auf die doch erheblichen Unterschiede zwischen Konzeption und Bedingungen des Tests im Vergleich zur Realsituation. Diese Kritik bezieht sich ebenso auf die Konzeption der EFL als Einpunktmessung wie auf die mangelnde Unterscheidung zwischen „limitations“ (das, was ein Patient nicht tun *kann*) und „restrictions“ (das, was ein Patient nicht tun *darf*) und die zum Teil wenig realistische Nachbildung der individuellen Arbeitsbedingungen des Probanden.

Im Klinikalltag scheint die EFL in gewissen Grenzen durchaus **praktikabel**. Vor allem aufgrund des hohen Zeitbedarfs wird jedoch vermutlich eher restriktiv verordnet und die Verordnung oft auf besondere Zielgruppen, beispielsweise potenzielle Umschüler, fokussiert.

In der Literatur zeigt sich, dass eine ganze Reihe von Forschern und Anwendern die EFL durchaus für **nützlich** halten. Insbesondere die ihr zugeschriebene Testgüte führt zu einer positiven Bewertung, obwohl diese, wie ein genauerer Blick auf die Befundlage zeigt, noch gar nicht in allen Aspekten eindeutig nachgewiesen ist. Empirische Studien zum Nutzen der EFL vor allem für Ärzte, die mit sozialmedizinischen Einschätzungen ihrer Patienten befasst sind, und für die Probanden selbst, stehen noch aus.

Dieser Überblick über bisherige Erkenntnisse über die Güte der EFL wirkt gegenüber den hohen Erwartungen an ihre Aussagekraft ziemlich ernüchternd. Trotzdem kann und soll ein eigentlich ungenügend abgesicherter Test durchgeführt werden, wenn sein praktischer Nutzen den Aufwand übersteigt:

„Dennoch lässt sich der Einsatz eines [...] Tests generell pragmatisch rechtfertigen, wenn die Entscheidungen und Vorhersagen, die auf der Basis des Tests getroffen werden, tauglicher sind als Entscheidungen und Vorhersagen, die ohne den Test möglich wären – es sei denn, der mit dem Test verbundene Aufwand steht in keinem Verhältnis zum Informationsgewinn.“

(Bortz & Döring, 2006, S. 202)

Daraus ergibt sich als neue Kernfrage:

☞ Rechtfertigt der Nutzen von EFL-Tests deren Einsatz für stationäre orthopädische Reha-Patienten?

4. Ziele und Fragestellungen der Studie

Ziel der vorliegenden Studie war es, der Forderung der Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung – SOMEKO – (2004) entsprechend auf dem Hintergrund des dargestellten Forschungsstandes den Nutzen von EFL-Tests im Einsatz bei stationären orthopädischen Reha-Patienten zu überprüfen und ihren Stellenwert in der sozialmedizinischen Einschätzung zu beleuchten.

4.1 Nutzen für die ärztliche Einschätzung

FCE-Systeme sollen in erster Linie Ärzten dazu dienen, zu einer valideren Einschätzung der Leistungsfähigkeit der Probanden zu kommen. Tatsächlich fanden Erbstößer et al. (2003) in einer Befragung von EFL-Anwendern heraus, dass in ca. 30% der Fälle die EFL den Probanden eine bessere, in ca. 10% der Fälle eine schlechtere Leistungsfähigkeit bescheinigt als die zuvor um eine Einschätzung gebetenen Ärzte. In der Regel wird daraufhin das ursprüngliche ärztliche Urteil korrigiert. Doch welche Zusatzinformationen bringt die EFL den Ärzten konkret?

→ Frage 1: Inwiefern entsprechen die sozialmedizinischen Beurteilungen der behandelnden Ärzte den Ergebnissen der EFL? In welchem Maße und an welchen Punkten unterscheiden sie sich? Wovon hängen mögliche Unterschiede ab?

4.2 Nutzen für den Patienten

Viele EFL-Patienten sind schon lange arbeitsunfähig. Es ist nicht auszuschließen, dass in der Zeit seit dem Beginn ihrer Erkrankung oder ihrer Arbeitsunfähigkeit ihre Leistungsfähigkeit aufgrund von Genesungsprozessen wieder gestiegen oder aufgrund von Dekonditionierungsprozessen gesunken ist.

Ohne Anlass, die eigenen berufsrelevanten Fähigkeiten auszuprobieren, fällt es den Patienten jedoch möglicherweise schwer, die eigene arbeitsbezogene Leistungsfähigkeit korrekt einzuschätzen (siehe auch SOMEKO, 2004) und daraus angemessene berufliche Perspektiven abzuleiten. Ein strukturiertes Ausprobieren der eigenen Fähigkeiten und eine detaillierte Rückmeldung könnten dafür hilfreich sein.

→ Frage 2: Trägt die EFL dazu bei, dass Patienten ein realistischeres Bild von ihrer eigenen funktionellen Leistungsfähigkeit bekommen? Wenn ja, hat das Auswirkungen auf ihre beruflichen Zielsetzungen?

4.3 Intangible Kosten: Beschwerdezunahme durch EFL?

Dem Nutzen eines Verfahrens stehen immer seine – monetären und intangiblen – Kosten gegenüber. Wie aus Gesprächen mit Ärzten, Therapeuten und Patienten deutlich geworden ist, schlägt an dieser Stelle häufig eine zeitweilige Zunahme der Beschwerden zu Buche, die bei der Verordnung von EFLs berücksichtigt werden muss. Dieses deckt sich mit der Aussage von Hart et al. (1993), dass vor allem bei Patienten, die seit mehr als 4 Wochen inaktiv sind, mit einer Schmerzzunahme gerechnet werden müsse. Den einzigen empirischen Hinweis auf ein Problem dieser Art haben wir bei Gross und Battié (2002) gefunden. Sie berichten, dass von 28 Probanden 3 am zweiten Testtag nicht erschienen sind, und sich 3 weitere aufgrund von Beschwerden nicht in der Lage fühlten, mit Gegenständen zu hantieren. Sie betonen, dass keiner von ihnen den ersten EFL-Tag als Ursache der Beschwerdezunahme angeführt hat. Unserer Ansicht nach deutet diese Quote jedoch ebenso wie die Rückmeldungen, die wir in Gesprächen mit Betroffenen bekommen haben, auf ein ernstzunehmendes Problem hin, das nach unserem Kenntnisstand noch nicht explizit untersucht worden ist.

→ Frage 3: Führt die EFL zu einer erheblichen Zunahme der Beschwerden?

5. Methode

5.1 Design

Die Studie war als Kombination von Quer- und Längsschnittuntersuchung mit einem sehr kurzen Intervall angelegt, wobei unterschiedliche Komponenten für die einzelnen Forschungsfragen herangezogen wurden. Der Übersichtlichkeit halber sollen an dieser Stelle das Gesamtdesign und die eingesetzten Instrumente dargestellt und jeweils zu Beginn der Kapitel zu den einzelnen Forschungsfragen die für die entsprechende Frage relevanten Teile noch einmal aufgegriffen werden.

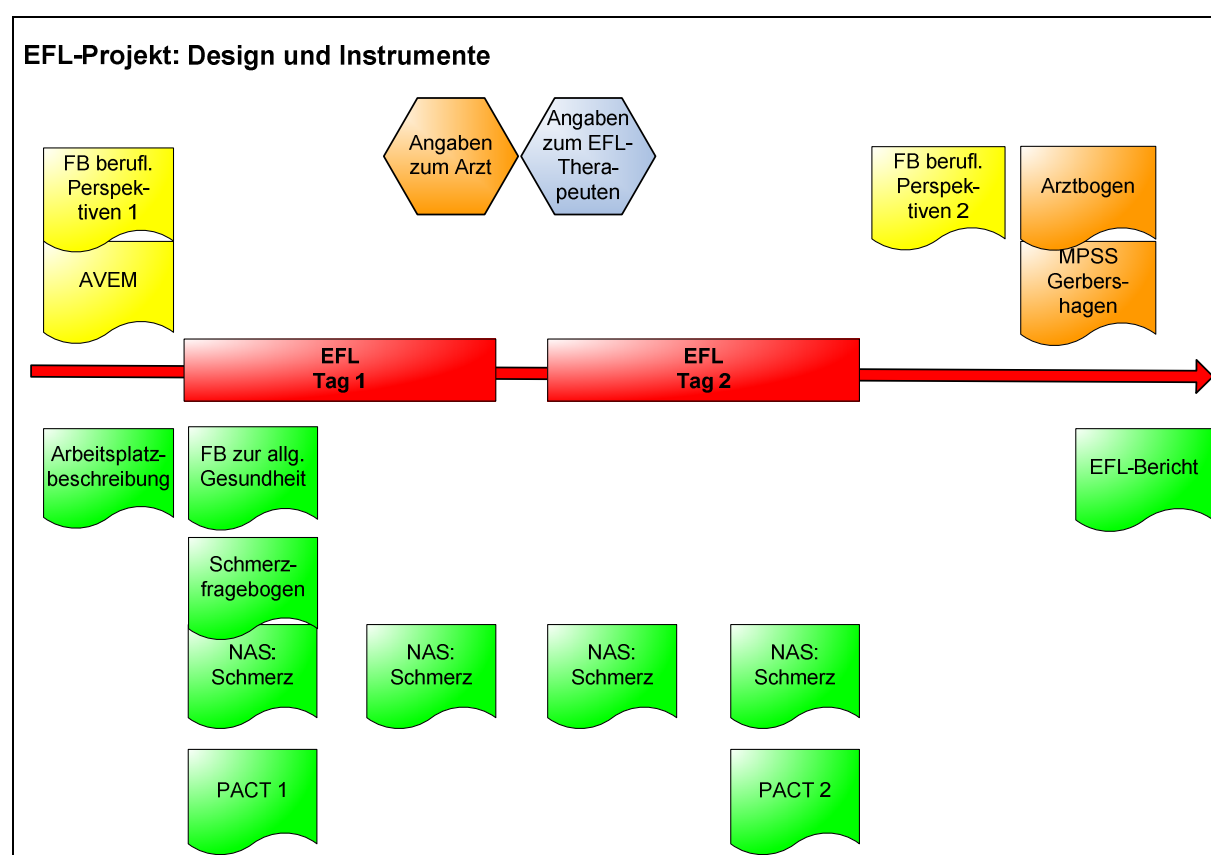


Abbildung 3: EFL-Projekt: Design und Instrumente

5.2 Instrumente

Wie in Abbildung 3 deutlich wurde, waren die eingesetzten Instrumente vielfältig. Sie lassen sich in vier Kategorien untergliedern und sollen im Folgenden vorgestellt und erläutert werden.

EFL-übliche Instrumente (in Abbildung 3 grün):

- In der „**Arbeitsplatzbeschreibung – Selbstauskunft durch Arbeitnehmer**“ geben die Patienten Organisationsform, Schwere und Bedingungen ihrer Arbeit an.
- Der „**Fragebogen zur allgemeinen Gesundheit**“ fragt das Vorhandensein von Risikofaktoren für körperliche Belastungen ab.
- Im „**Schmerzfragebogen**“ geben die Patienten neben der Beeinträchtigung durch Schmerzen in den letzten 7 Tagen auf numerischen Analogskalen (0=kein Schmerz; 10=unerträglicher Schmerz) am Anfang und Ende der Testtage ihre jeweilige aktuelle Schmerzstärke an.
- Die „**Selbstbeurteilung der körperlichen Fähigkeiten – PACT**“ (Schweizerische Arbeitsgemeinschaft für Rehabilitation, 1996) ist die deutsche Version des „Spinal Function Sort - Performance Assessment Capacity Testing (PACT)“ von Matheson & Matheson (1989) und erfasst die subjektiv eingeschätzten Fähigkeiten des Patienten, unterschiedlich schwere Alltagstätigkeiten zu verrichten.
- Im **EFL-Bericht** finden sich detaillierte Angaben zum Patienten, die Tabelle mit den Angaben zu einzelnen Testdaten und Belastbarkeitswerten sowie die zusammenfassende Einschätzung der Leistungsfähigkeit nach REFA.

Diese Instrumente werden von der EFL-Akademie vorgegeben oder empfohlen und im Rahmen der EFL standardmäßig eingesetzt.

Weitere Bögen für Patienten (in Abbildung 3 gelb):

- Der „**Fragebogen zu beruflichen Perspektiven I**“ umfasst neben den üblichen soziodemographischen Angaben Fragen zur gesundheitlichen und beruflichen Situation des Patienten und seiner wahrgenommenen Beeinträchtigung im Beruf durch seine Krankheit. Fragen zu beruflichen Perspektiven, Belastungen und Ressourcen sowie Hoffnungen und Befürchtungen im Hinblick auf den EFL-Test ergänzten das Spektrum. Dieser Bogen sollte am Tag vor der EFL ausgefüllt werden.
- Der Fragebogen „**Arbeitsbezogene Verhaltens- und Erlebensmuster (AVEM)**“ von Schaarschmidt und Fischer (2003) beschäftigt sich mit der Bewältigung von Arbeitsanforderungen. Dabei werden 11 Dimensionen von gesundheitsförderlichen oder -schädlichen Verhaltensweisen erhoben und daraus die Wahrscheinlichkeit der Zugehörigkeit zu den vier Bewältigungsmustern *Gesundheit, Schonung, Risiko im Sinne der Selbstüberforderung* und *Risiko im Sinne von Resignation und Depression* abgeleitet. Der AVEM wurde zusammen mit dem Fragebogen zu beruflichen Perspektiven I ausgeteilt.

- Im „**Fragebogen zu beruflichen Perspektiven II**“ wurden die Erfahrungen des Patienten mit dem EFL-Test erfragt und sowie die meisten Items aus dem Fragebogen zu beruflichen Perspektiven I zur indirekten Veränderungsmessung wiederholt. Dieser Bogen sollte am Tag nach der EFL ausgefüllt werden. Beide Fragebögen zu beruflichen Perspektiven wurden speziell für das Projekt von uns entwickelt.

Bögen für Ärzte (in Abbildung 3 orange):

- In den „**Angaben zum Arzt**“ wurden die Stations-, Ober- und EFL-Ärzte gebeten, uns Informationen über ihre Ausbildung und Erfahrung in ihrem Beruf, mit sozialmedizinischen Fragestellungen und mit der EFL zu geben, damit diese Variablen als mögliche Einflussfaktoren identifiziert werden können. Dieser Bogen ist eine Eigenentwicklung und wurde von jedem Arzt zu Beginn der Studie einmal ausgefüllt.
- Der „**Arztbogen**“ war analog zur EFL-Tabelle aufgebaut und erfasste die ärztliche Einschätzung der Einzelfähigkeiten eines Patienten sowie seiner allgemeinen Leistungsfähigkeit nach REFA. Die Einschätzung erfolgte durch eine Auswahl einer der durch die EFL-Tabelle vorgegeben Belastbarkeitskategorien *Nie/Keine*, *<10 kg*, *10-15 kg*, *15-40 kg*, *>40 kg* für die Maximalgewichte beim Hantieren von Lasten beziehungsweise *Nie (0%)*, *Gelegentlich (Bis zu 5%)*, *Zeitweise (Ca. 10%)*, *Überwiegend (51-90%)* und *Ständig (>90%)* für Haltung/Beweglichkeit und Fortbewegung. Handkraft, Gleichgewicht und Handkoordination wurden lediglich dichotom (*normal*, *<Norm*) beurteilt. Zusätzlich zu diesen Einschätzungen wurden die Ärzte nach ihrem Wissen um zusätzliche physische und psychische Belastungen des Patienten an seinem Arbeitsplatz gefragt. Auch dieser Bogen wurde von uns entworfen.
- Mit dem „**Mainz Pain Staging System (MPSS)**“ (Gerbershagen, keine Jahresangabe) wird das Chronifizierungsstadium eines Patienten ermittelt. Die Summe der Achsenstadien und das sich daraus ergebende Chronifizierungsstadium wurden im Arztbogen eingetragen.

Bogen für EFL-Therapeuten (in Abbildung 3 blau):

- Der Fragebogen „**Angaben zum EFL-Therapeuten**“ war analog zu „Angaben zum Arzt“ aufgebaut und fragte ebenfalls Ausbildung und Erfahrung im Beruf und mit EFLs ab. Dieser Bogen wurde von uns entwickelt und von jedem EFL-Therapeuten zu Beginn der Studie einmal ausgefüllt.

5.3 Ablauf

Der Ablauf war streng standardisiert um die normale Routine eines EFL-Tests organisiert. Durch ihn wurde sichergestellt, dass zu jeder Zeit auf der einen Seite alle nötigen Informationen bereitlagen, auf der anderen Seite aber keiner der Beteiligten über Fremd-Informationen verfügte, die sein eigenes Urteil hätten beeinflussen können.

Der erste Schritt bestand in der üblichen Verordnung eines EFL-Tests. Dieser folgte eine ausführliche Information des Patienten über die EFL und anschließend über die Studie, die in die Bitte einer freiwilligen Teilnahme mündete. Dieses Informationsgespräch wurde in Bad Pyrmont von Herrn Dr. Kasproski, in der Klinik Münsterland von den Projektmitarbeiterinnen übernommen. Stimmt der Patient der Teilnahme zu, wurden ihm eine schriftliche Darstellung des Projektes, Einverständniserklärungen für EFL und Studie sowie die vor der EFL auszufüllenden Fragebögen, also der Fragebogen zu beruflichen Perspektiven 1, der AVEM und die Arbeitsplatzbeschreibung, ausgehändigt. Diese Bögen sollte er zum ersten EFL-Termin wieder mitbringen. Die Einverständniserklärung für den EFL-Test und die Arbeitsplatzbeschreibung bekam der EFL-Therapeut, die beiden anderen Bögen wurden von den Projektmitarbeiterinnen abgeholt, bzw. in einem Briefkasten gesammelt, so dass der Therapeut keine Informationen daraus entnehmen konnte.

Zu Beginn der EFL füllte der Patient den Fragebogen zur allgemeinen Gesundheit, den Schmerzfragebogen inklusive der numerischen Analogskala zur Erfassung der aktuellen Schmerzstärke (NAS) und den PACT (PACT 1) aus. Die Angabe der aktuellen Schmerzstärke wurde zum Ende des 1. Testtages sowie zu Beginn und zum Ende des 2. Testtages wiederholt. Ebenso füllten die Probanden zum Ende des 2. Testtages den PACT noch ein weiteres Mal aus (PACT 2) und am Tag nach der EFL eine leicht veränderte Version des Fragebogens zu beruflichen Perspektiven (Fragebogen zu beruflichen Perspektiven 2). Dieser Bogen wurde ihnen erst am Ende des zweiten EFL-Tages ausgehändigt, so dass ein Abschreiben der Angaben aus dem ersten Bogen nicht möglich war. Er sollte am Tag nach der EFL ausgefüllt und in einem verschlossenen Umschlag in der Forschung (Klinik Münsterland) beziehungsweise dem Chefarztsekretariat (Bad Pyrmont) abgegeben werden.

Nach der EFL wurde der EFL-Bericht wie üblich verfasst, jedoch nicht den behandelnden Reha-Ärzten, sondern der Forschungsabteilung zur Verfügung gestellt. Dort bekamen ihn die Ärzte im Austausch gegen den ausgefüllten Arztbogen einschließlich der Ergebnisse des MPSS ausgehändigt und konnten ihn dann mit dem Patienten in der Abschlussuntersuchung besprechen.

5.4 Probanden

Eine Power-Analyse mit $\alpha=0,01$, $\beta=0,2$ und $d=0,5$ ergab eine optimale Stichprobengröße von 82 Versuchspersonen (Bortz & Döring, 1995). Diese sollte mit einer Vollerhebung aller stationären orthopädischen Reha-Patienten der Klinik Münsterland und des Reha-Zentrums Bad Pyrmont, die ab Dezember 2004 einen EFL-Test verordnet bekamen, erreicht werden. Leider verlief die Datenerhebung langsamer als geplant, so dass sie aus organisatorischen Gründen Ende Juni 2006 mit einer Stichprobengröße von $N=76$ abgebrochen werden musste. Erfreulicherweise hat vermutlich die intensive Vorinformation der Patienten im 1:1-Kontakt dazu beigetragen, dass fast alle einer Teilnahme bereitwillig zustimmten. Nur in wenigen Einzelfällen verweigerten Patienten die Studie, weil ihnen das Ausfüllen der Fragebögen zu viel Arbeit erschien oder sie nicht in ausreichendem Maße alphabetisiert waren.

6. Datenbasis

Insgesamt konnten 76 Patienten für die Studie rekrutiert werden. Drei von ihnen mussten die EFL aus medizinischen Gründen abbrechen, diese werden im Kapitel 9.1 intensiver besprochen. Da wir von diesen Probanden nicht alle Daten vorliegen haben, bezieht sich die Darstellung der Stichprobe an dieser Stelle auf die verbliebenen 73 Probanden.

6.1 Fehlende Daten

Bei einigen der in die Auswertung einbezogenen Probanden fehlen einzelne Fragebögen, so dass sich in den entsprechenden Analysen die Probandenanzahl reduziert:

- Fragebogen zu beruflichen Perspektiven I und II: 1 (beide Analphabetin)¹⁰
- AVEM: 8
 - davon 6 Patienten aus der Probephase, in der wir den AVEM noch nicht eingesetzt haben, inklusive Analphabetin
 - und 2, deren AVEM eigentlich vorliegen sollte (echte Missings)
- Fragebogen zur allgemeinen Gesundheit: 1 (Analphabetin)
- Arbeitsplatzbeschreibung 8
 - davon 5, bei denen das Nicht-Vorliegen plausibel ist
(1 Vp Analphabetin; 2 Vpn arbeitslos; 2 Vpn befristet EU-/BU-Rente)
 - und 3, die erwerbstätig sind und deren Bögen eigentlich vorliegen sollten (echte Missings)
- PACT 1: 1; PACT 2: 1 (beide Analphabetin)
- MMPS nach Gerbershagen: 14
 - davon 5 Patienten aus der Probephase, in der wir den MMPS noch nicht von Anfang an eingesetzt haben
 - und 9, deren MMPS eigentlich vorliegen sollten (echte Missings)
- Arztbogen: 3
- EFL-Bericht: 0

¹⁰ Bei dieser Patientin fehlen alle selbst auszufüllenden Bögen – sie ist Analphabetin und hat uns dieses leider so spät mitgeteilt, dass wir beim Ausfüllen nicht mehr helfen konnten.

6.2 Soziodemographische Daten

Die Projektplanung sah vor, in Bad Pyrmont und in Bad Rothenfelde etwa gleich viele Probanden zu rekrutieren – mit Probandenzahlen von 32 (43,8%) in Bad Pyrmont und 41 in Bad Rothenfelde (56,2%) konnte kein signifikanter Ortseffekt ermittelt werden.

In beiden Orten waren die Männer den Frauen gegenüber erheblich überrepräsentiert (jeweils $p=.000^{**}$). Insgesamt nahmen 69 Männer (94,5%) und 4 (5,5%) Frauen an der Studie teil. Die Frauen hatten einen Altersdurchschnitt von 42,75 Jahren (SD=7,14; Min=34; Max=51), die Männer waren im Mittel 39,16 Jahre alt (SD=8,55; Min=20; Max=57). Signifikante Geschlechts- oder Ortseffekte bezüglich des Alters gab es nicht.

Geschlecht x Ort		Klinik Münsterland	Bad Pyrmont	Gesamt
Männer	Anzahl	39	30	69
	Prozent	53,4%	41,1%	94,5%
Frauen	Anzahl	2	2	4
	Prozent	2,7%	2,7%	5,5%
Gesamt	Anzahl	41	32	73
	Prozent	56,2%	43,8%	100%

Tabelle 6: Geschlechterverteilung an den Erhebungsorten

Alter		Klinik Münsterland	Bad Pyrmont	Gesamt
Männer	MW	39,44	38,80	39,16
	SD	6,60	10,69	8,55
	Min. – Max.	22 – 53	20 – 57	20 – 57
Frauen	MW	46,00	39,50	42,75
	SD	7,07	7,78	7,14
	Min. – Max.	41 – 51	34 – 45	34 – 51
Gesamt	MW	39,76	38,84	39,36
	SD	6,61	10,43	8,48
	Min. – Max.	22 – 53	20 – 57	20 – 57

Tabelle 7: Altersdurchschnitte in Abhängigkeit von Geschlecht und Ort

72 Probanden (98,6%) hatten die deutsche Staatsbürgerschaft, mit ihnen war eine Verständigung in deutscher Sprache für den EFL-Test in ausreichendem Maß möglich. Bei einem ausländischen Probanden wurde die Kommunikation als nicht ausreichend angegeben – die Fragebögen wurden jedoch korrekt ausgefüllt.

Auf die Frage nach dem Familienstand gaben 39 Probanden (53,4%) an, verheiratet zu sein, 24 (32,9%) waren ledig. Die Scheidungsquote war mit 75% (n=3) bei den Frauen erheblich höher als bei den Männern mit 5,8% (n=3; $\chi^2=23,691$; $p=.000^{**}$).

43 Probanden (58,9%) gaben an in einer Lebensgemeinschaft zu leben, 24 (32,9%) gaben keine weitere Person in ihrem Haushalt an. Im Mittel lebten 1,32 Personen unter 18 Jahren mit im Haushalt, und 0,86 Kinder wurden als unterstützungspflichtig angegeben. Für die meisten Probanden ist die Familie eine wichtige Ressource: 53 (72,6%) gaben an, sich auf Unterstützung durch die Familie sehr stark verlassen zu können, nur 4 (5,5%) fühlten sich gar nicht unterstützt.

6.3 Ausbildung und Beruf

Die Schulausbildung unsere Probanden ist eher als basal zu bezeichnen: Über 2/3 (n=51) hatten einen Hauptschulabschluss und etwa 1/5 (n=14) hatten die Realschule beendet. Knapp 10% (n=7) haben ohne Abschluss die Schule verlassen – Frauen (N=2; 50% der Frauen) signifikant häufiger als Männer (N=5; 7,2% der Männer) (Exakter Test nach Fisher; $\chi^2=7,971$; $p=.044^*$). Kein Proband hatte das Abitur gemacht.

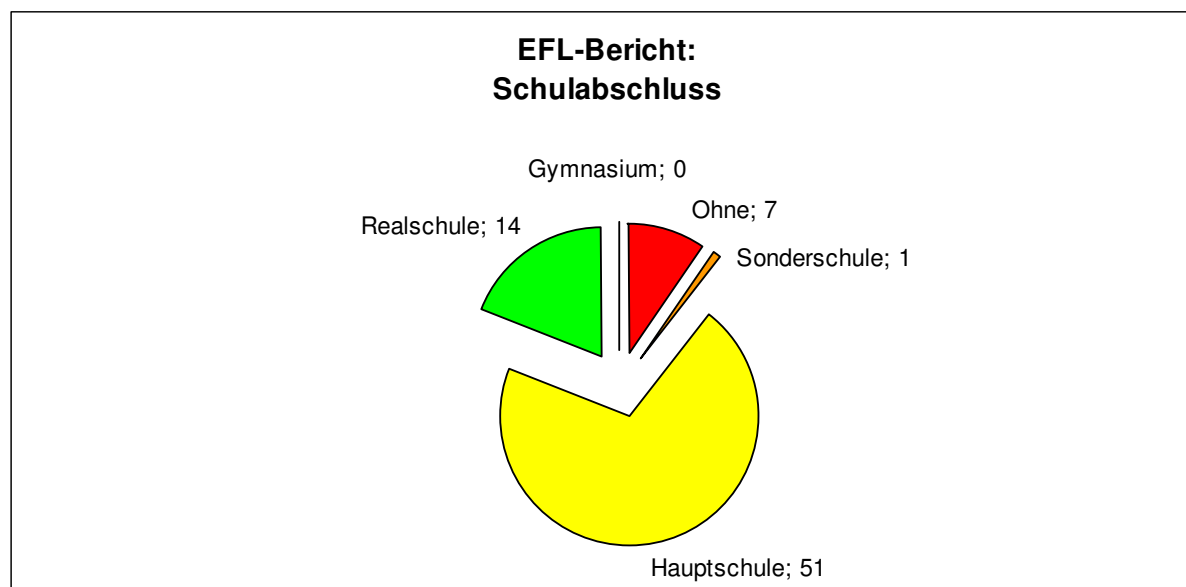


Abbildung 4: Schulabschluss

Die Angaben zu den Berufsausbildungen der Probanden korrespondierten gut mit den Angaben zur Schulbildung: Mit 56 Probanden (76,7%) hat die überwiegende Mehrheit eine Lehre absolviert, 3 (4,1%) erlangten eine höhere Qualifikation. Ein Proband (1,4%) war angelernt und 13 (17,8%) haben keinerlei Ausbildung durchlaufen. Diese Daten sind den EFL-Berichten entnommen.

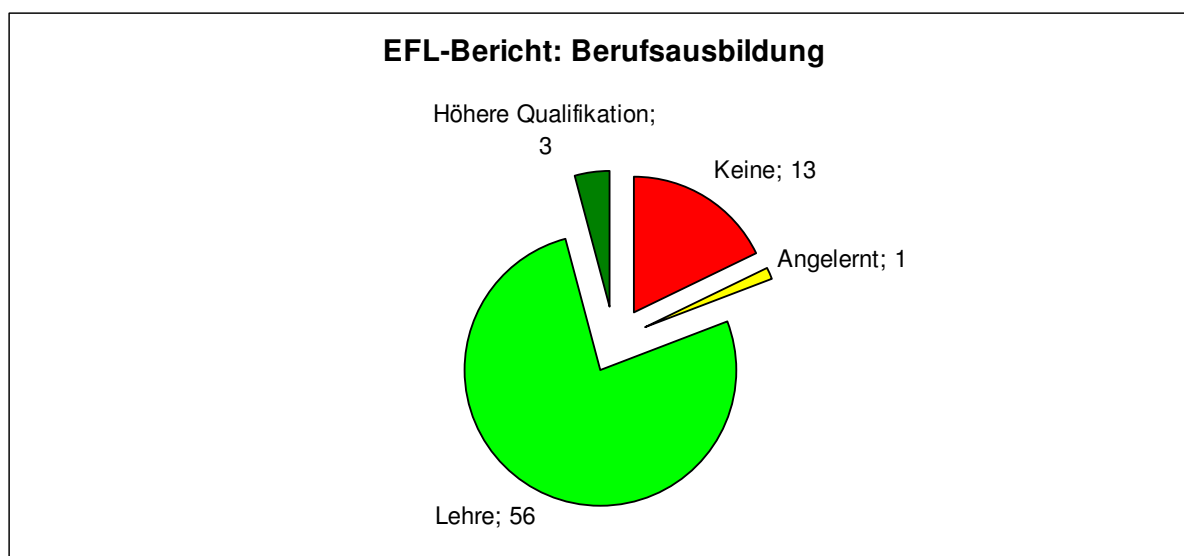


Abbildung 5: Berufsausbildung

Erfreulicherweise waren mit 57 (78,1%) über 3/4 unserer Probanden erwerbstätig, mit 13,7% (N=10) lag die Erwerbslosenquote leicht über dem Bundesdurchschnitt. Zwei Probanden (2,7%) waren in Ausbildung oder Umschulung, weitere 2 (2,7%) befristet in Erwerbs- oder Berufsunfähigkeitsrente (Abbildung 6). Laut EFL-Bericht hatten 44 Probanden (60,3%) einen gültigen Arbeitsvertrag, 7 (9,6%) ebenso, mussten aber mit einer Kündigung rechnen. 20 Probanden (27,4%) waren nicht vertraglich geschützt und von zwei Probanden (2,7%) liegt keine Angabe dazu vor.

Erwartungsgemäß hatten die meisten Probanden in ihrer letzten Tätigkeit einen Facharbeiter- oder Fachangestelltenstatus (N=29; 39,7%) oder waren angelernt (N=23; 31,5%).

Laut den EFL-Berichten haben 70 der Probanden (95,9%) an ihrem letzten Arbeitsplatz vollzeitig gearbeitet, 2 (2,7%) waren teilzeitbeschäftigt und von einem Probanden (1,4%) fehlt die Angabe. Einen Schonarbeitsplatz hatte niemand inne.

Die Männer in unserer Studie arbeiteten durchschnittlich 8,53 Stunden pro Tag (SD=1,29; Min=4,50; Max=14,00) beziehungsweise 41,94 Stunden pro Woche (SD=12,20; Min=5,00; Max=70,00). Damit waren sie durch ihren Beruf zeitlich signifikant stärker eingebunden als die Frauen, die durchschnittliche Arbeitsdauern von 6,50 Stunden pro Tag (SD=0,71; Min=6,00; Max=7,00) und 26,00 Stunden pro Woche (SD=18,33; Min=6,00; Max=42,00) angaben ($T_{\text{Tag}}=2,189$; $p=.033^*$; $T_{\text{Woche}}=2,160$; $p=.035^*$).

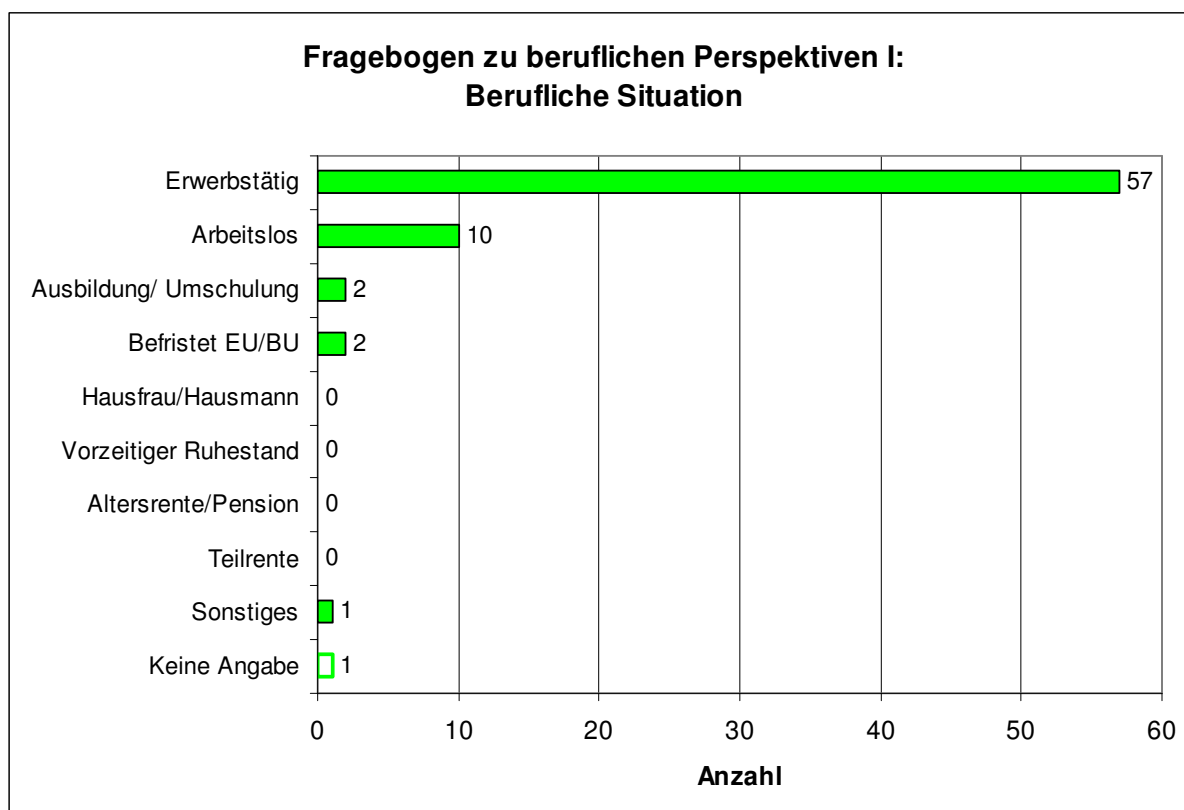


Abbildung 6: Berufliche Situation

Die Dauer der Betriebszugehörigkeit variierte sehr stark, lag aber eher im mittleren bis höheren Bereich: 33 Probanden (45,2%) gehörten ihrem Betrieb schon über 10 Jahre an, ebenfalls 33 Probanden (45,2%) berichten eine Firmenzugehörigkeit von mehr als einem Jahr. Weniger als ein Jahr waren nur 6 Patienten (8,2%) bei ihrem letzten Arbeitgeber beschäftigt.

Trotzdem zeichnet sich eine besondere arbeitsbezogene Problemlage der EFL-Patienten ab, denn zur Zeit des EFL-Tests waren 62 Probanden (84,9%) - und damit erheblich mehr als in der Gesamtheit der Patienten der Klinik Münsterland¹¹ - arbeitsunfähig. Diese Quote war in beiden Erhebungsorten vergleichbar, die Dauer der Arbeitsunfähigkeit bei den betroffenen Probanden war jedoch in der Klinik Münsterland signifikant höher als in Bad Pyrmont ($MW_{KML}=313,5$ Tage; $SD_{KML}=311,4$; $MW_{BP}=161,4$ Tage; $SD_{BP}=114,1$; $T=2,566$; $p=.014^*$). Ebenso waren Männer im Schnitt länger arbeitsunfähig als Frauen ($MW_M=251,7$ Tage; $SD_M=258,9$; $MW_W=136,5$ Tage; $SD_W=68,1$), wobei dieser Unterschied aufgrund der geringen Stichprobengröße der Frauen die Signifikanz verfehlte ($T=0,881$; $p=.382$).

¹¹ Insgesamt sind etwa 70% der Patienten der Klinik Münsterland arbeitsunfähig, aus dem Reha-Zentrum Bad Pyrmont liegen uns leider keine Zahlen vor.

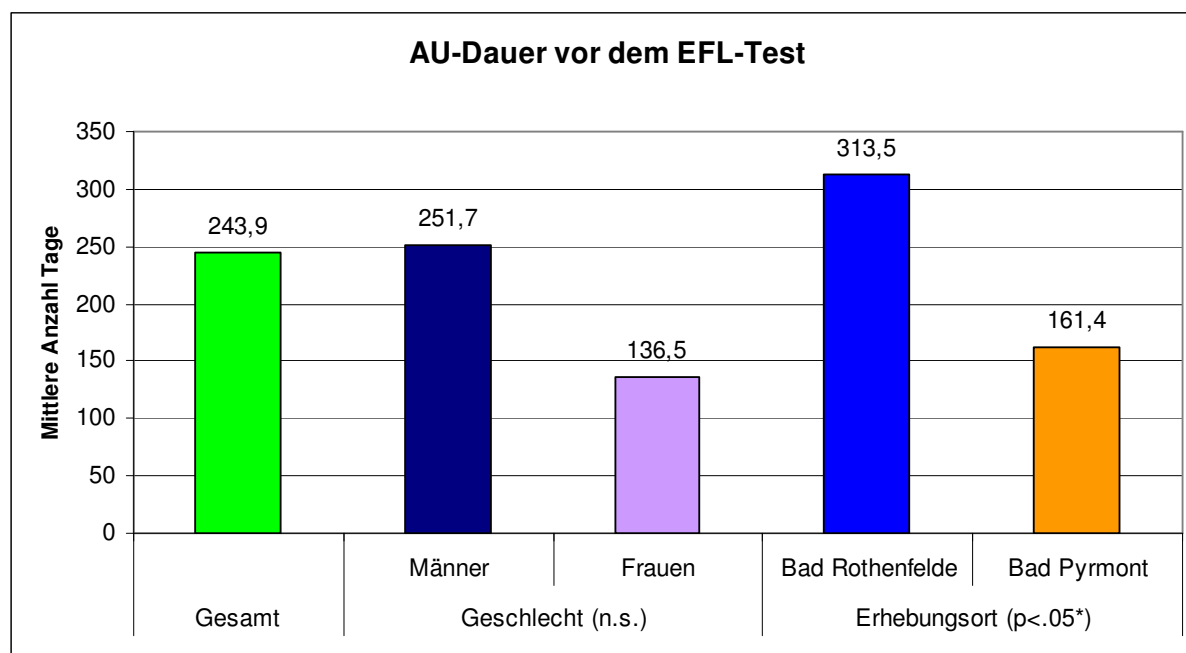


Abbildung 7: Dauer der Arbeitsunfähigkeit der zum EFL-Test arbeitsunfähigen Probanden

Aufschlussreich ist die Kategorisierung der AU-Dauer nach medizinisch und verwaltungstechnisch relevanten Zeiträumen. Dabei gilt das Kriterium eines halben Jahres als besonders interessant, da spätestens ab einer solchen Beschwerdedauer von Chronifizierung gesprochen wird und sich ab 6 Monaten Arbeitsunfähigkeit auch die berufliche Prognose der Patienten erheblich verschlechtert. Es zeigt sich, dass hiervon in der Klinik Münsterland erheblich mehr Patienten betroffen waren als in Bad Pymont, auch dieser Unterschied ist statistisch signifikant ($\chi^2=10,192$; $p=.037^*$) (Abbildung 8).

Die Angaben zu einem aktuellen Rentenbezug im EFL-Bericht deckten sich mit denen im Fragebogen zu beruflichen Perspektiven: 2 Probanden (2,7%) bekamen zur Zeit des Fragebogenausfüllens eine Rente, in beiden Fällen handelte es sich um eine Erwerbs- oder Berufsunfähigkeitsrente.

Einer der beiden Patienten sowie 6 weitere (insgesamt also insgesamt 7; 11%) hatten einen Rentenanspruch gestellt, über den noch nicht entschieden worden war.

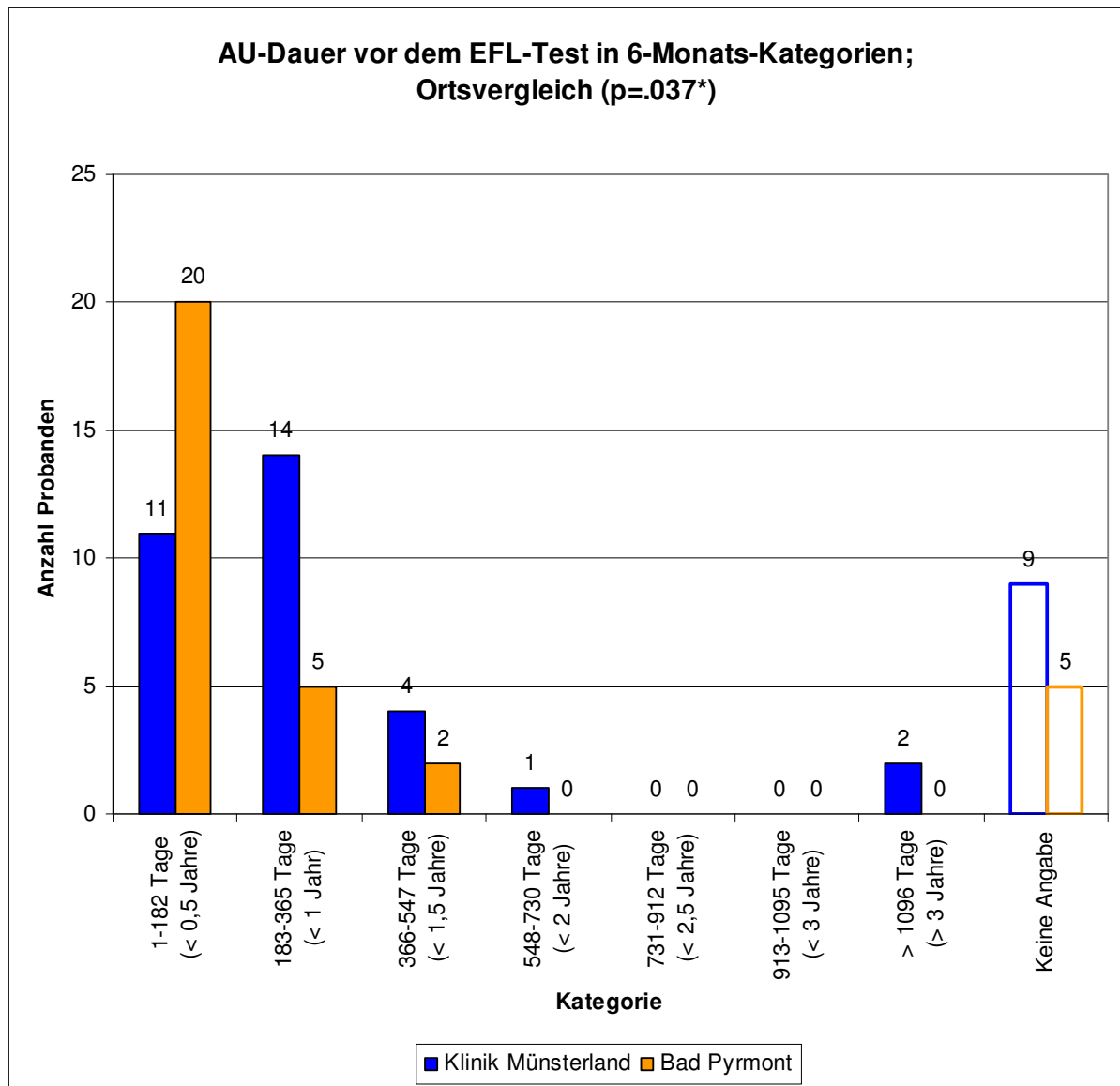


Abbildung 8: Dauer der Arbeitsunfähigkeit vor dem EFL-Test in 6-Monatskategorien im Ortsvergleich

6.4 Gesundheit

Im Fragebogen zur allgemeinen Gesundheit wurden die Risikofaktoren der Patienten erfragt. Mit 43 Nennungen (58,9%) führte Trainingsmangel die „Hitliste“ an, gefolgt von Übergewicht (N=36; 49,3%). Nimmt man nicht die Einschätzung der Patienten, sondern den BMI als Kriterium, müssen nach der auf der WHO-Website (2006) angegebenen Klassifikation sogar 40 Probanden (54,8%) als übergewichtig und 16 weitere (21,9%) als adipös bezeichnet werden.

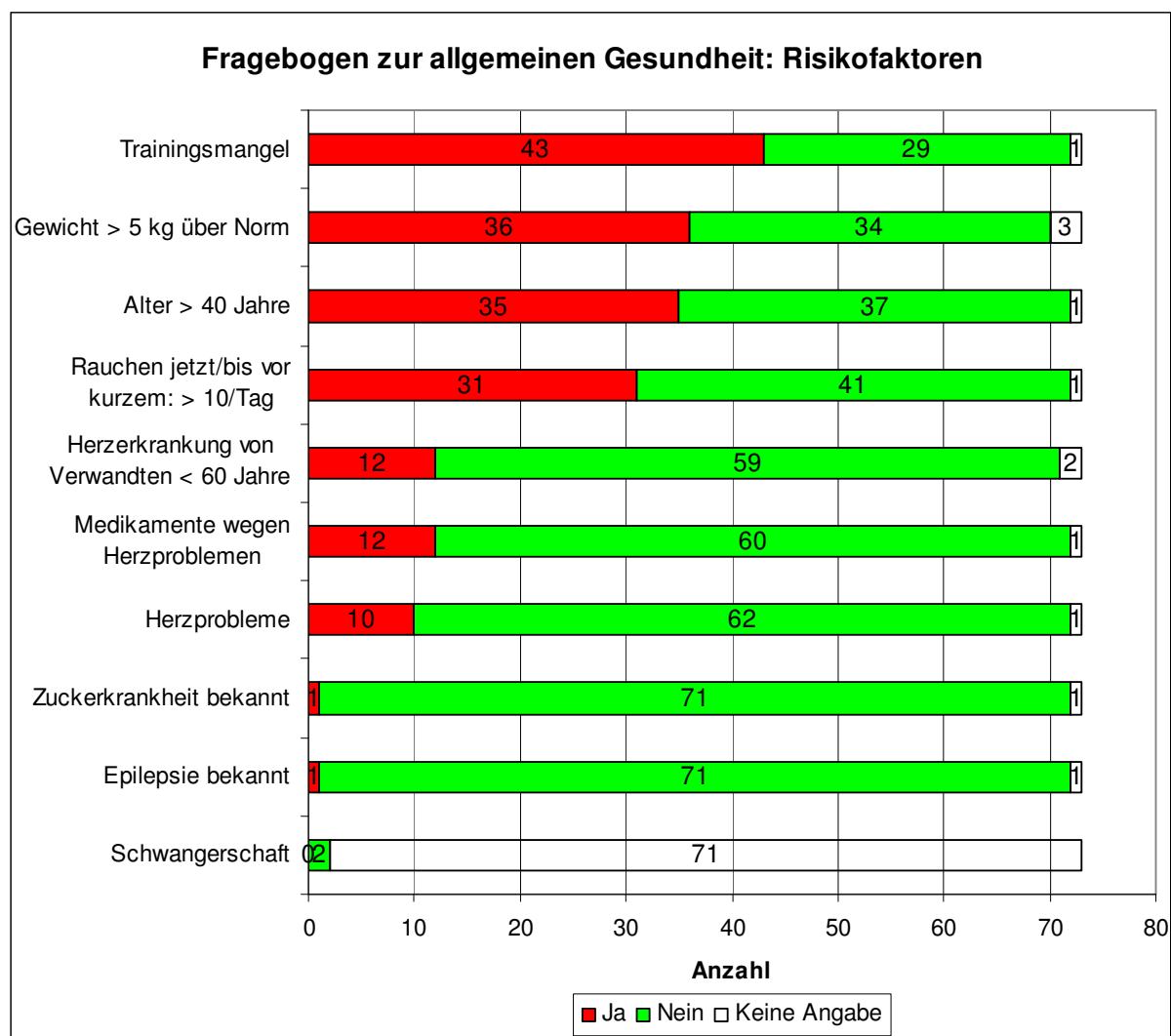


Abbildung 9: Risikofaktoren

Im EFL-Bericht wird für jeden Patienten das arbeitsbezogene Problem per Freitext eingetragen. Für die Auswertung haben wir diese Freitext-Angaben kategorisiert und den Gruppen „Obere Extremität“, „Rücken“; „Untere Extremität“, „Allgemeine Kraft-/Ausdauerschwäche“, „Psychosoziale Probleme“ und „Kein arbeitsbezogenes Problem“ zugeordnet. Dabei waren Mehrfachangaben erlaubt. Bei Rückenbeschwerden mit Ausstrahlungen in die Extremitäten haben wir nur die Rückenbeschwerden gewertet, da die Ausstrahlungen sekundär dem Rückenproblem zugeordnet werden können.

Insgesamt sind erwartungsgemäß Beschwerden im Bereich des Rückens das häufigste arbeitsbezogene Problem. Sie werden in der Klinik Münsterland mit 32 Nennungen (78,0% der dortigen Patienten) signifikant häufiger angeführt als in Bad Pyrmont (N=15; 46,9% der dortigen Patienten; $\text{Chi}^2=7,800$; $p=.011^*$), während in Bad Pyrmont öfter allgemeine Kraft- und Ausdauerschwächen diagnostiziert werden (Bad Pyrmont: N=8; 25%; Klinik Münsterland: N=1; 2,4%; $\text{Chi}^2=8,570$; $p=.008^{**}$). Zu

bedenken ist, dass diese Ergebnisse nicht nur den Gesundheitszustand der Patienten abbilden, sondern sicherlich auch diagnostische Traditionen und unterschiedliche Grade von Bewusstseinsnähe dieser Diagnosen mit einfließen.

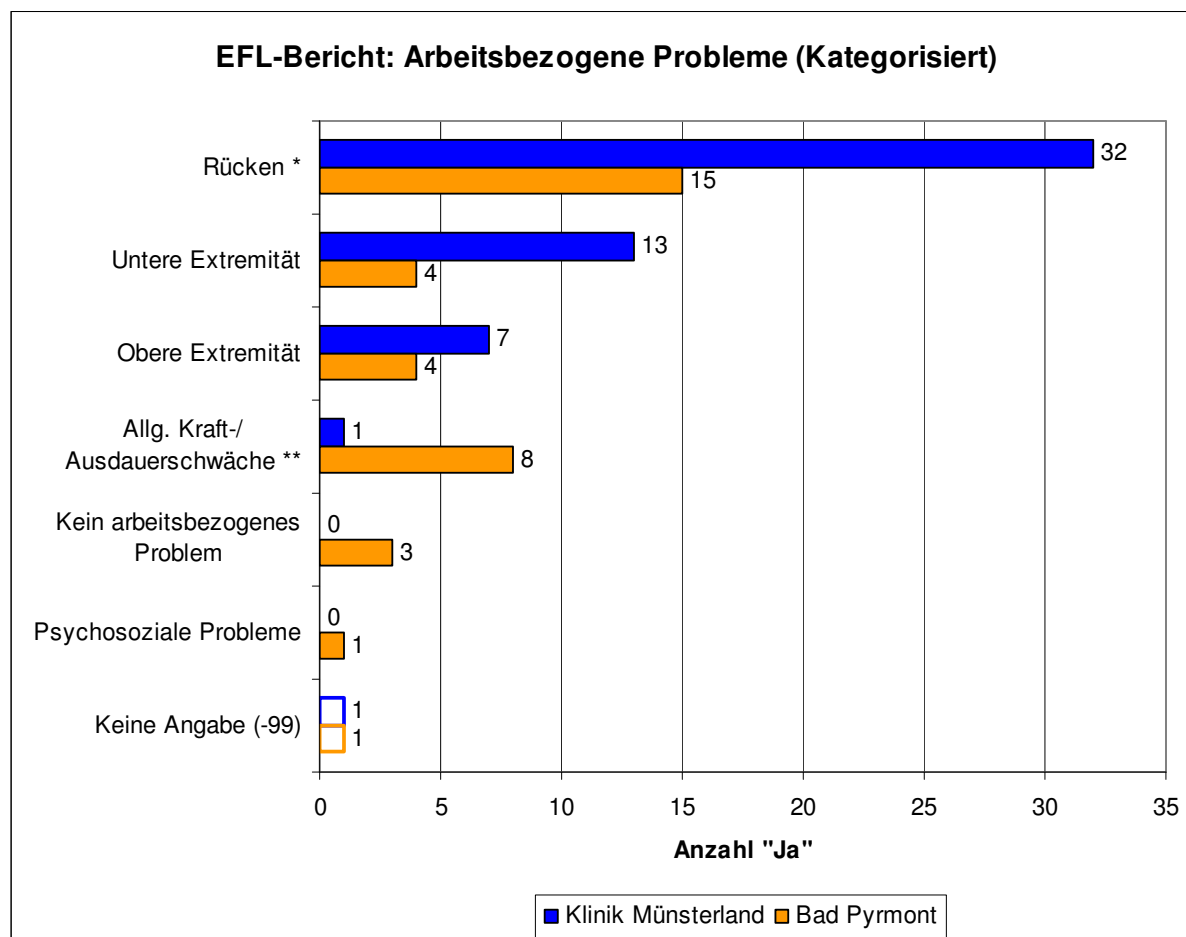


Abbildung 10: Arbeitsbezogene Probleme im Ortsvergleich

Die Patienten aus beiden Kliniken litten in den 7 Tagen vor dem EFL-Test unter erheblichen Schmerzen; der schlimmste Schmerz war bei den Probanden aus Bad Rothenfelde mit einem mittleren Wert von 6,67 (SD=2,03) auf einer Skala von 0 (kein Schmerz) bis 10 (unerträglicher Schmerz) noch signifikant stärker als bei den Probanden aus Bad Pyrmont mit 4,86 (SD=2,09) (T=3,745; p=.000**).

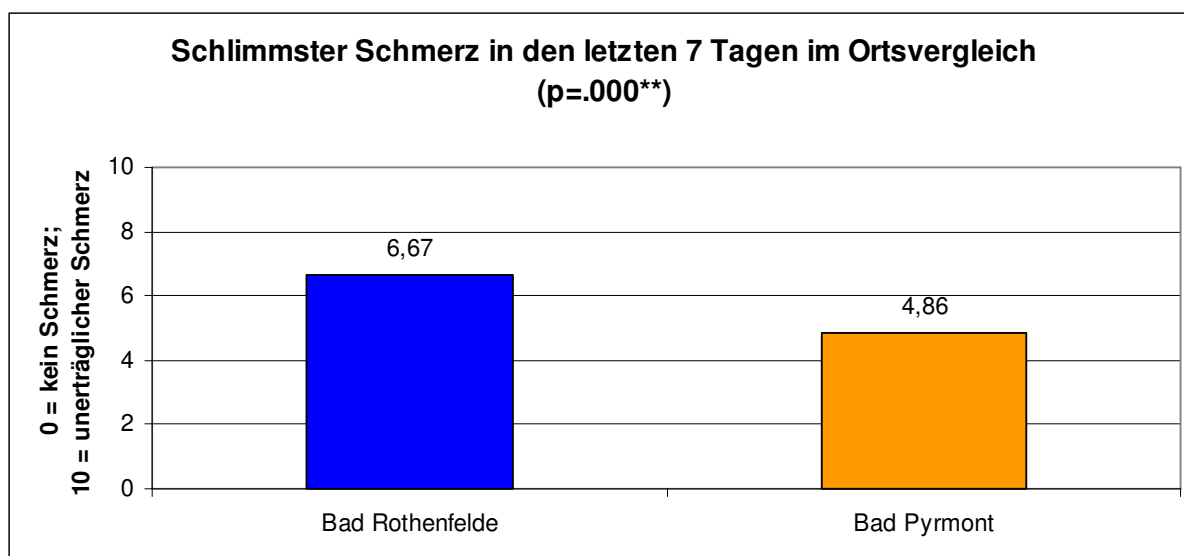


Abbildung 11: Schlimmster Schmerz in den letzten 7 Tagen im Ortsvergleich

Die Anzahl Tage mit Schmerz, mit sehr starkem Schmerz, mit Schmerzmedikamenten und mit schmerzbedingten Schlafstörungen unterschieden sich bei den beiden Probandengruppen nicht. Die deutlichen Unterschiede zwischen den Angaben der Tage mit Schmerz und der Tagen mit sehr starkem Schmerz deuten darauf hin, dass die Patienten durchaus in der Lage sind, ihre Beschwerden differenziert wahrzunehmen und anzugeben – eine Fähigkeit, die manchen chronifizierten Patienten verloren gegangen ist und dennoch sowohl die Selbsteinschätzung als auch die Prognose mitbestimmen kann.

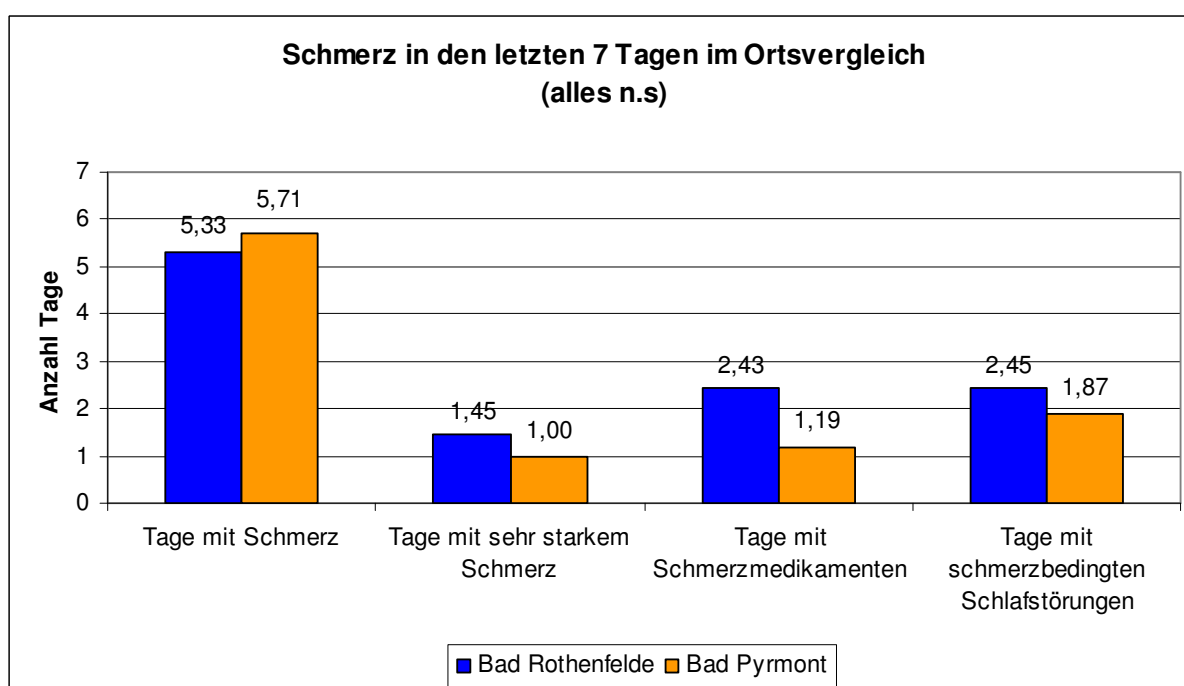


Abbildung 12: Schmerz in den letzten 7 Tagen im Ortsvergleich

Die momentane Einschränkung der Leistungsfähigkeit hatte bei 54 Probanden (73,9%) eine Erkrankung als Ursache; von 43 (58,9%) dieser Patienten ist bekannt, in welchem Jahr die Krankheit begonnen hat. Vergleicht man dieses mit dem Jahr, in dem die EFL durchgeführt wurde, ergibt sich im Mittel ein Abstand von 2,14 (SD=3,41) Jahren. In der Regel waren 0 bis 12 Jahre vergangen, bei einem Probanden brach die Krankheit bereits 20 Jahre vor der EFL aus.

Bei 10 Probanden (13,7%) wurde die Leistungseinschränkung auf einen Unfall zurückgeführt. Von 6 (8,2%) dieser Patienten ist das konkrete Unfalldatum bekannt, es lag im Vergleich zur EFL im Mittel 419,5 Tage (SD=361,58) zurück. Ein Proband hatte seinen Unfall 1129 Tage vor der EFL, die anderen weisen Werte zwischen 156 und 455 Tage auf.

Im Ortsvergleich zeigten sich erhebliche Unterschiede, ohne dass diese jedoch statistische Signifikanz erreicht hätten:

		Bad Rothenfelde	Bad Pyrmont	T	p
Mittlere Anzahl Jahre zwischen Krankheits- beginn und EFL	Incl. Ausreißer (20 Jahre)	1,40 Jahre	2,54 Jahre	-1,317	.196
	Excl. Ausreißer (20 Jahre)	1,40 Jahre	1,89 Jahre	-0,841	.405
Mittlere Anzahl Tage zwischen Unfall und EFL	Incl. Ausreißer (1129 Tage)	532,0 Tage	194,5 Tage	1,099	.333
	Excl. Ausreißer (1129 Tage)	333,5 Tage	194,5 Tage	1,582	.212

Tabelle 8: Zeiträume zwischen Krankheitsbeginn oder Unfall und EFL im Ortsvergleich

Vor dem jeweiligen Ereignis waren 45 Probanden (61,6%) voll und 8 (11,0%) eingeschränkt leistungsfähig, bei 20 (27,4%) fehlt die Angabe.

In Anbetracht der Zeiträume seit dem Erkrankungsbeginn oder dem Unfall scheinen die Probanden relativ stark chronifiziert. Da jedoch inzwischen nicht nur die Zeit als Indikator genutzt, sondern die Chronifizierung als multifaktorielles Geschehen aufgefasst wird, wurde der MPSS nach Gerbershagen eingesetzt. Danach waren 27 (37,0%) unserer Probanden dem Stadium I, 30 (41,0%) dem Stadium II und nur 2 (2,7%) dem Stadium III zuzuordnen. Von 14 Probanden (19,2%) fehlt diese Angabe.

6.5 Leistungsbereitschaft

Die Leistungsbereitschaft der Probanden wurde von den EFL-Therapeuten in 56 Fällen (76,7%) als zuverlässig, in 15 Fällen (20,5%) als fraglich und nur in 2 Fällen (2,7%) als nicht zuverlässig beurteilt; ein signifikanter Unterschied zwischen den Erhebungsorten bestand nicht ($\chi^2=4,357$; $df=2$; $p=.113$).

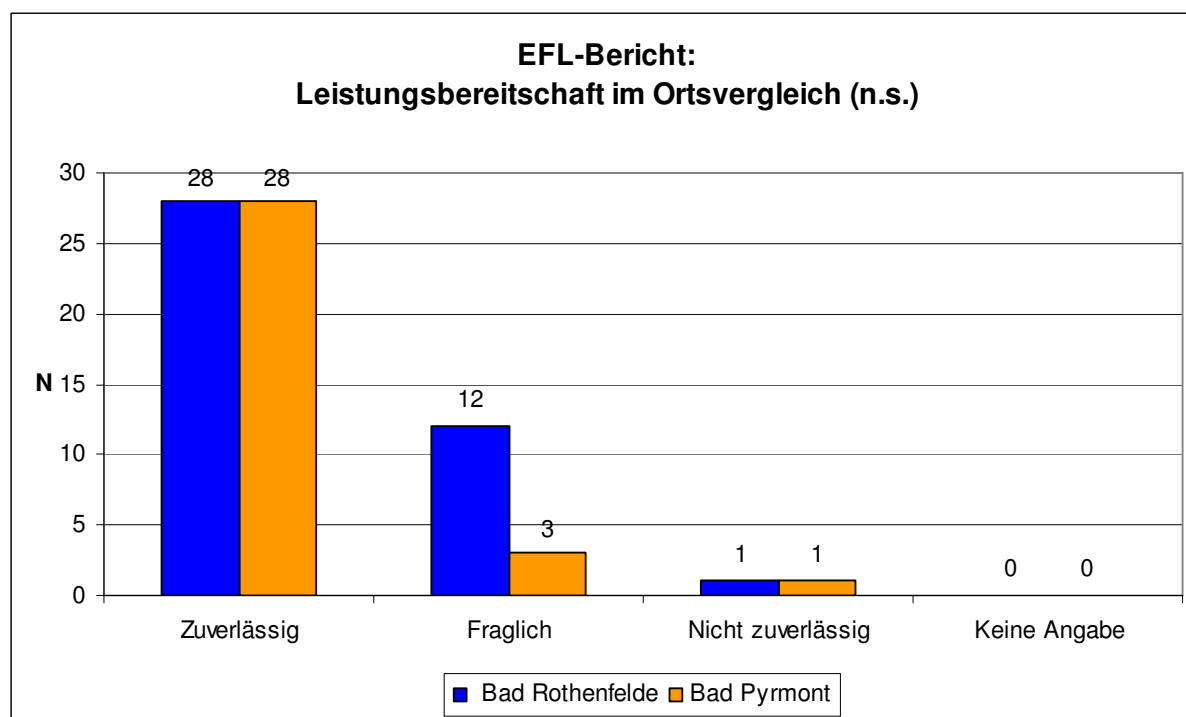


Abbildung 13: Zuverlässigkeit der Leistungsbereitschaft nach Einschätzung der EFL-Therapeuten

Diese Ergebnisse gehen konform mit den Befunden, dass bei 63 Probanden (86,3%) die Konsistenz der Subtestergebnisse als gut, bei 9 Probanden (12,3%) als mäßig und nur bei einem Probanden (1,4%) als schlecht beurteilt wurde.

Auch die Werte der Hebetests wurden im Großen und Ganzen am 2. Testtag bei der überwiegenden Mehrzahl der Probanden reproduziert, nur jeweils 7 (9,6%) zeigten eine vermehrte oder verminderte Belastbarkeit in diesem Bereich.

Neben den EFL-Ergebnissen liegen uns Daten des AVEM vor, mit dessen Hilfe Selbsteinschätzungen der Probanden hinsichtlich ihrer arbeitsbezogenen Verhaltens- und Erlebensmuster erhoben werden. Die für uns wichtigste Einzelskala ist die Verausgabungsbereitschaft. In dieser wiesen unsere Probanden einen mittleren Prozentrang von 52,2 im Vergleich zu einer heterogenen Normierungsstichprobe und einen mittleren Prozentrang von 37,4 im Vergleich zu einer Gruppe psychosomatischer Reha-Patienten auf. Fasst man das Konzept der Risikomuster etwas größer, so ist es plausibel, jeden Patienten dem Muster zuzuordnen, dem er mit

der höchsten Wahrscheinlichkeit angehört, ohne die konkreten Ausprägungen aller Wahrscheinlichkeiten zu berücksichtigen. Bei einer solchen Vorgehensweise ergibt sich für die Gesamtgruppe folgendes Bild:

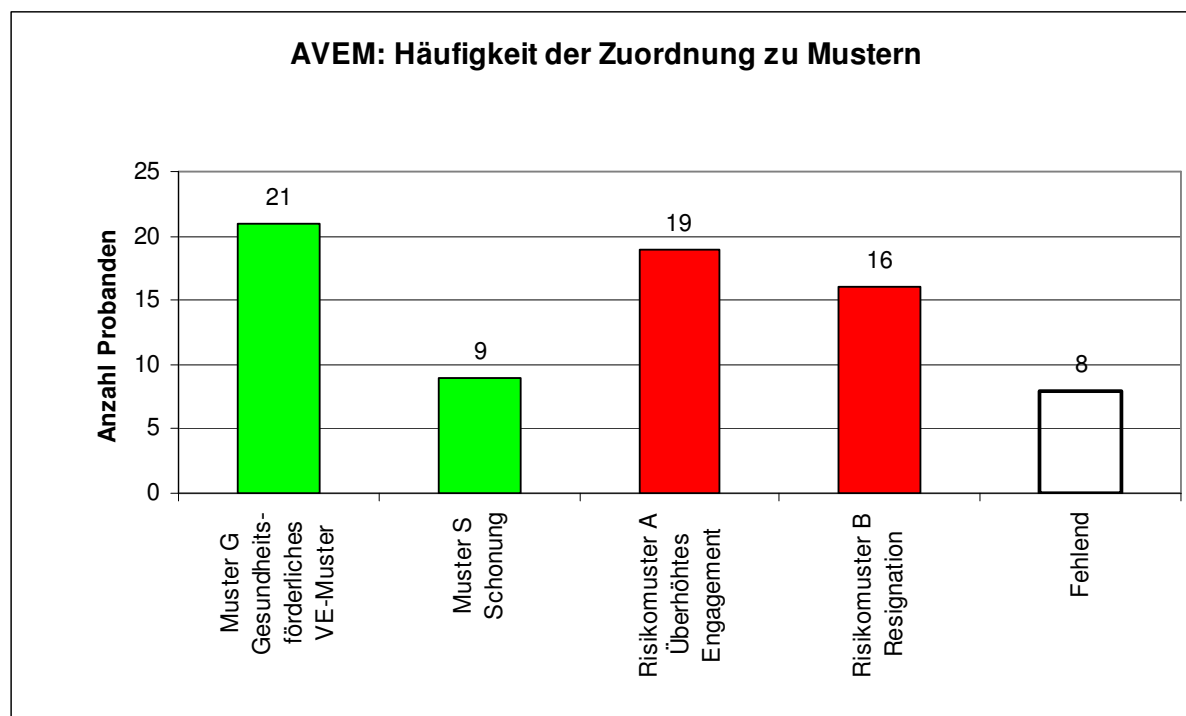


Abbildung 14: AVEM: Zuordnung der Probanden zu Mustern

Über ein Viertel der Probanden ($n=21$; 28,8%) weist ein Muster gesundheitsförderlichen Verhaltens und Erlebens auf, bei dem mit einer hohen, aber nicht übertriebenen Arbeitsmotivation und einer ebensolchen Einsatzbereitschaft gerechnet werden kann. 9 Probanden (12,3%) neigen eher zur Schonung. Auch dieses Muster wird von Schaarschmidt und Fischer positiv bewertet, denn solche Probanden sind durchaus zufrieden und haben ein eher positives Lebensgefühl. „Allerdings dürfte die Quelle dafür eher außerhalb der Arbeit zu finden sein“ (Schaarschmidt & Fischer, 2003, S. 11). Die beiden im AVEM definierten Risikomuster gehen beide mit einem Überforderungserleben bei eingeschränkter Erholungs- und Distanzierungsfähigkeit sowie vermehrten psychischen und körperlichen Beschwerden einher. Unterschieden werden das Risikomuster A „Überhöhtes Engagement“, das viel Ähnlichkeit mit dem Typ A - Verhalten bei gleichzeitiger Gratifikationskrise hat, und das Risikomuster B „Resignation“, das ungefähr dem Burnout entspricht. Beide sind mit 1/4 ($n=19$; 26,0%) beziehungsweise 1/5 ($n=16$; 21,9%) der Probanden ebenfalls maßgeblich vertreten und scheinen die beiden in der Schmerztherapie oft unterschiedenen Gruppen der „fröhlichen (?; die Autoren) Durchhalter“ und der „Katastrophisierer“ (Kürten, 2004) abzubilden. Ein signifikanter Ortseffekt hinsichtlich der Muster bestand nicht ($\chi^2=1,248$; $df=3$; $p=.741$).

7. Frage 1: Inwiefern entsprechen die sozialmedizinischen Beurteilungen der behandelnden Ärzte den Ergebnissen der EFL? In welchem Maße und an welchen Punkten unterscheiden sie sich? Wovon hängen mögliche Unterschiede ab?

Der Hauptzweck von FCE-Systemen, auch der EFL nach Isernhagen, wird darin gesehen, Beurteiler der funktionellen und in der Regel arbeitsbezogenen Leistungsfähigkeit von Menschen in ihrer Aufgabe zu unterstützen, indem sie deren Einschätzung auf eine reliablere und validere Basis stellen. In der Literatur werden vor allem Befunde zur Inhaltsvalidität und zur Kriteriumsvalidität im Hinblick auf eine berufliche Wiedereingliederung von Patienten präsentiert. Aussagen zum konkreten Nutzen für die Beurteiler hingegen haben wir nur in der Publikation von Erbstößer et al. (2003) gefunden. Die Autoren berichten aus einer Befragung von EFL-Anwendern, dass in etwa 60% der Fälle die Einschätzung der Leistungsfähigkeit der Probanden durch Ärzte mit den EFL-Ergebnissen übereinstimmen. In ca. 30% der Fälle kommt die die EFL zu einem besseren, in ca. 10% der Fälle zu einem schlechteren Testergebnis als die vorher getroffenen Urteile der Ärzte. In der Regel werden in diesen Fällen die ursprünglichen Einschätzungen korrigiert. Leider beschreiben die Autoren nicht genauer, worauf sich die Einschätzung der Leistungsfähigkeit in ihrer Befragung bezieht, vermutlich ist die allgemeine Leistungsfähigkeit nach REFA gemeint.

Es stellt sich also die Frage, ob sich diese Ergebnisse replizieren und differenzieren lassen. Sind die Diskrepanzen in den einzelnen Subtests unterschiedlich groß? Auf welche Subtests sollte man keinesfalls verzichten? Wovon hängen Unterschiede in den Urteilen ab? Sind sie rein als „Messfehler“ zu werten oder sind sie erklärlich? Kann man sie vielleicht sogar nutzen?

7.1 Design und Instrumente

Dieser erste Fragenkomplex wurde im Kern über einen Vergleich der EFL-Ergebnisse mit verblindeten Arzturteilen über die gleichen Patienten beantwortet. Weitere Daten wurden für die Prädiktorenanalyse einbezogen. Die Kerninstrumente sind in der Abbildung 15 satt, die Instrumente für die Prädiktorenanalyse blass und die für diese Frage nicht verwendeten Instrumente grau dargestellt.

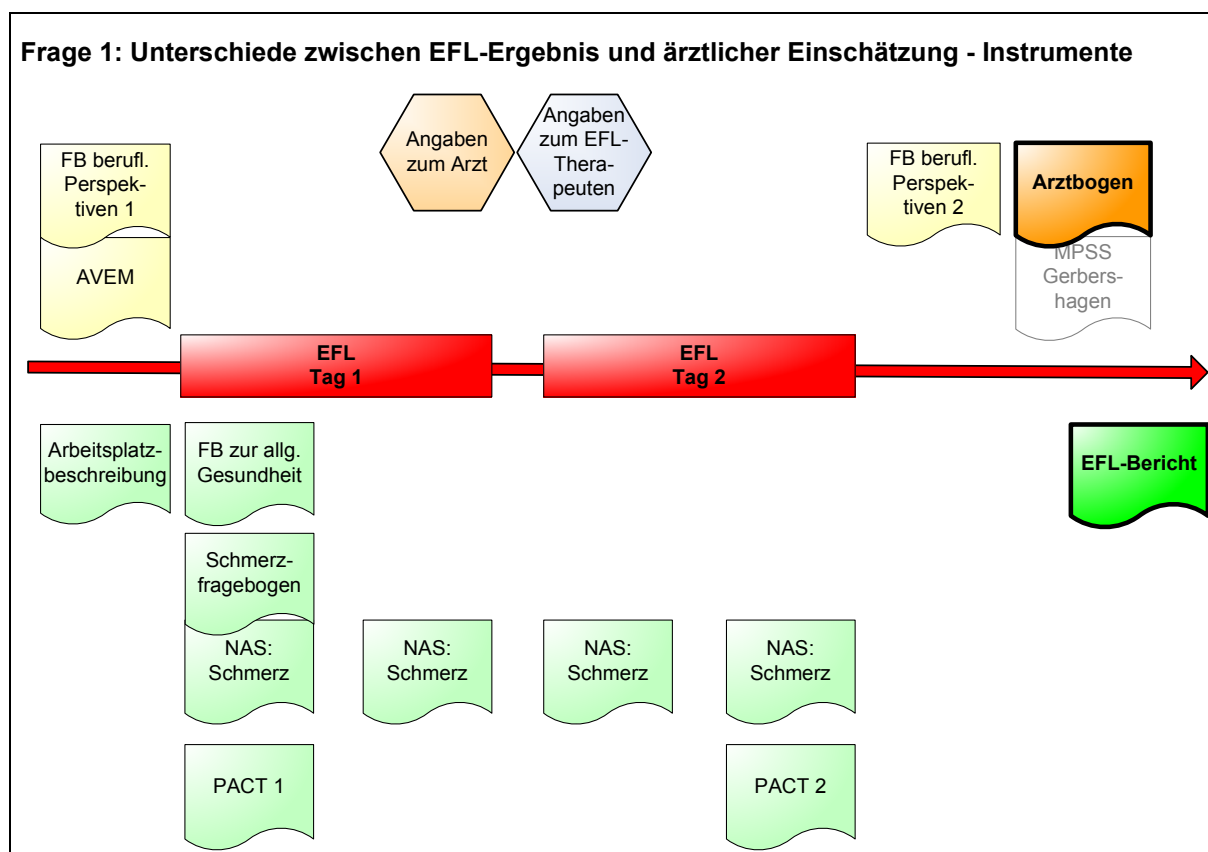


Abbildung 15: Frage 1: Unterschiede zwischen EFL-Ergebnis und ärztlicher Einschätzung – Instrumente
Die Kerninstrumente sind satt, die Instrumente für die Prädiktorenanalyse blass und die für diese Frage nicht verwendeten Instrumente grau dargestellt.

7.2 Ergebnisse

Die berichteten Ergebnisse beruhen auf den Datensätzen der 70 Probanden, von denen sowohl EFL-Bericht als auch Arztbogen vorliegen.

Allgemeine Leistungsfähigkeit nach REFA

Bei dem Vergleich der EFL-Ergebnisse mit den ärztlichen Einschätzungen standen wir vor der Schwierigkeit, dass im EFL-Bericht die REFA-Klassifikation nicht exakt abgebildet wird, so dass wir mit einer Mischkategorie „leicht bis mittelschwer“ arbeiten mussten. Wir haben uns für die in Tabelle 9 dargestellte Codierung entschieden, die dieser Mischkategorie auch einen dem Inhalt entsprechenden Zahlenwert zuweist. In der Folge können auch Differenzen zwischen den Einschätzungen von einer halben Kategorie auftauchen.

EFL	Arztbogen (= REFA-Kategorien)	Codierung für Auswertung
	Sitzend Heben bis maximal 5 kg	1
Leicht	Leicht Heben bis maximal 10 kg	2
Leicht bis mittelschwer		2,5
Mittelschwer	Mittelschwer Heben bis maximal 15 kg	3
Schwer	Schwer Heben bis maximal 40 kg	4
	Sehr schwer Heben bis über 40 kg	5

Tabelle 9: Codierung der Kategorien der allgemeinen Leistungsfähigkeit in EFL-Bericht und Arztbogen

Diese Codierung zugrunde gelegt, stimmen in 35 Fällen (50%) beide Einschätzungen exakt überein. Bei 22 Probanden (31,4%) vermuteten die Ärzte eine höhere Leistungsfähigkeit als das EFL-Ergebnis nahe legt, bei 6 Probanden (8,6%) war es umgekehrt. Im Mittel überstieg die ärztliche Einschätzung das EFL-Ergebnis um 0,21 Kategorien (SD=0,58), was bei einer Korrelation nach Pearson von $r=.62$ ($p=.000^{**}$) einer Effektstärke von 0,59 entspricht. In 7 Arztbögen (10%) fehlt die Angabe der allgemeinen Leistungsfähigkeit. Damit gehen Unterschiede zwischen den beiden Einschätzungen eher in die entgegengesetzte Richtung als bei Erbstöber et al. (2003).

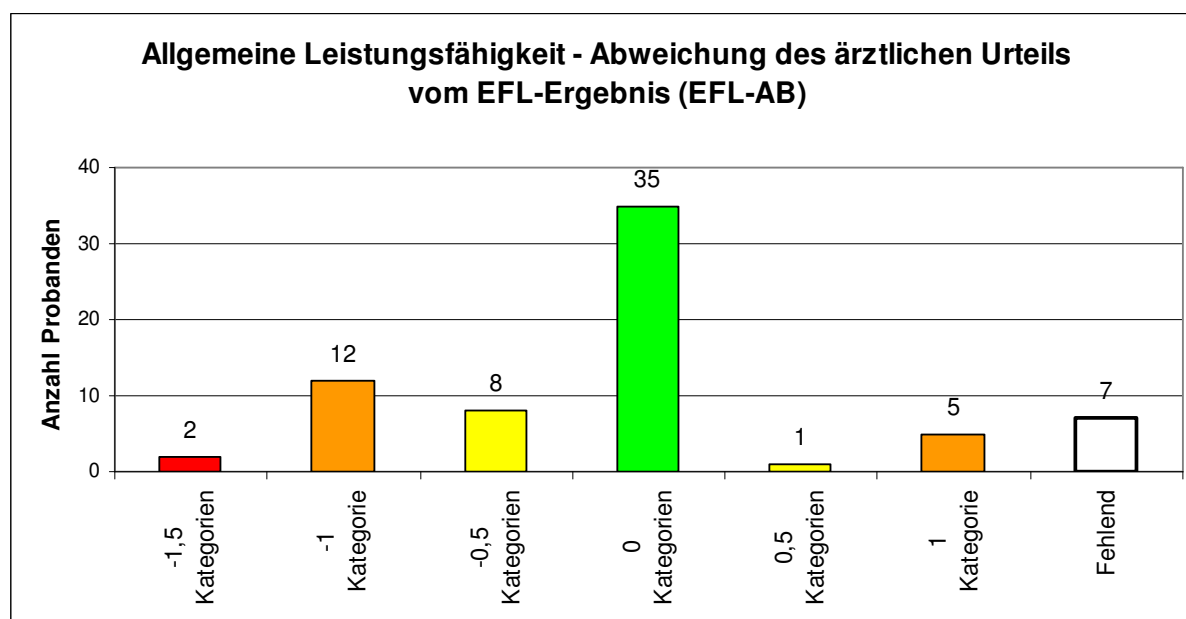


Abbildung 16: Allgemeine Leistungsfähigkeit – Abweichung des ärztlichen Urteils vom EFL-Ergebnis Dargestellt ist die Differenz EFL-Ergebnis minus ärztlicher Einschätzung laut Arztbogen.

Subtests der EFL – Einzelfähigkeiten

Im Vergleich zu der allgemeinen Leistungsfähigkeit kehrte sich das Bild in den Einzelfähigkeiten um: In allen Subtests ergaben die EFL-Ergebnisse im Mittel höhere Leistungsgrenzen als die Einschätzung der Ärzte, fast alle Differenzen erreichen statistische Signifikanz. In den Einschätzungen für die Fähigkeiten *Sitzen vorgeneigt*, *Rotation im Sitzen*, *Kriechen*, *Knien*, *Hockstellung*, *wiederholte Kniebeugen*, *längeres Sitzen* und *Leiter steigen* kamen sogar die maximal möglichen Abweichungen von 4 Kategorien zustande. Den davon betroffenen Patienten würde die entsprechende Anforderung durch die Ärzte komplett verboten, die EFL kam hingegen zu keinerlei Einschränkungen!

Die zum Teil krassen Unterschiede in den Einschätzungen der einzelnen Patienten spiegeln sich auch in den Korrelationen wider. Bedenkt man, dass in beiden Einschätzungen eine professionelle Beurteilung der gleichen Fähigkeiten der gleichen Probanden vorgenommen wurde, so sollte man mit Korrelationen nahe 1 rechnen. Berücksichtigt man, dass Korrelationen von etwa .5 lediglich als mittelstark gelten und Brouwer et al. (2005) eine Mindestkorrelation von .75 von ärztlicher Einschätzung und FCE-Urteil für akzeptabel halten, so erstaunen die insgesamt niedrigen Werte für Kendalls Tau von .57 für die allgemeine Leistungsfähigkeit und maximal .40 für die Einzelfähigkeiten umso mehr. Die Korrelationen für *Tragen linke Hand*, *Sitzen vorgeneigt*, *längeres Stehen*, *Knien*, *Arbeit über Kopf*, *Rotation im Sitzen* sowie *Gehen* und *Leiter steigen* sind mit Werten zwischen .04 und .20 sogar so gering, dass sie noch nicht einmal statistische Signifikanz erreichen. In diesen Einzeltests weisen die Beurteilungen der gleichen Patienten durch Arzt und EFL also keinen überzufälligen Zusammenhang auf!

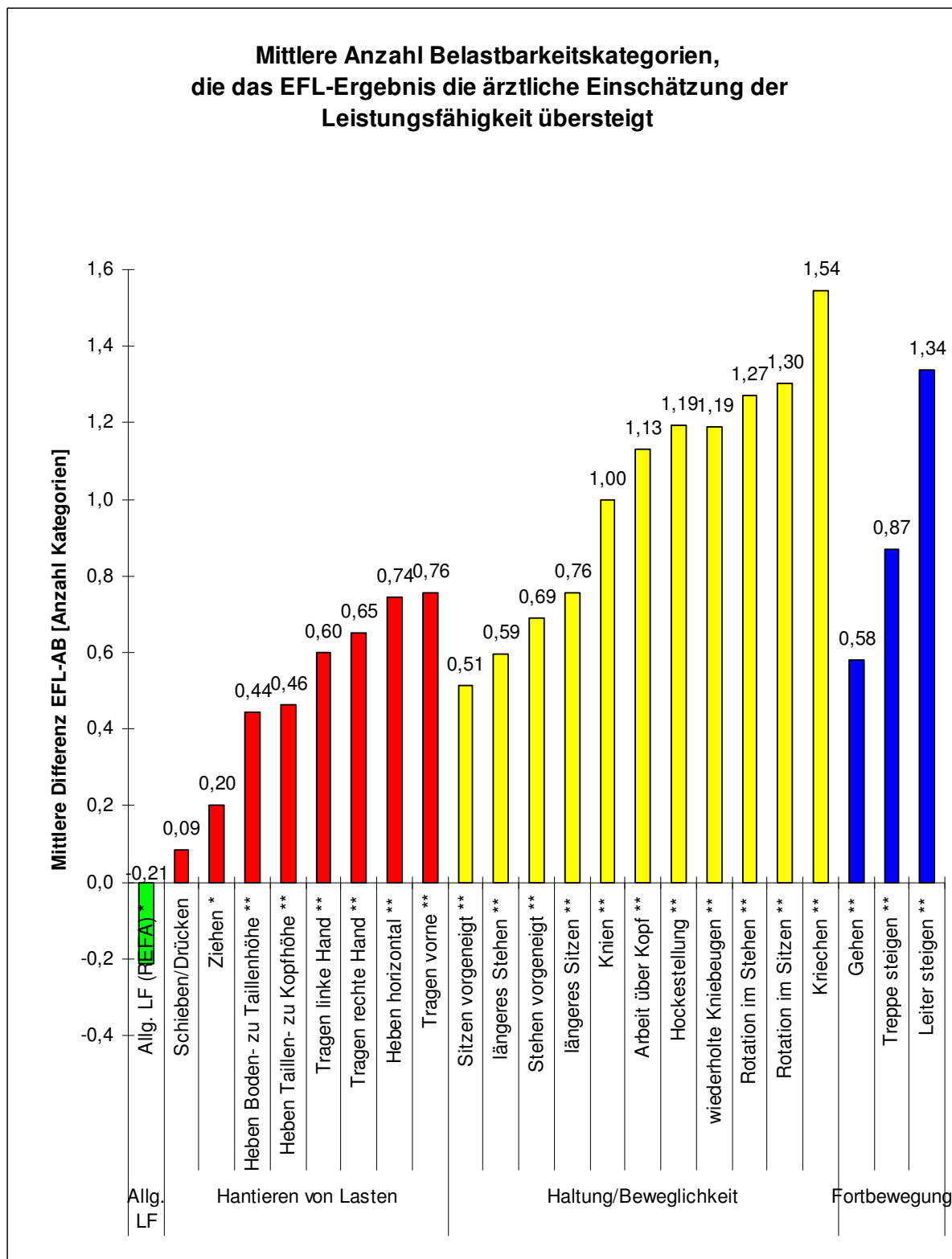


Abbildung 17: Vergleich der EFL-Ergebnisse mit den ärztlichen Einschätzungen: Mittlere Anzahl Belastbarkeitskategorien, die das EFL-Ergebnis die ärztliche Einschätzung überstieg (U-Test nach Wilcoxon; * p<.05; ** p<.01)

Anmerkung: Belastbarkeitskategorien für die Maximalgewichte beim Hantieren von Lasten: *Nie/Keine*, <10 kg, 10-15 kg, 15-40 kg, >40 kg; Kategorien für Haltung/Beweglichkeit und Fortbewegung: *Nie (0%)*, *Gelegentlich (Bis zu 5%)*, *Zeitweise (Ca. 10%)*, *Überwiegend (51-90%)* und *Ständig (>90%)*

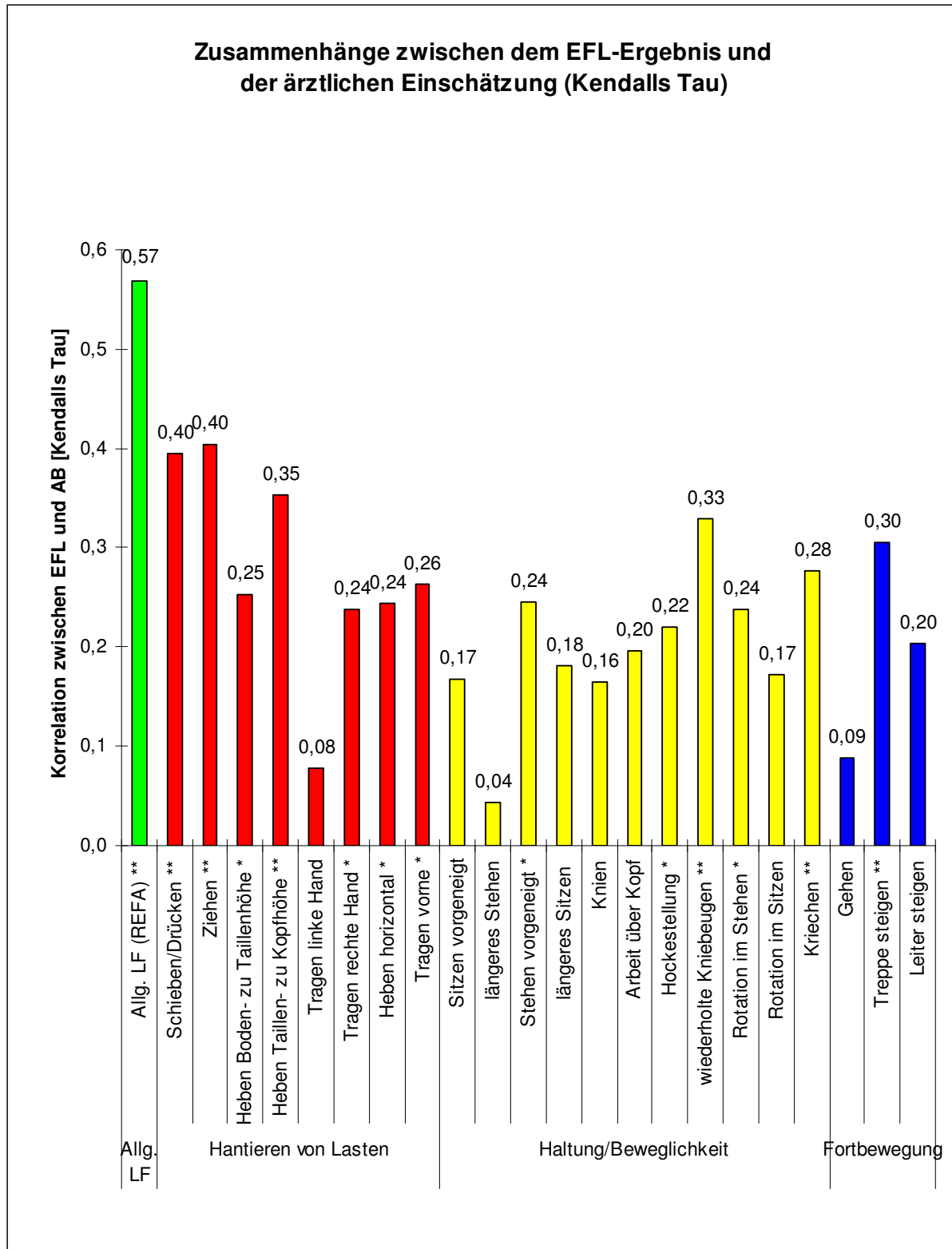


Abbildung 18: Zusammenhänge zwischen EFL-Ergebnis und ärztlicher Übereinstimmung (Kendalls Tau; * p<.05; ** p<.01)

In einigen Tests wird die Leistungsfähigkeit des Probanden nicht in abgestuften Kategorien, sondern nur dichotom als „Norm“ oder „<Norm“ beurteilt. In diesen Tests waren die Diskrepanzen zwischen EFL und Arzturteil erheblich geringer und nur für

die Handkoordination rechts signifikant. Für diese Aufgabe bescheinigten die Ärzte mehr Probanden eine normale Leistungsfähigkeit als sich mit der EFL bestätigen ließ.

	EFL-Ergebnis = Arzturteil		EFL-Ergebnis ≠ Arzturteil		McNemar-Test
	EFL: normal Arzt: normal	EFL: <Norm Arzt: <Norm	EFL: normal Arzt: <Norm	EFL: <Norm Arzt: normal	
Handkraft rechts	57	0	3	10	p=.092
Handkraft links	60	0	1	7	p=.070
Gleichgewicht	64	1	0	3	p=.250
Handkoordination rechts	59	1	1	8	p=.039*
Handkoordination links	62	0	0	5	p=.063

Tabelle 10: Übereinstimmung von EFL und Arzt in den dichotom codierten Variablen
Dargestellt ist die Anzahl der Patienten, für die die jeweilige Urteilskombination zutrifft.

Anmerkung: In den dichotom codierten Variablen wird unterschieden zwischen „normal“ und „< Norm“.

7.3 Diskussion

In den nur dichotom als *normal* oder *<Norm* einzustufenden Fähigkeiten der Handkraft, der Handkoordination und des Gleichgewichtes entsprechen die ärztlichen Einschätzungen weitgehend den EFL-Ergebnissen. Die EFL deckt tendenziell mehr Schwächen auf als die Ärzte vermuten, jedoch ist dieser Unterschied nur in Bezug auf die *Handkoordination rechts* signifikant. Für diese Fähigkeiten liefert die EFL nur in begrenztem Maße Zusatzinformationen. Allerdings kann es zielführend sein, entsprechende Tests – gegebenenfalls isoliert – durchführen zu lassen, wenn diese Fähigkeiten für den Beruf des Patienten besonders wichtig sind und auch nur geringe Zweifel an seiner Leistungsfähigkeit bestehen.

Die dargestellten Ergebnisse für die kategorialen Tests wirken auf den ersten Blick erstaunlich, doch liegen die Übereinstimmungen ungefähr in dem Bereich, den auch Brouwer et al. (2005) in einer ähnlichen Studie, ebenfalls unter Einsatz der EFL, berichtet haben. Die Abweichungen scheinen also nicht auf Messungenauigkeiten zu beruhen, sondern systematisch verursacht zu sein. Als Gründe für die Abweichungen diskutieren Brouwer et al. neben statistischen Besonderheiten die unterschiedlichen Standpunkte und Informationsquellen, die Patient, Arzt und EFL-Testleiter zur

Verfügung haben. Wir gehen über ihre Arbeit hinaus, indem wir die Gesamtkonzeptionen der EFL und des Arzturteils einbeziehen. Dazu leiten wir aus der Diskussion der Testgüte der EFL (Kapitel 3) einige mögliche Erklärungen für die erheblichen Unterschiede in der Einschätzung der Leistungsfähigkeit durch EFL und behandelnden Arzt ab und überprüfen diese soweit möglich mit den in der Studie erhobenen Daten.

Erklärungsversuch 1: Im Gegensatz zu den Ärzten berücksichtigt die EFL keine weiteren Belastungen.

Wie in der Beschreibung der EFL bereits ausgeführt, findet der Test in der Regel unter freundlichen Bedingungen statt. Die Apparaturen sind in einem warmen, trockenen Raum aufgestellt, sie sind sauber und haben einigermaßen ergonomische Formen. Während der Testdurchführung sollte der Therapeut zwar durchaus bestimmt sein, auf der anderen Seite aber auch freundlich bleiben, um die Compliance des Patienten nicht zu gefährden. Auch Zeitdruck, Auseinandersetzungen mit Vorgesetzten, Kollegen und gegebenenfalls Kunden fallen im EFL-Test weg. Geht es um die (Wieder-) Aufnahme einer bestimmten Tätigkeit, so sollen im EFL-Test die Arbeitsbedingungen an dem konkreten Arbeitsplatz zwar so gut wie möglich simuliert werden, jedoch ist das nur begrenzt durchführbar, zumal unter jeder Individualisierung die Standardisierung leidet. Insgesamt sind die Bedingungen also so gestaltet, dass der Proband seine maximale Leistungsfähigkeit entfalten kann. Tatsache ist jedoch, dass der Arbeitsalltag vieler Patienten, gerade der ehemaligen LVA, ganz anders aussieht.

Den Ärzten hingegen ist in der Regel klar, dass Patienten bestimmter Indikationsbereiche beispielsweise mit einigen Umgebungsbedingungen weniger gut zurechtkommen als mit anderen. So führen häufig Arbeiten in Kälte und Nässe zu einer Zunahme der Beschwerden bei Patienten mit muskuloskelettalen Erkrankungen. Vermutlich wissen die Ärzte aufgrund ihrer Erfahrung und aufgrund von Informationen, die sie von den Probanden erhalten, auch von ungünstigen Arbeitsbedingungen ihrer Patienten und beziehen dieses Wissen in ihre Einschätzung der Leistungsfähigkeit mit ein. Wenn diese Hypothese zutrifft, müssten die Ärzte besondere Arbeitsbelastungen ihrer Patienten angeben können und die Abstände der ärztlichen Einschätzung vom EFL-Ergebnis mit zunehmendem Wissen um ungünstige Arbeitsbedingungen zunehmen.

Die Auswertung des Arztbogens zeigt, dass den Ärzten tatsächlich eine ganze Reihe körperlicher und psychischer Zusatzbelastungen bekannt sind. Erwartungsgemäß führen die orthopädisch besonders relevanten Belastungen, die bei Patienten der ehemaligen LVA vermehrt vermutet werden können, die „Hitliste“ an.

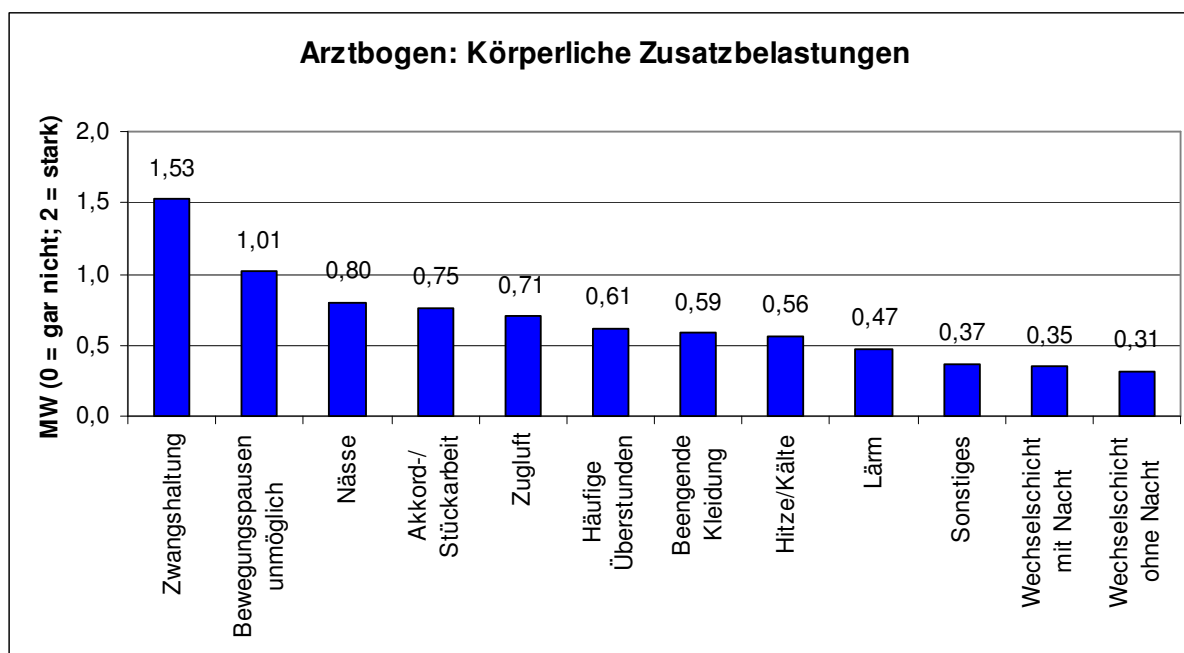


Abbildung 19: Ärztliche Angaben über körperliche Zusatzbelastungen der Patienten an ihren Arbeitsplätzen

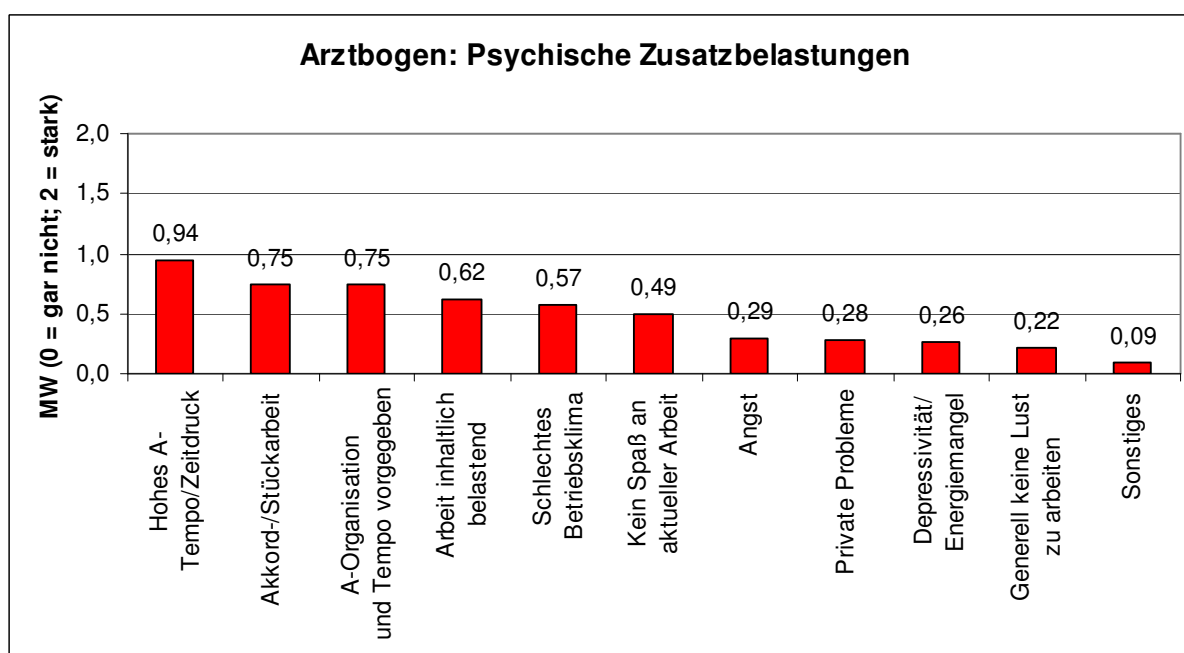


Abbildung 20: Ärztliche Angaben über psychische Zusatzbelastungen der Patienten an ihren Arbeitsplätzen

Die Daten stützen unsere Hypothese: In der Regel weicht die Einschätzung der Ärzte umso stärker vom EFL-Ergebnis ab, je mehr Zusatzbelastungen ihnen bekannt sind – die signifikanten Zusammenhänge (Kendalls Tau) sind großenteils positiv (in der Tabelle grün unterlegt; die signifikant negativen Zusammenhänge sind orange unterlegt).

	Körperliche Belastungen					Psychische Belastungen				
	Zwangshaltungen	Bewegungspausen unmöglich	Nässe	Akkord-/Stückarbeit (k)	Zugluft	Hohes Arbeits-tempo/Zeitdruck	Akkord-/Stückarbeit (p)	A-Organisation und Tempo vorgegeben	Arbeit inhaltlich belastend	Schlechtes Betriebsklima
Allg. LF nach REFA	-.05	-.15	.09	-.01	.02	-.06	-.01	-.03	-.23*	-.10
Hantieren von Lasten										
Schieben/Drücken	.44**	.21	.21	.14	.17	.09	.15	.08	-.07	-.04
Ziehen	.23*	.03	.09	-.14	.07	-.17	-.15	-.12	-.24*	-.21
Heben Boden- zu Taillenhöhe	.29**	.19	.15	.11	.19	.09	.14	.05	-.09	-.07
Heben Taillen- zu Kopfhöhe	.22*	.05	.16	.13	.18	.08	.12	-.01	-.02	-.09
Tragen linke Hand	.28**	.23*	.29**	.08	.18	.16	.05	.12	-.02	.02
Tragen rechte Hand	.23*	.28**	.30**	.13	.23*	.17	.09	.16	.02	.02
Heben horizontal	.29**	.02	.29**	.04	.26*	.06	.17	.12	-.12	.02
Tragen vorne	.29**	.01	.26*	.00	.28*	.04	.09	.10	-.10	-.06
Haltung/ Beweglichkeit										
Sitzen vorgeneigt	-.10	-.02	-.11	-.14	-.01	-.01	-.07	-.01	-.07	-.23*
Längeres Stehen	.37**	.27*	.31**	.27*	.41**	.38**	.38**	.35**	.28*	.23*
Stehen vorgeneigt	-.01	-.09	-.09	-.14	-.01	-.08	-.08	-.03	-.06	-.24*
Längeres Sitzen	.31**	.25*	.30**	.20	.27*	.24*	.19	.11	.09	.00
Knien	.10	.11	.17	-.03	.25*	.22*	.10	.17	.07	.04
Arbeit über Kopf	.10	.08	.11	-.01	.25*	.10	.11	.13	.04	.04
Hockstellung	.12	.16	.27*	.09	.39**	.31**	.24*	.20	.14	.08
Wiederholte Kniebeugen	.13	.09	.18	.02	.24*	.16	.16	.12	.01	.03
Rotation im Stehen	.11	-.06	.09	-.02	.17	.11	.11	.04	-.06	-.04
Rotation im Sitzen	.16	.06	.08	.05	.18	.19	.11	.06	.05	-.03
Kriechen	.24*	.16	.34**	.19	.42**	.42**	.32**	.33**	.25*	.24*
Fortbewegung										
Gehen	.27*	.14	.36**	.25*	.45**	.22	.39**	.28*	.04	.11
Treppe steigen	.16	.13	.18	.09	.29**	.13	.21	.10	-.02	-.01
Leiter steigen	.33*	.34**	.48**	.32**	.43**	.35**	.41**	.38**	.19	0,16

Tabelle 11: Zusammenhänge zwischen dem Wissen der Ärzte um Zusatzbelastungen der Patienten an ihren Arbeitsplätzen und der Abweichung ihrer Einschätzung vom EFL-Ergebnis (Kendalls Tau; * p<.05; ** p<.01)

Es ist unstrittig, dass eine valide Aussage über die zumindest mittelfristige Fähigkeit eines Patienten, an seine Arbeitsstelle zurückzukehren, nur unter Berücksichtigung der individuellen Bedingungen möglich ist. Da jedoch die sozialmedizinische Beurteilung unter Umständen weitreichende Bedeutung für den Patienten hat, scheint es umso wichtiger, dass die Ärzte gut über entsprechende Arbeitsbedingungen informiert sind. Um einen Eindruck davon zu bekommen, in welchem Maß diese Forderung erfüllt ist, haben wir für einige Zusatzbelastungen Angaben von Patienten und Ärzten gleichermaßen erhoben. Ein Vergleich zeigt, dass in den Fragen der Arbeit unter Zwangshaltungen und des Zeitdrucks eine relativ hohe Übereinstimmung besteht, für die meisten Belastungen aber gravierende Unterschiede zu finden sind. Häufig bejahen mehr Patienten als Ärzte eine bestimmte Zusatzbelastung. Es kommt aber durchaus auch vor, dass Ärzte von Belastungen ausgehen, die die Patienten selbst nicht angeben, so dass keine generelle Aggravationstendenz der Patienten angenommen werden muss. In der Abbildung grün dargestellt ist die Anzahl der Fälle, in denen Arzt und Patient das Vorhandensein einer Zusatzbelastung gleichermaßen verneinen (hellgrün) oder bejahen (dunkelgrün). Nimmt der Arzt eine Belastung an, die der Patient nicht bestätigt, wird dieser Fall dem orangefarbenen Bereich zugeordnet. Gibt hingegen der Patient eine Belastung an, die der Arzt nicht sieht, wird der Fall rot markiert.

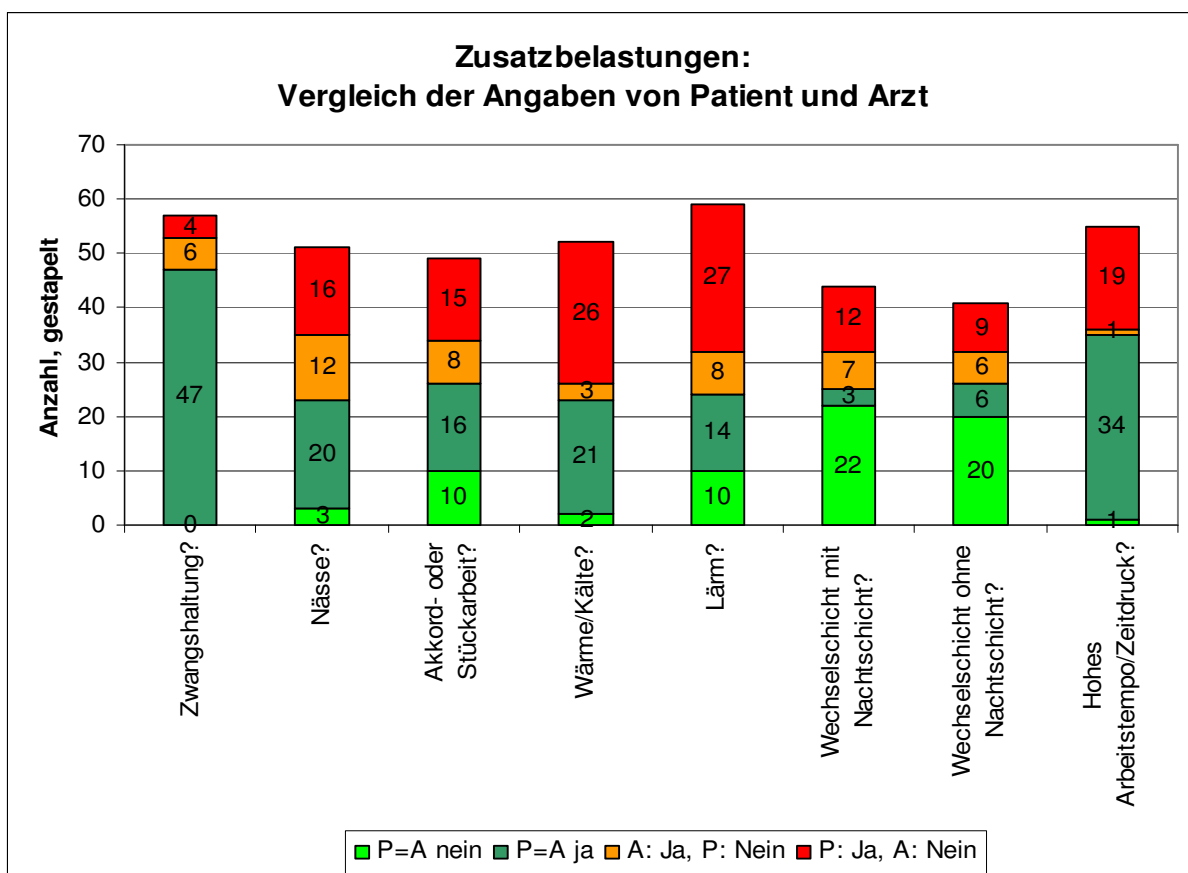


Abbildung 21: Vergleich der Angaben von Ärzten und Patienten über Zusatzbelastungen der Patienten an ihren Arbeitsplätzen

Die Diskrepanzen sind zum Teil erheblich. Es ist zu vermuten, dass selbst bei Patienten, deren Rückkehr an ihren bisherigen Arbeitsplatz aus medizinischen Gründen fraglich ist, dem Thema der individuellen Arbeitsbedingungen momentan nicht ausreichend Raum im Arzt-Patienten-Gespräch zugestanden wird. Dieses ist besonders kritisch zu sehen, wenn dem Arzt lediglich die subjektive – und damit nicht nur möglicherweise verzerrte, sondern in der zeitlichen Begrenzung des Gesprächs vermutlich auch unvollständige – Schilderung des Arbeitsplatzes durch den Patienten vorliegt. Der daraus resultierende Unterschied in der Bewertung der dortigen Belastungen durch Arzt und Patient wird vermutlich zu diskrepanten Überzeugungen führen, ob und wann eine Rückkehr in den Beruf möglich sein wird – und könnte eine Quelle der immer wieder vorkommenden Uneinigkeit über die sozialmedizinische Einschätzung darstellen. Da es auch für den Patienten nicht immer leicht ist, unvorbereitet den Arbeitsplatz exakt zu beschreiben¹², wäre eine Nutzung weiterer Informationsquellen, am besten unter Einsatz standardisierter Erhebungsverfahren, sicherlich von Vorteil. Denkbar wäre neben einer – wünschenswerten, aber in der Regel zu aufwändigen – Begehung des Arbeitsplatzes die Befragung beispielsweise von Betriebsärzten, Vorgesetzten oder Kollegen des betroffenen Patienten. Damit bestünde nicht nur die Chance, Fehleinschätzungen zu revidieren, sondern auch mögliche sprachliche Grenzen des Patienten auszugleichen und eine bessere Basis für einen wichtigen Teil der sozialmedizinischen Einschätzung zu erhalten.

Erklärungsversuch 2: Im Gegensatz zu den Ärzten berücksichtigt die EFL keine Entwicklungen und Prognosen.

Die EFL ermittelt die funktionelle Leistungsfähigkeit eines Probanden an zwei aufeinander folgenden Tagen. Damit ist sie anderen FCE-Systemen sicherlich erheblich überlegen, die die Messung auf nur einen einzigen Tag begrenzen. Dennoch bleibt sie eine Einpunkt-Messung, eine Art Schnappschuss, aufgenommen an einem bestimmten Termin – und streng genommen darf ihre Aussage in der Interpretation auch nur auf diesen Termin bezogen werden. Allerdings ist als Indikationsvoraussetzung das Erreichen medizinischer Stabilität formuliert, die ein gewisses Niveau als gegeben annehmen lässt. Kaum beachtet werden jedoch Tagesschwankungen, die auch bei medizinisch stabilen Patienten durchaus vorkommen, und gar keine Berücksichtigung finden langfristige Prognosen. Abweichungen des ärztlichen Urteils vom EFL-Ergebnis nach unten könnten also auch darin begründet sein, dass gerade erfahrene Ärzte den Patienten über die Zeit der Reha hinweg aufmerksam beobachten (lassen) und auf diese Weise einen Eindruck über seine Leistungsfähigkeit und deren Entwicklung über immerhin knapp

¹² Wissen Sie genau, wie viel Kilogramm groß die schwerste Last wiegt, die Sie heben müssen, oder wie hoch die Platte Ihres Schreibtischs eingestellt ist, ohne beides nachzuprüfen?

drei Wochen in verschiedenen Kontexten erhalten, der die Grundlage ihrer Einschätzung darstellt. Ebenso können erfahrene Ärzte vermutlich am besten abschätzen, dass ein Patient momentan einer bestimmten Belastung durchaus standhalten kann („limitations“), dass er dieser aber trotzdem nicht dauerhaft ausgesetzt werden sollte, weil dann in absehbarer Zeit erhebliche Schäden drohen („restrictions“). Vermutlich beziehen sie auch solche Vorhersagen in ihre Einschätzungen mit ein und korrigieren diese nach unten – umso stärker, je mehr Berufserfahrung sie haben, die ihnen hilft, Prognosen auch für längere Zeiträume zu stellen.

Mit durchschnittlich 20,7 (SD=7,6; Min=12; Max=33) Berufsjahren haben sehr erfahrene Ärzte die Leistungsfähigkeit der Patienten beurteilt. Sie hatten im Schnitt seit 14 Jahren mit sozialmedizinischen Beurteilungen zu tun (SD=8; Min=3; Max=31) und arbeiteten seit 3,75 Jahren (SD=1,7; Min=2; Max=6) mit EFL-Tests.

Die in Tabelle 12 dargestellten Befunde bestätigen die oben aufgestellte Hypothese: Je länger die Ärzte ihren Beruf bereits ausübten, umso stärker wichen ihre Einschätzungen von den Ergebnissen der EFL ab. Ein umgekehrtes Bild ergab sich hingegen für die Arbeit mit EFLs: Größere EFL-Erfahrung der Ärzte geht mit einer geringeren Abweichung ihrer Urteile von den EFL-Ergebnissen einher – offenbar übernehmen sie zum Teil die Sichtweise der EFL. Der dritte Komplex, die sozialmedizinische Erfahrung, zeigt keine eindeutige Effektrichtung.

Eine weitere Erklärungsmöglichkeit für die mit zunehmender Berufserfahrung steigenden Abweichung der ärztlichen Einschätzung vom EFL-Ergebnis liegt auf einer ganz anderen Ebene: Erfahrene Ärzte sind zwangsläufig im Schnitt älter als weniger erfahrene – und haben vermutlich in ihrem Leben selbst mehr Erfahrung mit Beschwerden, Schmerzen und Beeinträchtigungen gemacht. Können sie die Situation der Patienten einfach besser nachvollziehen und wollen ihnen deswegen eher weniger zumuten?

	Arztbogen-Arzt		
	Berufserfahrung (Jahre)	Erfahrung mit Sozialmedizin (Jahre)	Erfahrung mit EFL (Jahre)
Allgemeine LF nach REFA	.23*	.05	-.15
Hantieren von Lasten			
Schieben/Drücken	.07	-.10	-.02
Ziehen	-.03	-.27*	-.01
Heben Boden- zu Taillenhöhe	.07	.09	-.04
Heben Taillen- zu Kopfhöhe	.27**	.15	-.30*
Tragen linke Hand	-.06	.09	.05
Tragen rechte Hand	-.07	.10	.06
Heben horizontal	.28**	.10	-.29*
Tragen vorne	.13	.02	-.04
Haltung/Beweglichkeit			
Sitzen vorgeneigt	.22*	-.01	-.14
Längeres Stehen	.15	-.02	.02
Stehen vorgeneigt	.07	-.23*	-.08
Längeres Sitzen	.28**	.23*	-.09
Knien	.35**	.14	-.14
Arbeit über Kopf	.15	.03	-.12
Wiederholte Kniebeugen	.28**	.03	-.12
Hockstellung	.34**	.17	-.11
Rotation im Stehen	.30**	.08	-.34**
Rotation im Sitzen	.41**	.29**	-.33**
Kriechen	.33**	.25*	-.12
Fortbewegung			
Gehen	.15	.04	.03
Treppe steigen	.34**	.07	-.08
Leiter steigen	.36**	.18	-.06

Tabelle 12: Zusammenhänge zwischen der Erfahrung der Ärzte und der Abweichung ihrer Einschätzung vom EFL-Ergebnis (Kendalls Tau; * p<.05; ** p<.01)

Erklärungsversuch 3: Patientenbezogene Prädiktoren

Wie in Kapitel 3.4 über die Konstruktvalidität der EFL dargestellt, haben einige Studien gezeigt, dass das EFL-Ergebnis nicht nur von der reinen funktionellen Leistungsfähigkeit des Probanden abhängt, sondern auch von einigen psychosozialen Merkmalen, die er mitbringt.

Beeinflussen solche Variablen auch die Unterschiede zwischen EFL-Ergebnis und ärztlicher Einschätzung? Aus unseren Daten konnten wir eine ganze Reihe Probandencharakteristika auf ihren prädiktiven Wert hinsichtlich dieser Frage überprüfen. Die wichtigsten Ergebnisse möchten wir an dieser Stelle berichten, verzichten aber auf Vollständigkeit, um den Rahmen der Arbeit nicht zu sprengen.

Soziodemographische Daten

Für die Unterschiedlichkeit von EFL-Ergebnis und Arzturteil spielten Geschlecht, Alter, Partnerschaft, Anzahl der Personen im Haushalt und Berufsausbildung keine wesentliche Rolle.

Bedeutsam war jedoch, an welchem unserer Erhebungsorte der Proband seine Reha und den EFL-Test absolvierte – anhand von U-Tests nach Mann-Whitney ergeben sich signifikant höhere Abweichungen des Arzturteils von EFL-Ergebnis für die Tests *Heben Boden- zu Taillenhöhe*, *Heben Taillen- zu Kopfhöhe*, *Tragen rechte Hand*, *Tragen linke Hand*, *Längeres Stehen*, *Gehen* und *Leiter steigen* für die Probanden aus der Klinik Münsterland. Lediglich für den Test *Sitzen vorgeneigt* wichen die Ärzte in Bad Pyrmont stärker vom EFL-Ergebnis ab als ihre Kollegen in Bad Rothenfelde (Tabelle 13).

Subtest	MW _{KML}	Mittlerer Rang _{KML}	MW _{BP}	Mittlerer Rang _{BP}	U	p
Heben Boden-Taillenhöhe	0,684	41,22	0,156	28,70	390,5	.006**
Taillen- Kopfhöhe	0,649	39,32	0,250	30,00	432,0	.040*
Tragen rechte Hand	0,947	41,92	0,290	26,52	326,0	.001**
Tragen linke Hand	0,947	42,96	0,188	26,64	324,5	.000*
Sitzen vorgeneigt	0,211	30,71	0,875	41,19	426,0	.027*
Längeres Stehen	0,811	40,26	0,344	28,92	397,5	.011*
Gehen	0,771	39,19	0,375	28,33	397,5	.011*
Leiter steigen	1,611	39,65	1,031	28,70	378,5	.010**

Tabelle 13: Abweichungen der ärztlichen Einschätzungen vom EFL-Ergebnis im Ortsvergleich
Dargestellt sind die mittleren Anzahlen von Kategorien, die die Arzturteile von den EFL-Ergebnissen abweichen, bzw. deren mittlere Ränge, aus denen die Teststatistiken für U-Tests errechnet werden. (U-Tests nach Wilcoxon; * p<.05; ** p<.01)

Berufliche Situation

Die erhobenen „harten“ Daten der Erwerbssituation der Probanden zeigten höchstens niedrige und unsystematische Zusammenhänge zu den Differenzen in den Einschätzungen von Arzt und EFL, sie schienen eher statistischen als inhaltlichen Ursachen zu unterliegen¹³. Dieses galt sowohl für die berufliche Situation im engeren Sinn einschließlich der Fragen nach dem Bezug von Rente oder dem Laufen eines Rentenverfahrens wie auch für die Dauer der Firmenzugehörigkeit, die Stellung in der Firma und das Vorhandensein eines gültigen Arbeitsvertrages.

Bedeutsam hingegen war das subjektive Erleben der Arbeit: Bewerteten die Probanden ihre Arbeit positiv in dem Sinne, dass sie ihnen Freude machte und sie sich nicht überlastet fühlten, sondern mit der Arbeit an sich, der Beziehung zu ihrem Vorgesetzten und ihrem Lohn oder Gehalt zufrieden waren, überstieg das EFL-Ergebnis die ärztliche Einschätzung in besonderem Maße. Möglicherweise strengten sich diese Probanden in der EFL stärker als andere an, um an ihren Arbeitsplatz zurückkehren zu dürfen – und belasteten sich damit über das Niveau hinaus, zu dem ihre Ärzte ihnen raten würden. Ebenso ging eine zuverlässige Leistungsbereitschaft damit einher, dass das EFL-Ergebnis zum Teil die ärztliche Einschätzung überstieg. Dieses galt für die Tests *Heben horizontal*, *Tragen vorne*, *Arbeit über Kopf*, *Sitzen vorgeneigt*, *Stehen vorgeneigt*, *Rotation im Stehen*, *Knien*, *Wiederholte Kniebeugen* und *Treppe steigen*.

Bei Patienten mit ungünstiger Motivationslage in Bezug auf ihre Arbeit waren die Abweichungen zwischen EFL-Ergebnis und ärztlicher Einschätzung eher gering. Je sicherer ein Patient im AVEM dem Risikomuster Resignation zugeordnet wurde, umso besser stimmte die Einschätzung seines Arztes mit seinem EFL-Ergebnis in Bezug auf die allgemeine Leistungsfähigkeit nach REFA und verschiedenen Einzeltests überein. Die Werte für Kendalls Tau bewegten sich in diesen Fällen zwischen $-.201$ ($p=.034^*$; *Knien*) und $-.299$ ($p=.002^{**}$; *Treppe steigen*).

Eigene Einschätzung der beruflichen Fähigkeiten und berufliche Perspektive

Nicht nur die Freude am Beruf, auch die durch die Patienten selbst vor und nach der EFL eingeschätzten berufsbezogenen Fähigkeiten zeigten Zusammenhänge zur Abweichung der ärztlichen Einschätzung vom EFL-Ergebnis. Vor allem Unterschiede in den Einschätzungen der Fähigkeiten *Tragen*, *Sitzen vorgeneigt* und *Stehen vorgeneigt* korrelierten mit den subjektiven Urteilen der Probanden in Bezug auf ihre *Beeinträchtigung durch ihre Krankheit in Arbeit und Beruf* und die *maximale Schwere der Tätigkeit, die sie meinen, ohne Gesundheitsgefährdung ausführen zu können*.

¹³ Bei einem Signifikanzniveau von 5% muss man von 5% irrtümlich signifikant gewordenen Ergebnissen ausgehen. Ergeben sich in einer Ergebnismatrix Signifikanzen dieses Umfangs, die eher unsystematisch gestreut erscheinen, so kann man einen verfahrensspezifischen statistischen Fehler vermuten.

Die entsprechenden Korrelationen liegen zwischen -0,26 und 0,23. Weitere Zusammenhänge ließen sich zu dem *vermuteten Zeitpunkt der Fähigkeit zur Wiederaufnahme der/einer Erwerbstätigkeit nach der Reha* nachweisen sowie zu der Wahrscheinlichkeit, dass ein Patient die *Tätigkeit, die er in einem Jahr innehat, ohne gesundheitliche Bedenken* verrichten können (Kendalls Tau zwischen -0,22 und 0,19). Da diese Korrelationen jedoch in Höhe und Richtung stark wechselten, ist eine zusammenfassende Aussage kaum möglich und der prädiktive Wert scheint begrenzt. Die stärkste Vorhersagekraft hat der *vermutete Zeitpunkt der Fähigkeit zur Wiederaufnahme der/einer Erwerbstätigkeit nach der Reha* in der Einschätzung nach der EFL. Er erklärt bis zu 21% der Differenzen zwischen EFL und ärztlicher Einschätzung in den Subtests.

Die konkreteren beruflichen Perspektiven, wie der vermutete berufliche Status oder die berufliche Situation in einem Jahr, spielten eine ebenso geringe Rolle wie die Wunschtätigkeit nach der Reha.

Gesundheit

Eine potenzielle Erklärung für Differenzen zwischen EFL und Arztergebnis könnte darin liegen, dass die Ärzte im Gegensatz zur EFL in ihrer Einschätzung gesundheitliche **Risikofaktoren** ihrer Patienten berücksichtigen. Dieser Sachverhalt schlägt sich aber offenbar nicht in den Beurteilungen der Leistungsfähigkeit der Patienten nieder, denn die meisten erhobenen Risikofaktoren korrelierten kaum mit der Abweichung zwischen beiden Einschätzungen. Lediglich *Trainingsmangel*, das *Vorliegen einer schweren Herzerkrankung bei engen Verwandten vor dem 60. Lebensjahr* und *Lebensalter >40 Jahre* zeigten prädiktives Potenzial (Tabelle 14). *Übergewicht*, *eigene Herzprobleme* des Probanden mit und ohne entsprechender Medikation und *Rauchen* hingegen spielten keine Rolle. *Diabetes* und *Epilepsie* kamen nur jeweils einmal vor, so dass kein Signifikanztest durchgeführt werden konnte.

Die Frage, ob ein Proband zur Zeit des EFL-Tests **arbeitsunfähig** war oder nicht, war für die Divergenz zwischen EFL-Ergebnis und Arzturteil unerheblich. Im Gegensatz dazu war die Dauer der Arbeitsunfähigkeitszeiten durchaus bedeutsam: Je länger die Probanden vor dem EFL-Test ihre Tätigkeit nicht hatten ausüben können, umso stärker überschritt das EFL-Ergebnis die ärztliche Einschätzung. Signifikante Korrelationen ließen sich zwischen der Anzahl der AU-Tage vor der EFL und der Differenz der Einschätzungen in Bezug auf *Heben Taillen- zu Kopfhöhe* ($\tau=.236$; $p=.023^*$), *Heben horizontal* ($\tau=.212$; $p=.037^*$), *Gehen* ($\tau=.265$; $p=.013^*$) und *Leiter steigen* ($\tau=.203$; $p=.045^*$) nachweisen.

Risikofaktor Trainingsmangel (TM = Trainingsmangel)						
Subtest	MW_{kein TM}	Mittlerer Rang_{kein TM}	MW_{TM}	Mittlerer Rang_{TM}	U	p
Schieben/ Drücken	-0,111	29,19	0,214	38,74	410,0	.031*
Ziehen	-0,148	25,98	0,439	40,11	323,5	.001**
Rotation im Sitzen	1,630	39,96	1,073	30,90	406,0	.050*
Kriechen	1,852	40,02	1,275	29,94	377,5	.030*
Längeres Sitzen	1,148	41,93	0,476	30,55	380,0	.012*
Risikofaktor: Schwere Herzerkrankung von nahen Verwandten vor dem 60. Lebensjahr (HEV = Herzerkrankung von Verwandten)						
Subtest	MW_{keine HEV}	Mittlerer Rang_{keine HEV}	MW_{HEV}	Mittlerer Rang_{HEV}	U	p
Heben Boden- zu Taillenhöhe	0,509	36,55	-0,091	23,86	196,5	.037*
Gehen	0,630	35,09	0,182	22,73	184,0	.025*
Treppe steigen	0,965	36,60	0,364	23,64	194,0	.036*
Leiter steigen	1,436	36,43	0,455	18,86	141,5	.004**
Risikofaktor: Lebensalter >40 Jahre						
Subtest	MW_{<40}	Mittlerer Rang_{<40}	MW_{>40}	Mittlerer Rang_{>40}	U	p
Allgemeine LF (REFA)	-0,067	36,37	-0,375	26,94	1091,0	.022*

Tabelle 14: Abweichungen der ärztlichen Einschätzungen vom EFL-Ergebnis in Abhängigkeit von verschiedenen Risikofaktoren der Patienten

Dargestellt sind jeweils die mittlere Anzahl von Kategorien, die die Arzturteile von den EFL-Ergebnissen abweichen, bzw. die mittleren Ränge, aus denen die Teststatistiken für U-Tests errechnet werden. Es wurden nur die signifikanten Werte angegeben.

(U-Tests nach Wilcoxon; * p<.05; ** p<.01)

Auch die **Lokalisation des arbeitsbezogenen Problems** hatte nur teilweise Auswirkungen auf die Unterschiedlichkeit der beiden Einschätzungen: U-Tests nach Mann-Whitney ergaben, dass bei Patienten mit Rückenproblemen das Arzturteil stärker vom EFL-Ergebnis in Bezug auf *Sitzen vorgeneigt* und *Stehen vorgeneigt* abwich als bei Patienten ohne Rückenschmerzen. Bei Patienten mit allgemeiner Kraft- oder Ausdauerschwäche war die Richtung umgekehrt. Hatten Patienten Probleme mit den unteren Extremitäten, war die Abweichung in Bezug auf die

Fähigkeit, in der Hocke zu verharren, kleiner als bei Patienten ohne entsprechende Diagnose.

Arbeitsbezogenes Problem: Rücken						
Subtest	MW_{nein}	Mittlerer Rang_{nein}	MW_{ja}	Mittlerer Rang_{ja}	U	p
Sitzen vorgeneigt	0,952	42,95	0,277	30,72	316,0	.015*
Stehen vorgeneigt	1,350	44,90	0,326	28,54	232,0	.001**
Allgemeine Leistungsfähigkeit	-0,048	37,43	-0,329	28,46	306,0	.040*
Arbeitsbezogenes Problem: Untere Extremität						
Subtest	MW_{nein}	Mittlerer Rang_{nein}	MW_{ja}	Mittlerer Rang_{ja}	U	p
Hocke-stellung	1,385	35,94	0,308	21,23	185,00	.010**
Arbeitsbezogenes Problem: Kraft-, Ausdauerschwäche						
Subtest	MW_{nein}	Mittlerer Rang_{nein}	MW_{ja}	Mittlerer Rang_{ja}	U	p
Sitzen vorgeneigt	0,339	32,08	1,444	50,39	122,5	.008**
Stehen vorgeneigt	0,509	31,45	1,444	46,50	139,5	.024*

Tabelle 15: Mittlere Differenzen zwischen den EFL-Ergebnissen und den ärztlichen Einschätzungen in Abhängigkeit von der Lokalisation des arbeitsbezogenen Problems.

Dargestellt sind jeweils die mittlere Anzahl von Kategorien, die die Arzturteile von den EFL-Ergebnissen abweichen, bzw. die mittleren Ränge, aus denen die Teststatistiken für U-Tests errechnet werden. (U-Tests nach Wilcoxon; * $p < .05$; ** $p < .01$)

Der **Schmerz** gilt als eine Variable, die sowohl durch die körperliche Erkrankung als auch durch psychosoziale Faktoren beeinflusst wird. Er stellte sich als ausgesprochen wichtiger Prädiktor für die Abweichung des ärztlichen Urteils vom EFL-Ergebnis dar. Zusammenfassend lassen sich stärkere Unterschiede in den Einschätzungen mit geringeren Schmerzangaben der Patienten für die letzten 7 Tage vor der EFL wie auch zu Beginn und zum Ende eines jeden Testtages in Verbindung bringen. In Tabelle 16 und Tabelle 17) sind die Zusammenhänge zwischen den Schmerzitems und den Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis dargestellt. Der Übersichtlichkeit halber sind haben wir in den Tabellen nur die EFL-Einzeltests aufgenommen, die mindestens eine signifikante Korrelation aufgewiesen haben.

	Schlimmster Schmerz	Tage mit Schmerz	Tage mit sehr starkem Schmerz	Tage mit Schmerzmedikamenten
Schieben/Drücken	-.007	-.130	-.210*	-.101
Ziehen	.013	-.262*	-.149	-.219*
Arbeit über Kopf	-.185*	-.089	.025	-.183
Sitzen vorgeneigt	-.227*	-.137	-.089	-.179
Stehen vorgeneigt	-.136	-.131	-.034	-.307**
Rotation im Stehen	-.174	-.106	-.158	-.247*
Allgemeine LF (REFA)	-.159	-.103	-.154	-.257*

Tabelle 16: Signifikante Zusammenhänge zwischen Angaben zu Schmerz in den letzten 7 Tagen und Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis (Kendalls Tau; * p<.05; ** p<.01).

	Schmerz Anfang 1. Testtag	Schmerz Ende 1. Testtag	Schmerz Anfang 2. Testtag	Schmerz Ende 2. Testtag	Schmerzzunahme
Tragen rechte Hand	.097	.165	.170	.230*	.193*
Tragen linke Hand	.133	.168	.168	.225*	.146
Arbeit über Kopf	-.043	-.104	-.101	-.236*	-.230*
Sitzen vorgeneigt	-.249**	-.321**	-.328**	-.354**	-.196*
Stehen vorgeneigt	-.248**	-.254**	-.260**	-.263**	-.118
Rotation im Stehen	-.126	-.182	-.103	-.219*	-.154
Knien	-.145	-.221*	-.077	-.192*	-.083
Wiederholte Kniebeugen	-.191*	-.214*	-.184*	-.271**	-.155
Allgemeine LF (REFA)	-.135	-.178	-.237*	-.260**	-.250*

Tabelle 17: Signifikante Zusammenhänge zwischen Angaben zu Schmerz zu Beginn und Ende der EFL-Testtage und Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis (Kendalls Tau; * p<.05; ** p<.01).

EFL-Erwartungen und -Befürchtungen

Wie weiter oben dargestellt, korreliert eine positive Einstellung der Probanden zu seiner Arbeit positiv mit der Abweichung der ärztlichen Einschätzung vom EFL-Ergebnis. Spielen auch Erwartungen und Befürchtungen des Probanden hinsichtlich des Tests selbst, die sich vermutlich im Testverhalten niederschlagen, eine Rolle?

In den Daten lassen sich kaum signifikante Zusammenhänge zwischen den im Fragebogen zu beruflichen Perspektiven 1 erhobenen **Erwartungen und Befürchtungen** und den Differenzen in den Einschätzungen finden, nennenswert sind lediglich zwei Items:

- Je größer die Erwartung der Patienten war, dass der EFL-Test zu einem Nachweis der Erwerbsunfähigkeit und der Notwendigkeit einer Berentung führen würde, umso stärker überstieg das EFL-Ergebnis die ärztliche Einschätzung. Die signifikanten Korrelationen bewegten sich zwischen .234 ($p=.043^*$; zu *Hockstellung*) und .315 ($p=.007^{**}$; zu *Leiter steigen*).
- Eine stärkere Befürchtung, sich gefährlich zu verletzen, korrespondierte hingegen eher mit einer geringeren Abweichung des ärztlichen Urteils vom EFL-Ergebnis. Für diesen Prädiktor ergaben sich signifikante Korrelationen von -.221 ($p=.046^*$; zu *Kriechen*) bis -.395 ($p=.000^{**}$; zu *Knien*).

Auch einige Aspekte des tatsächlichen **Testerlebens** erwiesen sich als relevant. Das EFL-Ergebnis überstieg die ärztliche Einschätzung umso mehr,...

- je weniger schmerzhaft der Test erlebt wurde
($T_{\min}=-.229$; $p=.037^*$; zu *Rotation im Stehen*; $T_{\max}=-.359$; $p=.001^{**}$; zu *Sitzen vorgeneigt*).
- je stärker die EFL als Herausforderung gesehen wurde („Konnte anderen zeigen, was ich kann“)
($T_{\min}=.271$; $p=.012^*$; zu *Stehen vorgeneigt*; $T_{\max}=.306$; $p=.004^{**}$; zu *Sitzen vorgeneigt*).
- je weniger stark die Angst geblieben ist, sich gefährlich zu verletzen
($T_{\min}=-.226$; $p=.047^*$; zu *Heben Boden- zu Taillenhöhe.*; $T_{\max}=-.246$; $p=.034^*$; zu *Ziehen*)

Globale im Test gewonnene Erkenntnisse hingegen spielten kaum eine Rolle, ebenso wie die Veränderung der Einschätzung der eigenen Leistungsfähigkeit und die Bereitschaft, noch einmal an einem EFL-Test teilzunehmen.

Leistungsniveau der Probanden

Eine Überprüfung der Differenzen auf Abhängigkeit zwischen Arzturteil und EFL-Ergebnis vom Leistungsniveau des Probanden ist aus zwei Gründen wichtig:

Zum einen können sich bei Patienten mit extremen Leistungsniveaus statistische Verzerrungen einschleichen, die sich auf die mittleren Diskrepanzen beider Urteile niederschlagen, zum anderen könnten inhaltliche Ursachen wie unterschiedliche „Sensibilitätsspitzen“ der beiden „Messinstrumente“ Arzt und EFL zu systematischen Effekten führen.

Da die Differenz der beiden Einschätzungen eine direkte Funktion der einzelnen Beurteilungen der Leistungsfähigkeit darstellt, ist eine Regressionsanalyse vom Leistungsniveau auf die Unterschiedlichkeit der Urteile nicht möglich. Stattdessen haben wir uns für einen Mediansplit der Probanden bezüglich der Leistungsfähigkeit entschieden und geschaut, ob sich die auf diese Weise definierten Probandengruppen in der Abweichung des ärztlichen Urteils vom EFL-Ergebnis unterscheiden. Leider können wir nicht sagen, ob das EFL-Ergebnis oder die ärztliche Einschätzung das Leistungsniveau des Probanden valider beschreibt, so dass wir nacheinander beide als Prädiktorvariablen herangezogen und für den Mediansplit genutzt haben.

In Abbildung 22 und Abbildung 23 zeigt sich, dass in fast allen Subtests die Abweichungen der ärztlichen Einschätzungen vom EFL-Ergebnis bei den Probanden größer sind, die im EFL-Ergebnis ein hohes Leistungsniveau erreicht haben, und deren Leistungsfähigkeit von den Ärzten als eher gering eingestuft wird.

Nimmt man das EFL-Ergebnis als Operationalisierung der Leistungsfähigkeit, so wird die starke Schiefe der Verteilung der Leistungswerte in einigen Subtests deutlich – in 13 Tests entsprach der Median dem oberen Maximum, so dass gar kein Mediansplit durchgeführt werden konnte und ein Vergleich zweier Gruppen ausgeschlossen war. Oft haben die Probanden laut EFL keinerlei Einschränkungen in diesen Fähigkeiten, so dass eine Abweichung des ärztlichen Urteils nach oben gar nicht möglich ist (Deckeneffekt).

In der Regel sind von diesem Effekt Fähigkeiten betroffen, die nicht wie das Heben durch die Ermittlung des kurzfristigen Leistungsmaximums und eine Extrapolation auf langfristig zumutbare reduzierte Gewichte ermittelt werden. Vielmehr tritt die Problematik in den Haltungs-/Beweglichkeits- und Fortbewegungstests auf, in denen aus einer zeitweise tolerierten Belastung auf eine erheblich längere Toleranzdauer geschlossen wird, ohne dass die Belastungsstärke reduziert werden könnte. Damit scheinen gerade die Einschätzungen aufgrund von Testergebnissen, die auf einer solchen (fragwürdigen?; vgl. Kapitel 3.4) Berechnung beruhen, von den Ärzten nicht geteilt zu werden. Vermutlich „bestehen“ eine ganze Reihe Probanden, denen die

Ärzte die in Frage stehenden Fähigkeiten nicht grenzenlos zutrauen, den erheblich einfacheren, weil zeitlich begrenzten EFL-Subtest noch ohne Schwierigkeiten.

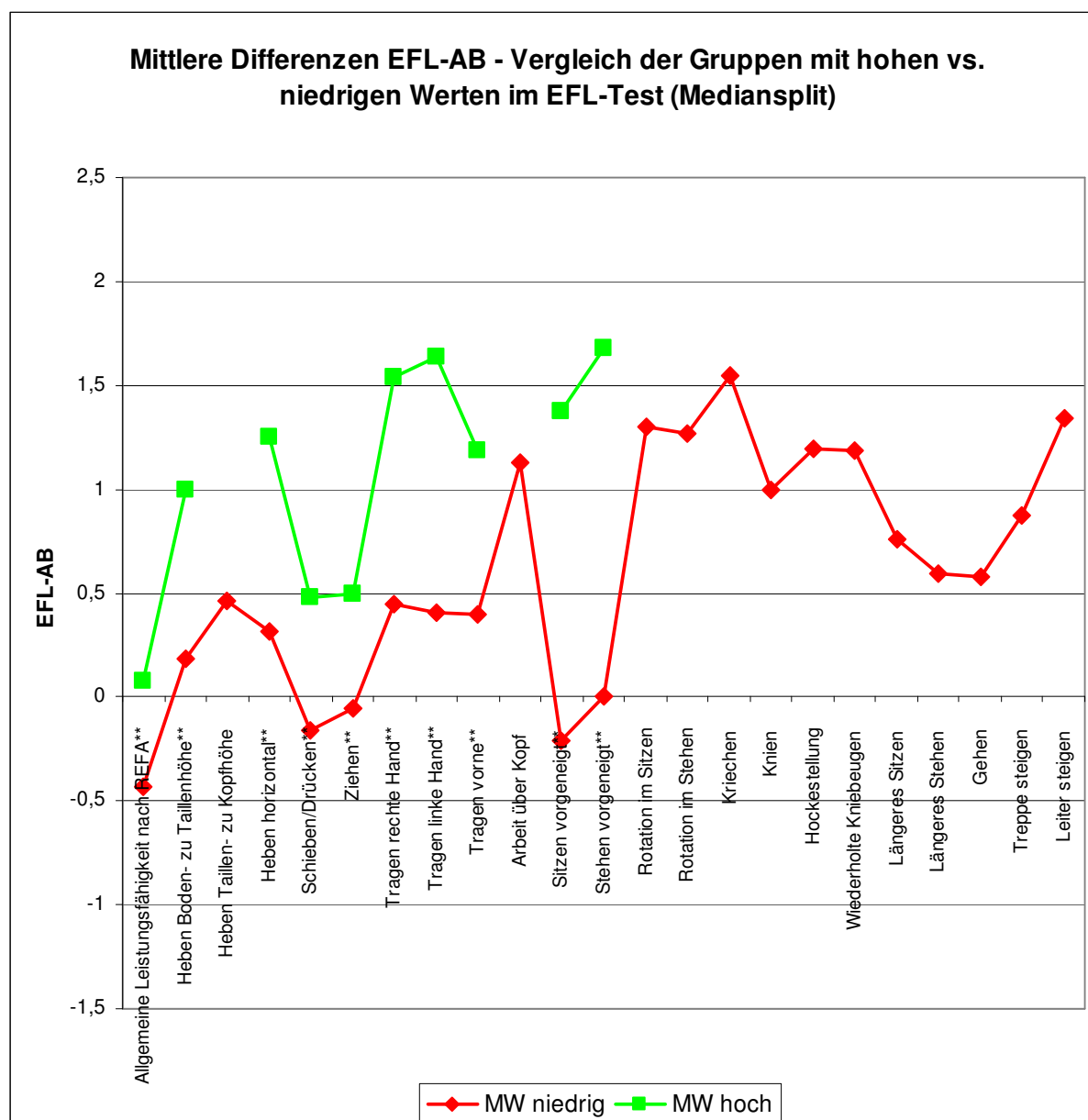


Abbildung 22: Mittlere Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis - Vergleich von Probanden mit niedrigen und hohen Leistungswerten in der EFL in den einzelnen Subtests (Mediansplit; $p < .05$; ** $p < .01$).

Bei den Einschätzungen der Ärzte, die ja gerade in diesen Tests im Schnitt erheblich niedriger liegen, tritt das Problem der Unmöglichkeit eines Mediansplits aufgrund von Decken- oder Bodeneffekten nicht auf.

Vielmehr kann die besonders große Diskrepanz zwischen EFL-Ergebnis und ärztlicher Einschätzung bei den „schwachen“ Patienten inhaltlich begründet und als Hinweis darauf gewertet werden, dass EFL und Arzt unterschiedliche Dinge

„messen“: Während die Ärzte einigen Patienten unter anderem im Hinblick auf längerfristige Prognosen eine nur geringe berufliche Leistungsfähigkeit zuschreiben und damit dem Patienten möglicherweise *restrictions* auferlegen, misst die EFL die aktuellen Leistungsgrenzen, die *limitations*, die durchaus erheblich höher liegen können (vgl. Kapitel 3.4).

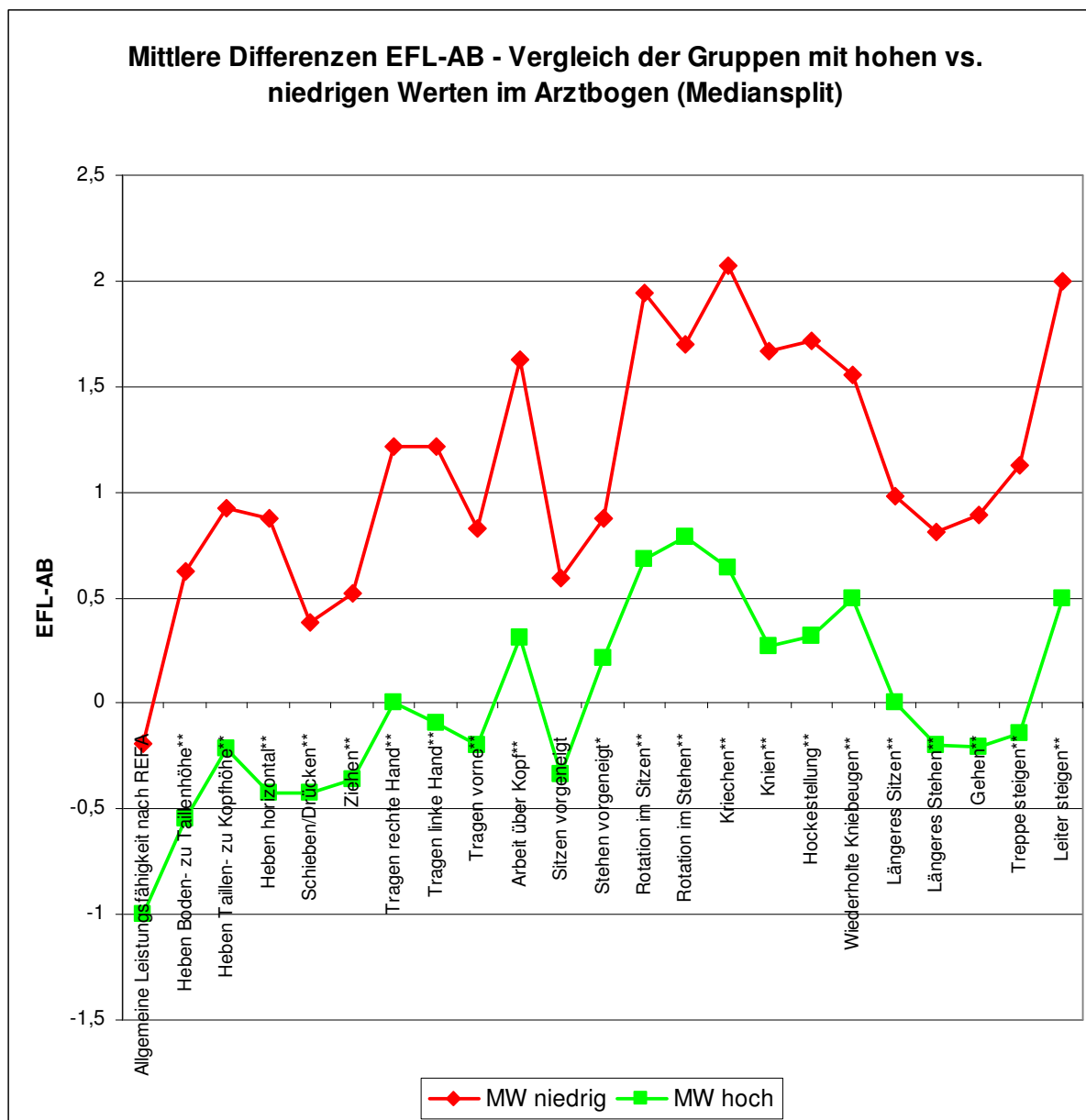


Abbildung 23: Mittlere Abweichungen der ärztlichen Einschätzung vom EFL-Ergebnis – Vergleich von Probanden mit niedrigen und hohen Leistungswerten laut ärztlicher Einschätzung in den einzelnen Subtests (Mediansplit; * $p < .05$; ** $p < .01$).

Erklärungsversuch 4: Die Ärzte lassen sich durch die Persönlichkeit des Patienten beeinflussen, die EFL nicht.

Die behandelnden Reha-Ärzte treffen normalerweise ihre sozialmedizinische Einschätzung auf der Basis von Anamnese, klinischer Untersuchung und Beobachtung des Patienten über den Zeitraum der Reha-Maßnahme. Dafür spielen neben medizinischen Daten die persönliche Einschätzung – beispielsweise die Vermutung von Aggravationstendenzen, aber sicherlich auch durchaus Faktoren wie Sympathie – eine wichtige Rolle. Wir haben die Ärzte im Bogen „Angaben zum Arzt“ gebeten, auf einer Skala von 1 (*überhaupt nicht*) bis 7 (*sehr stark*) anzugeben, in welchem Maß sie glauben, sich in der Beurteilung ihrer Leistungsfähigkeit von der Persönlichkeit des Patienten beeinflussen zu lassen. Mit Angaben zwischen 2 und 4 (MW=2,7; SD=0,71) bewegten sich die Ärzte geschlossen im unteren Bereich der Skala, der für eine eher geringe subjektiv wahrgenommene Beeinflussung steht. Diese Angaben standen in keinem Zusammenhang zu den Abweichungen der Ärzte von den EFL-Ergebnissen. Möglicherweise lässt sich dieses rein statistisch über die relativ konstanten Antworten der Ärzte bezüglich ihrer Beeinflussbarkeit erklären. Von den EFL-Therapeuten liegen zur Frage der Beeinflussung durch die Persönlichkeit des Patienten leider nur Daten von 3 Therapeuten vor; mit Angaben zwischen 2 und 3 und einem Mittelwert von 2,67 (SD=0,58) liegen diese jedoch ebenfalls genau in dem Bereich der ärztlichen Angaben.

Es gibt also in unseren – zugegebenermaßen recht schwachen – Daten keinen Hinweis darauf, dass sich die Ärzte „um den Finger wickeln“ lassen, während die EFL objektiv ist.

Erklärungsversuch 5: Die Extrapolation der Testergebnisse auf eine Dauerbelastung ist fragwürdig.

Leider kann auf diesen Kritikpunkt nur theoretisch eingegangen werden, da ohne Nachbefragung der Patienten nach einer gewissen Zeit nicht entschieden werden kann, ob sich die in der EFL ermittelten Werte der Leistungsfähigkeit auch im Alltag bestätigen oder ob die ärztliche – niedrigere – Einschätzung eher die dauerhaft realistische ist.

In der Literatur findet man jedoch eine ganze Reihe von Autoren, die die Extrapolation der Ergebnisse einer zeitlich begrenzten Einpunkt-Messung auf eine Dauerbelastung von 8 Arbeitsstunden pro Tag an jedem Tag über viele Jahre hinweg als Hauptproblem der Konzeption von FCE-Systemen ansehen. (Abdel-Moty et al., 1993; Saunders et al., 1997; Rivier & Seewer, 2002; Gatty, 2002; Brouwer et al., 2005).

Saunders et al. (1997) ermittelten empirisch optimale Extrapolationsformeln und führten für diese eine Kreuzvalidierung an einer zweiten Probandengruppe durch. Die Ergebnisse waren jedoch eher mäßig und sie kamen zu dem Schluss:

"The results of our study lead us to seriously question the general use of formulas in FCEs, unless the formulas have been shown to yield much smaller errors than those derived here. An alternative is to use direct observation of weight lifted in an extended FCE."

(Saunders et al., 1997, S. 1726)

River & Seewer (2002) fordern, dieses Problem bei der Interpretation der Ergebnisse zu beachten:

"Man muss sich auch bewusst sein, dass eines der Hauptprobleme dieser Systeme ist, dass sie die Hypothese schaffen, dass es möglich wäre, die Leistungen am Arbeitsplatz zu bestimmen, die im Verlaufe eines ganzen Tages erbracht werden, und dies gestützt auf Daten, die über einige Stunden in einem spezialisierten Zentrum gesammelt wurden."

(Rivier & Seewer, 2002, S. 39)

Eine große Gefahr liegt jedoch darin, dass diese Warnung nicht in alle FCE-Berichte explizit aufgenommen und nicht von allen Lesern entsprechend beachtet wird. Dann könnten Szenarien entstehen, wie das von Gatty beschriebene:

„It may be concluded that the client can sit „continuously for up to 1 hour.“ As such, the following return-to-work scenario is possible: the worker sits for an hour continuously with a 2-minute break, only to return to sit for another hour, and then another and another; the complaints begin after a few days followed by a sick day or two. And somewhere down the hall, a supervisor is screaming, “But according to the FCE..!” “

(Gatty, 2002, S. 17)

Exkurs: Ein Proband – zwei Ärzte

Eigentlich unbeabsichtigt haben wir für 2 Probanden von 2 unterschiedlichen Ärzten Arztbögen ausgefüllt bekommen. Stellt man einen Vergleich dieser Urteile den Abweichungen der ärztlichen Einschätzungen vom EFL-Ergebnis gegenüber, so sieht man, dass nicht nur die ärztlichen Einschätzungen vom EFL-Ergebnis abweichen, sondern es auch durchaus auch erhebliche Differenzen von bis zu zwei Kategorien zwischen den Urteilen einzelner Ärzte gibt.

(Hinweis zu den Grafiken: Die Achsen des dargestellten Rades bilden die einzelnen Subtests ab, die Positionen der Markierungen auf diesen Achsen geben die angegebene Kategorie wieder – je weiter außen die Marke liegt, umso größer ist die Leistungsfähigkeit, die dem Probanden zugeschrieben worden ist.)

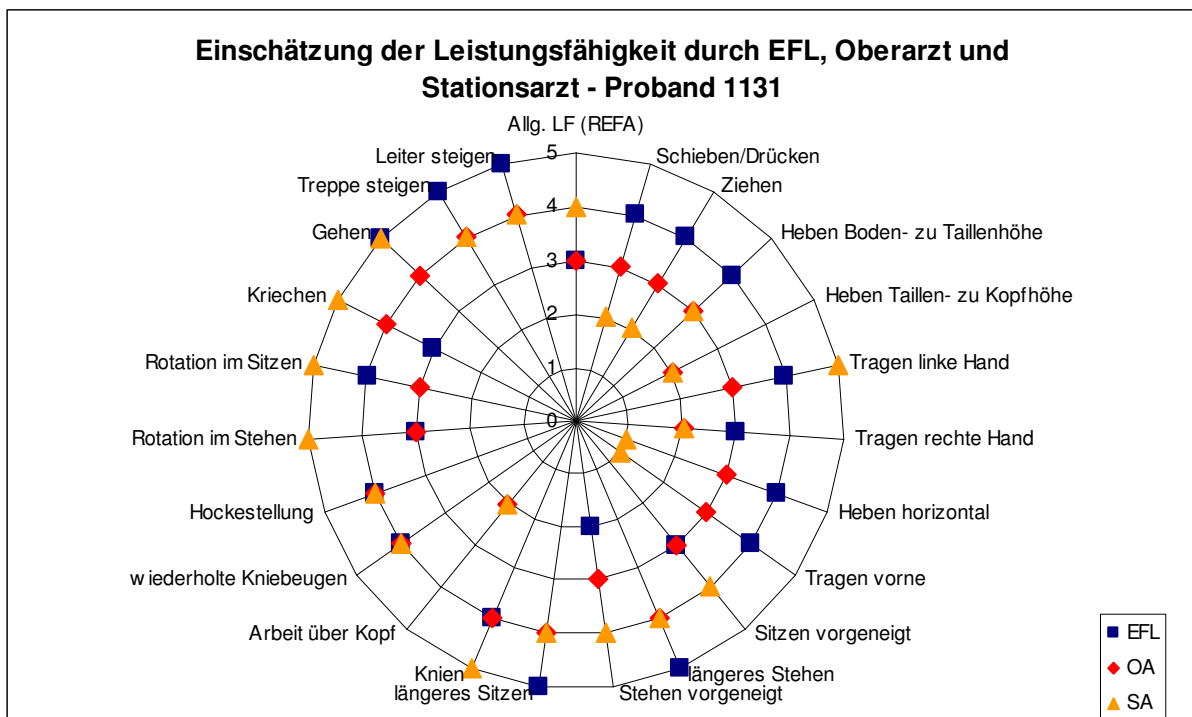
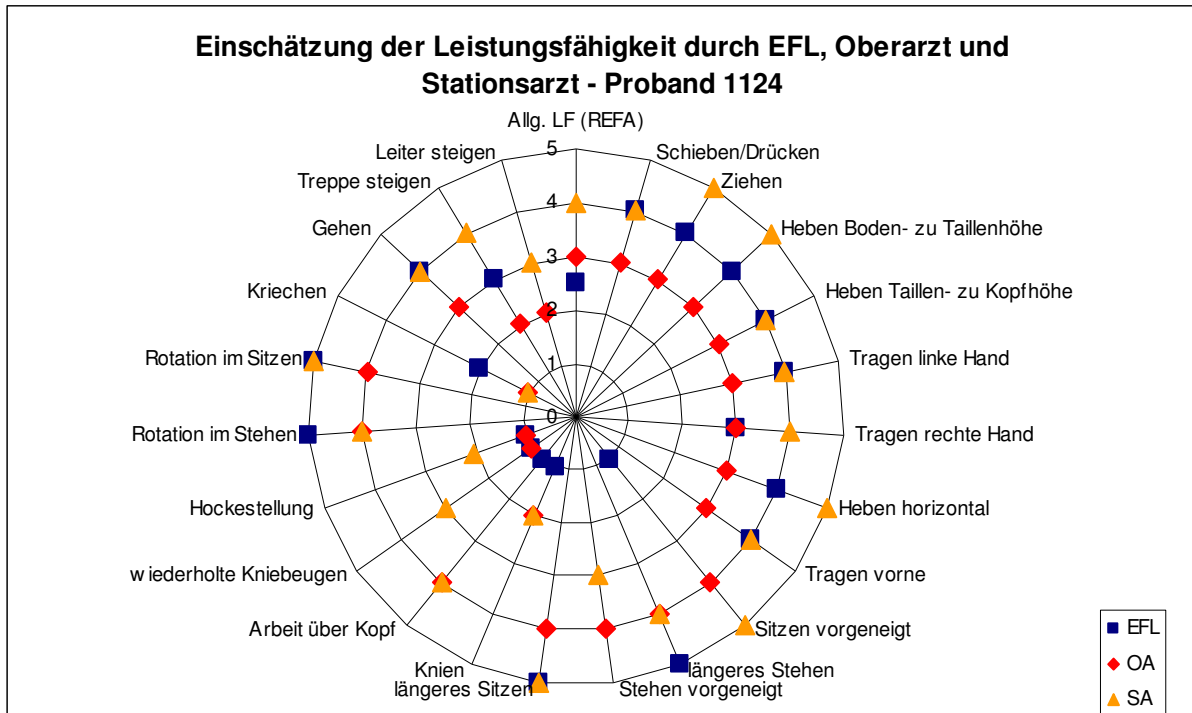


Abbildung 24: Vergleich der Einschätzungen von EFL, Ober- und Stationsarzt bei zwei Patienten

7.4 Zusammenfassung zu Frage 1

Bedenkt man, dass sowohl EFL als auch Arzt jeweils die gleichen Fähigkeiten der gleichen Probanden eingeschätzt haben, zeigen sich in unseren Daten mit Abweichungen von bis zu durchschnittlich 1,54 REFA-Kategorien unerwartet große Unterschiede zwischen beiden. Dabei trauen die Ärzte den Probanden im Schnitt eine höhere allgemeine Leistungsfähigkeit nach REFA zu, aber geringere Einzelfähigkeiten. Diese Abweichungen decken sich in etwa mit den Befunden von Brouwer et al. (2005), so dass von einer Systematik ausgegangen werden kann. Auf der Suche nach Erklärungen für die Diskrepanzen haben wir aus der Diskussion um die Validität von FCE-Systemen allgemein und der EFL im Besonderen Erklärungsversuche abgeleitet und diese anhand unserer Daten überprüft. Wir stellten fest, dass die Unterschiede zwischen den Einschätzungen besonders groß waren, wenn

- den Ärzten Arbeitsbedingungen ihrer Patienten bekannt waren, die deren Leistungsfähigkeit am Arbeitsplatz gegenüber ihrer optimalen Leistungsfähigkeit mindern könnten.
- die Ärzte mehr Berufserfahrung, aber weniger EFL-Erfahrung hatten.
- die Patienten in Bad Pyrmont ihre Reha und ihren EFL-Test absolvierten.
- die Patienten eine positive Einstellung zu ihrer Arbeit hatten.
- die Leistungsbereitschaft der Patienten in der EFL als zuverlässig eingeschätzt wurde.
- die Patienten, wenn sie zur Zeit der EFL arbeitsunfähig waren, schon eine relativ lange AU-Zeit hatten.
- die Patienten wenig Schmerz in den letzten 7 Tagen und während der EFL berichteten.
- die Patienten der EFL eher positiv begegneten.
- die Patienten die EFL als Herausforderung und Chance erlebten.
- die Patienten laut EFL ein hohes Leistungsniveau hatten.
- die Patienten laut ärztlicher Einschätzung ein niedriges Leistungsniveau hatten.

Die sozialmedizinische Erfahrung der Ärzte und einige patientenbezogene Variablen zeigten ebenfalls signifikante Zusammenhänge zu der Abweichung des ärztlichen Urteils vom EFL-Ergebnis. Diese Korrelationen waren jedoch in Bezug auf einige Subtests positiv, in Bezug auf andere negativ, so dass sie kaum als Prädiktoren für die Unterschiedlichkeit beider Urteile zu interpretieren waren.

Dazu gehörten

- die eigenen Einschätzungen der Probanden, inwieweit ihre berufliche Leistungsfähigkeit durch ihren Gesundheitszustand beeinträchtigt ist.
- ihre beruflichen Perspektiven.
- die Risikofaktoren *Trainingsmangel* und *schwere Herzerkrankung eines engen Verwandten*.
- die Lokalisation des arbeitsbezogenen Problems.

Kaum signifikante Korrelationen zu der Unterschiedlichkeit von ärztlicher Einschätzung und EFL-Ergebnis zeigten

- die Angabe der Ärzte, inwieweit sie sich durch die Persönlichkeit des Patienten in ihrem Urteil beeinflussen lassen.
- soziodemographische Variablen der Probanden: das Geschlecht, das Alter, das Bestehen einer Partnerschaft, die Anzahl der Personen im Haushalt.
- die Berufsausbildung der Probanden.
- die berufliche Situation der Probanden, beschrieben durch den beruflichen Status, den Bezug von Rente oder das Laufen eines Rentenantrags, die Dauer der Firmenzugehörigkeit, die Stellung in der Firma und das Vorhandensein eines gültigen Arbeitsvertrages.
- die konkretere berufliche Perspektive der Probanden, insbesondere ihr vermuteter beruflicher Status und ihre vermutete berufliche Situation in einem Jahr, sowie ihre Wunschtätigkeit nach der Reha.
- das Vorhandensein oder Nicht-Vorhandensein der Risikofaktoren *Übergewicht*, *eigene Herzprobleme* des Probanden mit und ohne entsprechender Medikation, *Rauchen*.
- die grobe Unterscheidung zwischen Probanden, die zurzeit des EFL-Tests arbeitsfähig und denen, die arbeitsunfähig waren.
- eine Veränderung der eigenen Einschätzung der Leistungsfähigkeit.
- die Bereitschaft, noch einmal an einem EFL-Test teilzunehmen.
- sowie eine ganze Reihe relativ global formulierter Testerkenntnisse.

Diese Variablen hatten folglich auch keinen prädiktiven Wert.

Diese Ergebnisse werten wir als Hinweise darauf, dass Unterschiede zwischen EFL-Ergebnis und ärztlicher Einschätzung zustande kommen können, weil

1. die EFL im Gegensatz zu den Ärzten keine weiteren Belastungen berücksichtigt.
2. die EFL im Gegensatz zu den Ärzten keine Entwicklungen und Prognosen berücksichtigt.
3. neben der reinen funktionellen Leistungsfähigkeit weitere patientenbezogene Variablen eine wichtige Rolle spielen.

Nicht bestätigen aber auch nicht empirisch widerlegen konnten wir den Anspruch der EFL auf stärkere Objektivität im Vergleich zum ärztlichen Urteil.

Zur Überprüfung der Kritik, dass die Extrapolation der Testergebnisse auf volle Arbeitstage kritisch sei, haben wir keine Daten sammeln können, halten diese Zweifel aber für ausgesprochen berechtigt und schließen uns der Kritik an.

7.5 Grenzen der Studie für die Beantwortung dieser Frage

Wie jede Studie hat auch die vorliegende einige Schwächen und Grenzen im Hinblick auf die Umsetzung des Forschungsauftrags:

- Ein Problem war die Abweichung der Leistungskategorien in der EFL-Tabelle von den REFA-Kategorien. Da die EFL-Kategorien jedoch die gängigen sind, haben wir uns dafür entschieden, diese auch im Arztbogen zu nutzen – und mussten dafür in der Auswertung mit einer künstlich erzeugten Mischkategorie arbeiten.
- Einige Ärzte haben angemerkt, dass es ihnen ausgesprochen schwer fällt, sich für eine der im Arztbogen angebotenen Kategorien für die Tests zu Haltung/Beweglichkeit und Fortbewegung zu entscheiden. Der Grund dafür ist eine uns unverständliche Lücke, die zwischen den in der EFL-Tabelle aufgeführten Kategorien *Zeitweise* (Ca. 10%) und *Überwiegend* (51-90%) klafft. Diese lässt die eine Zumutung der Belastung in beispielsweise etwa 1/3 des Arbeitstages nicht zu. Da wir den Arztbogen möglichst eng an die EFL-Tabelle angelehnt haben, hätten wir dieses nicht ändern können. Vielleicht wäre jedoch eine Zusatzangabe der Ärzte, wie gut ihre Angabe ihre tatsächliche Meinung widerspiegelt, interessant gewesen.
- Den größten Schwachpunkt unserer Studie in Bezug auf die Frage nach dem Stellenwert der EFL für die sozialmedizinische Leistungsbeurteilung sehen wir selbst darin, dass wir aufgrund der begrenzten Projektlaufzeit leider keine Katamnesen erheben konnten, um zu überprüfen, ob die ärztliche

Einschätzung oder das EFL-Ergebnis mittel- und langfristig die bessere Prognose darstellt. Ein Projekt, das genau darauf abzielt, ist unseres Wissens nach bisher nicht durchgeführt worden, würde aber in der Beurteilung der aktuellen Begutachtungsmethoden einen großen Schritt nach vorn bedeuten und sollte unbedingt baldmöglichst ins Auge gefasst werden.

7.6 Fazit zu Frage 1:

Vor allem bei schwer einzuschätzenden Patienten kann die EFL den Ärzten durchaus wertvolle Zusatzinformationen liefern. Diese müssen jedoch, damit sie zu langfristig validen Beurteilungen beitragen, auf die individuelle Realität der Patienten abgestimmt werden. Sie können die Einschätzung eines erfahrenen Arztes also ergänzen, aber keinesfalls ersetzen!

8. Frage 2: Trägt die EFL dazu bei, dass Patienten ein realistischeres Bild von ihrer eigenen funktionellen Leistungsfähigkeit bekommen? Wenn ja, hat das Auswirkungen auf ihre beruflichen Zielsetzungen?

Der Nutzen eines Verfahrens wie der EFL kann vielschichtig sein. Ein solches standardisiertes Assessment ist möglicherweise nicht nur für jemanden hilfreich, der die funktionelle Leistungsfähigkeit eines anderen einschätzen soll, sondern auch für den Probanden selbst. Viele Patienten, die einen EFL-Test durchlaufen, sind schon lange arbeitsunfähig und müssen zurzeit nicht die in ihrem Beruf alltäglichen Anforderungen bestehen¹⁴. Gehen weiterhin bestimmte Belastungen mit Beschwerden einher, haben viele von ihnen ihre eigenen Fähigkeiten vermutlich schon lange nicht mehr ausprobiert. Möglicherweise fällt es ihnen schwer, diese exakt einzuschätzen, weil sie Genesungs- oder Dekonditionierungsprozesse nicht angemessen berücksichtigen können. Eine unrealistische Selbsteinschätzung, möglicherweise verbunden mit einer ebenso unrealistischen beruflichen Perspektive, könnte die Folge sein.

Für diese Patienten hält ein Testverfahren wie die EFL die Möglichkeit bereit, die eigenen Fähigkeiten auszuprobieren und die eigenen Einschätzungen und Perspektiven daraufhin zu korrigieren. Doch lernen die Probanden tatsächlich etwas über und für sich selbst?

8.1 Design und Instrumente

Die Veränderung der Selbsteinschätzung und der beruflichen Perspektiven der Probanden wurde über Items aus den beiden Fragebögen zu beruflichen Perspektiven und über die beiden PACT-Tests zu Beginn und zum Ende der EFL erhoben. Realisiert wurde eine direkte Veränderungsmessung über Fragen nach konkreten Erkenntnissen durch den EFL-Test. Eine indirekte Veränderungsmessung war durch einen Vergleich der Angaben der Probanden zu der selbsteingeschätzten Leistungsfähigkeit und ihren beruflichen Perspektiven in den genannten Fragebögen im Sinne eines Prä-Post-Designs möglich.

Da diese Frage eher eine Nebenfrage des Projektes darstellte, wird an dieser Stelle auf eine umfangreiche Prädiktorenanalyse verzichtet.

¹⁴ Wie in Kapitel 6.3 dargestellt, waren 61 unserer 76 Probanden zur Zeit des EFL-Tests arbeitsunfähig, im Schnitt 244 Tage, also etwa 8 Monate.

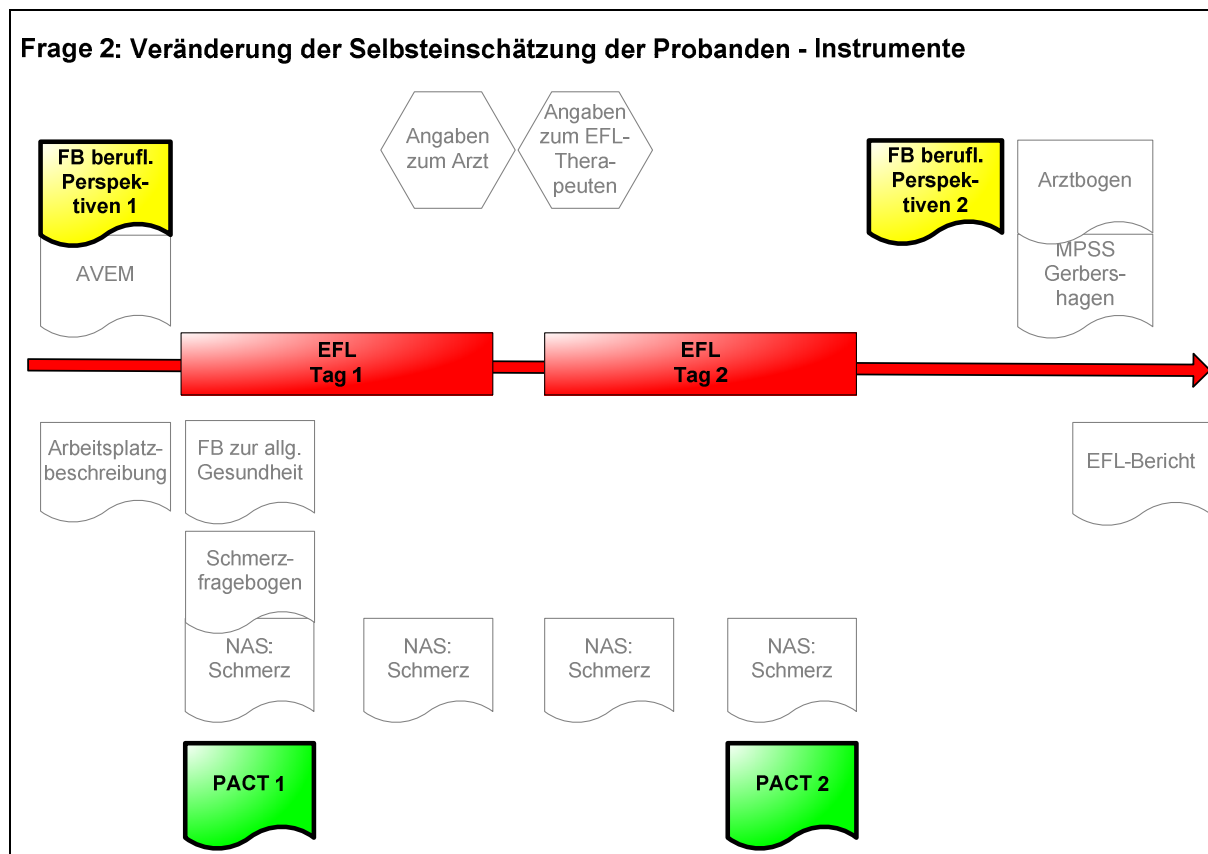


Abbildung 25: Frage 2: Veränderung der Selbsteinschätzung der Probanden – Instrumente
Die Kerninstrumente sind satt und die für diese Frage nicht verwendeten Instrumente grau dargestellt.

8.2 Ergebnisse

Direkt nach einer Veränderung ihrer Selbsteinschätzung durch die EFL gefragt, antworteten 29 Probanden (39,7%), dass sie positiv von ihrer Leistungsfähigkeit überrascht worden sind. 18 Probanden (24,7%) hatten vor der EFL vermutet, dass sie noch mehr können, 25 Probanden (34,2%) gaben an, dass sich ihre Einschätzung nicht verändert habe (Abbildung 26).

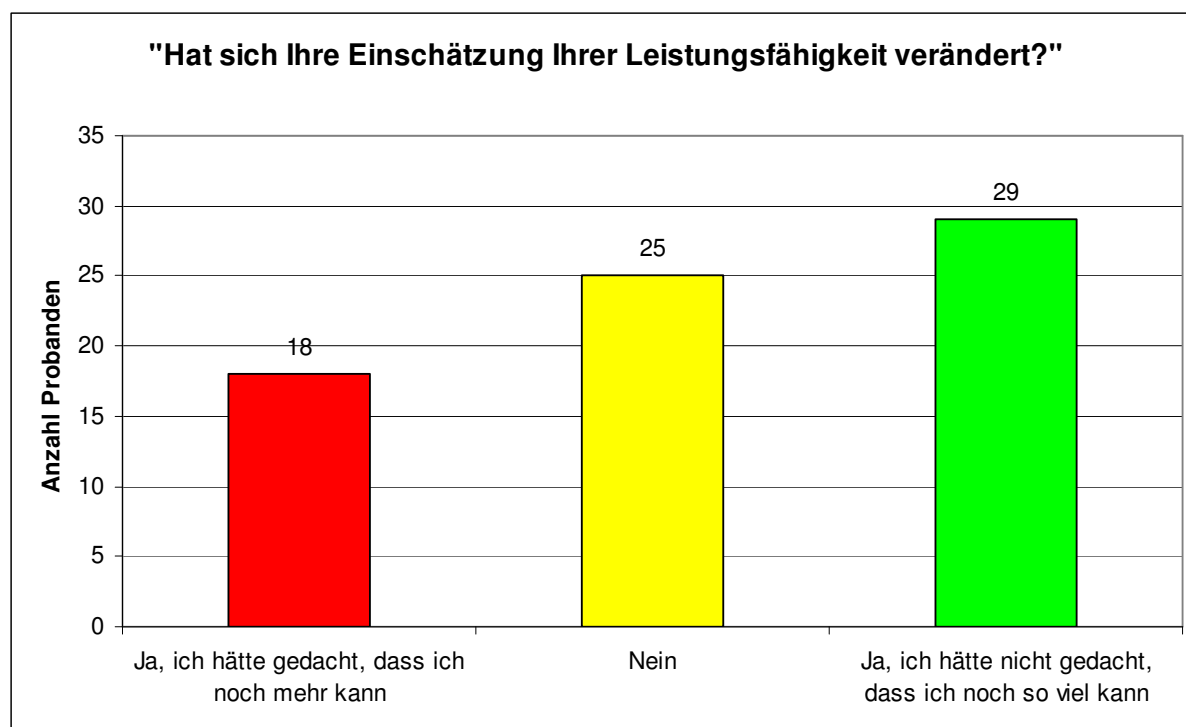


Abbildung 26: Veränderung der eigenen Einschätzung der Leistungsfähigkeit der Probanden

Die vor und nach der EFL erhobenen PACT-Werte bestätigten diesen Befund: Der Indexwert, der die gesamte Selbsteinschätzung der Patienten zusammenfasst, stieg nach der EFL von 143,7 signifikant auf 154,6 an ($T=4,268$; $p=.000^{**}$). Da jedoch komplexe Erfahrungen wie die eines EFL-Tests die Selbsteinschätzung und die persönlichen Perspektiven in verschiedene Richtungen beeinflussen können, reicht die Analyse der Veränderung eines Maßes der zentralen Tendenz nicht aus. Stattdessen muss eine differenziertere Betrachtung erfolgen, die die Entwicklungsrichtungen der einzelnen Probanden berücksichtigt. Im Falle der PACT-Werte zeigte sich, dass nach der EFL 29 Patienten etwa in dem Bereich ihres ursprünglichen Wertes (bis +/- 10 Punkte) blieben, 32 ihre Selbsteinschätzung um 11 bis 77 (!) Punkte verbesserten und 12 ihre selbst zugeschriebenen Fähigkeiten um 11 bis 38 Punkte reduzierten (Abbildung 27). Damit verdoppelte sich die Anzahl der Patienten, deren PACT-Ergebnis dem EFL-Ergebnis entsprach, so dass es im EFL-Bericht als realistisch eingestuft wurde (Abbildung 28).

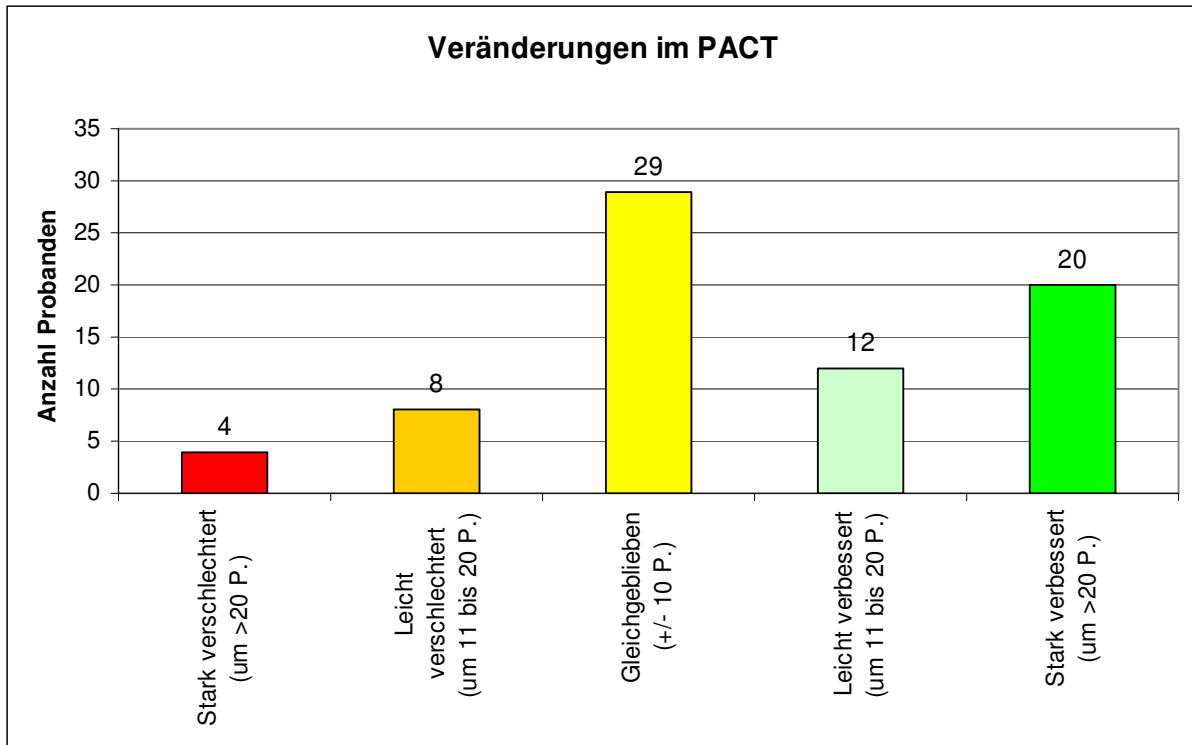


Abbildung 27: Veränderungen im PACT

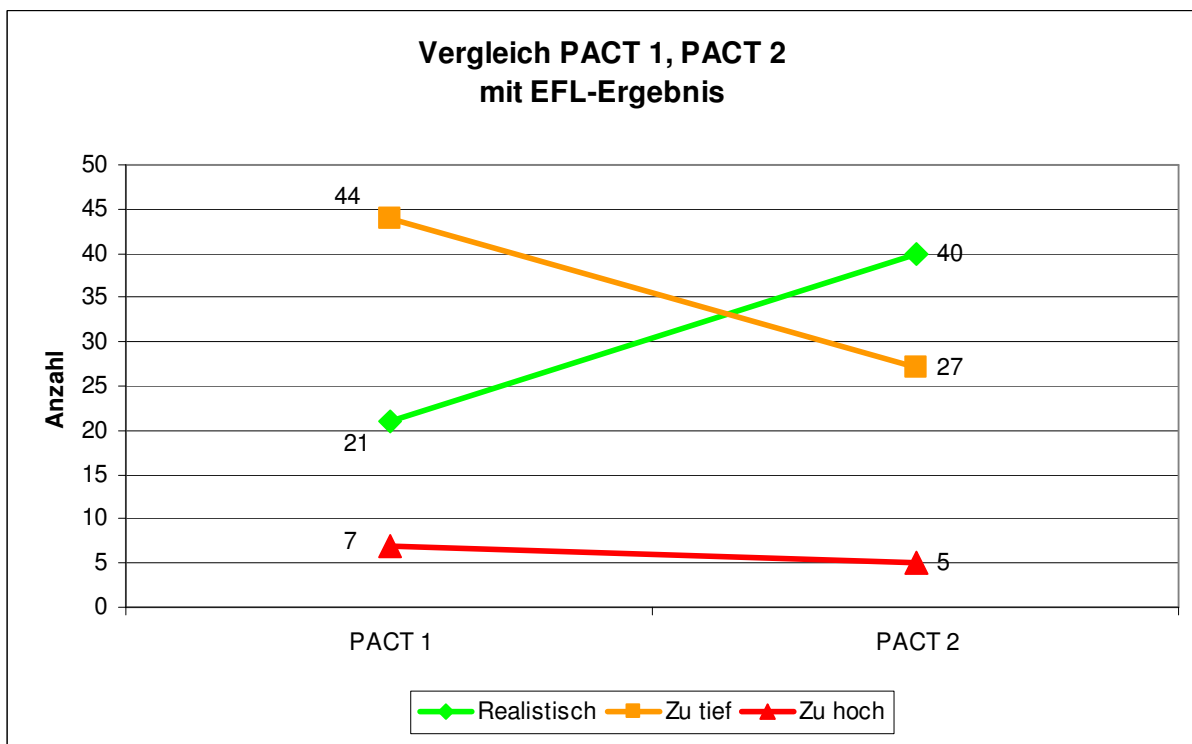


Abbildung 28: Selbsteinschätzung der Leistungsfähigkeit der Patienten im PACT im Vergleich zu ihrem EFL-Ergebnis

(Hinweis: PACT 1 wird zu Beginn des ersten, PACT 2 am Ende des zweiten Testtages ausgefüllt.)

Viele Probanden berichteten nicht nur, ihre Selbsteinschätzung verändert zu haben, sondern konnten konkrete Aspekte angeben, hinsichtlich derer sie etwas dazugelernt hatten. Erwartungsgemäß wurden an dieser Stelle am häufigsten genaue Kenntnisse der eigenen Fähigkeiten und Grenzen benannt. Erfreulicherweise relativ selten war das Fazit der Patienten, nicht mehr erwerbsfähig zu sein und eine Rente zu benötigen (Abbildung 29).

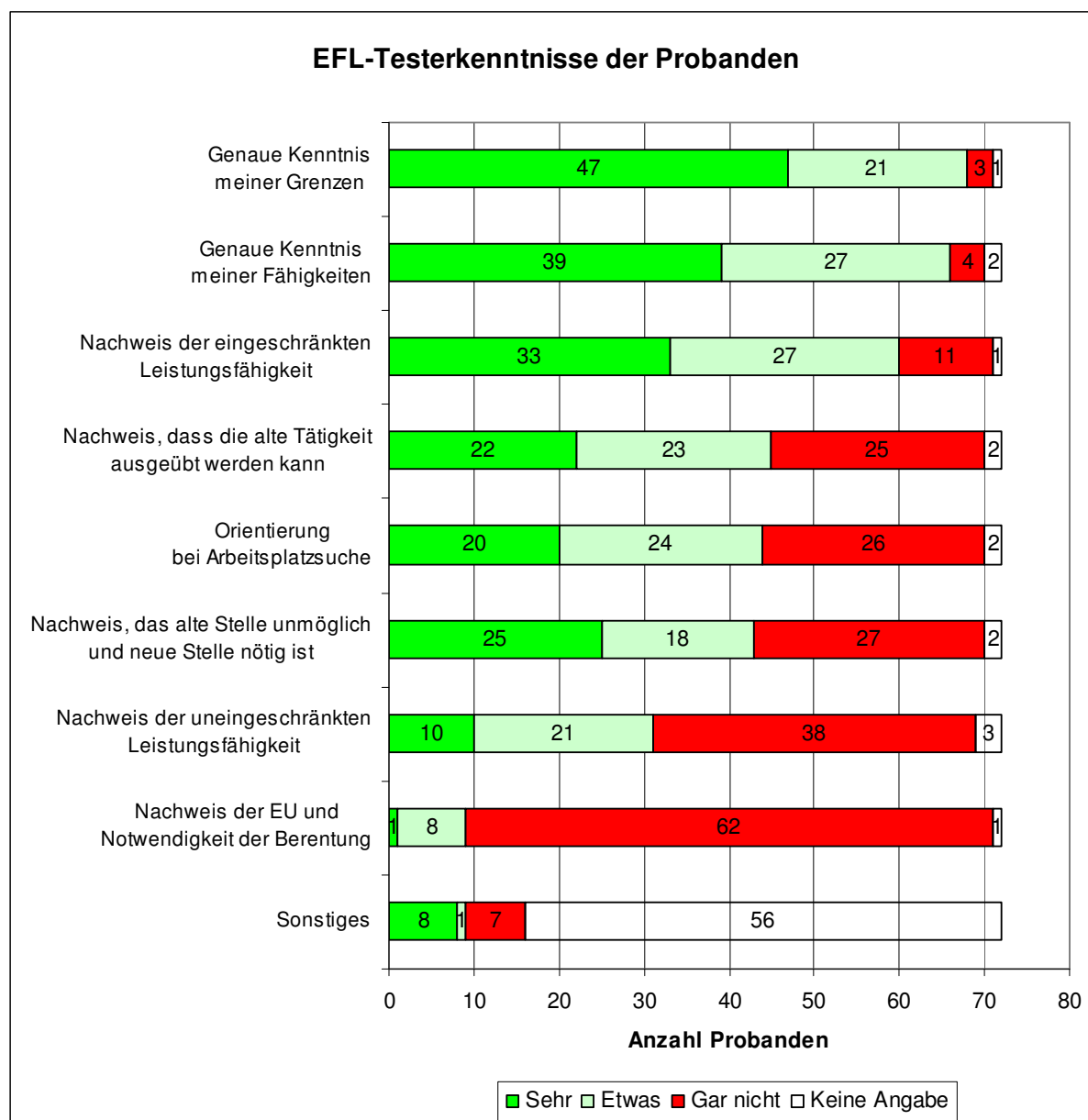


Abbildung 29: EFL-Testerkenntnisse der Probanden

In der Zusammenschau der Veränderungen in Abbildung 30 fällt auf, dass sich die starken Veränderungen in den grundsätzlichen Selbsteinschätzungen der Probanden in ihren konkreten beruflichen Plänen nicht widerspiegeln. So korrigierten 47 Probanden (64,4%) ihre Einschätzung ihrer eigenen Leistungsfähigkeit, jedoch nur 26 (35,6%) die maximale Arbeitsschwere, die sie sich ohne gesundheitliches Risiko

zutrauen, und nicht mehr als 13 (17,8%) ihre Perspektive hinsichtlich ihres Berufsstatus in einem Jahr¹⁵.

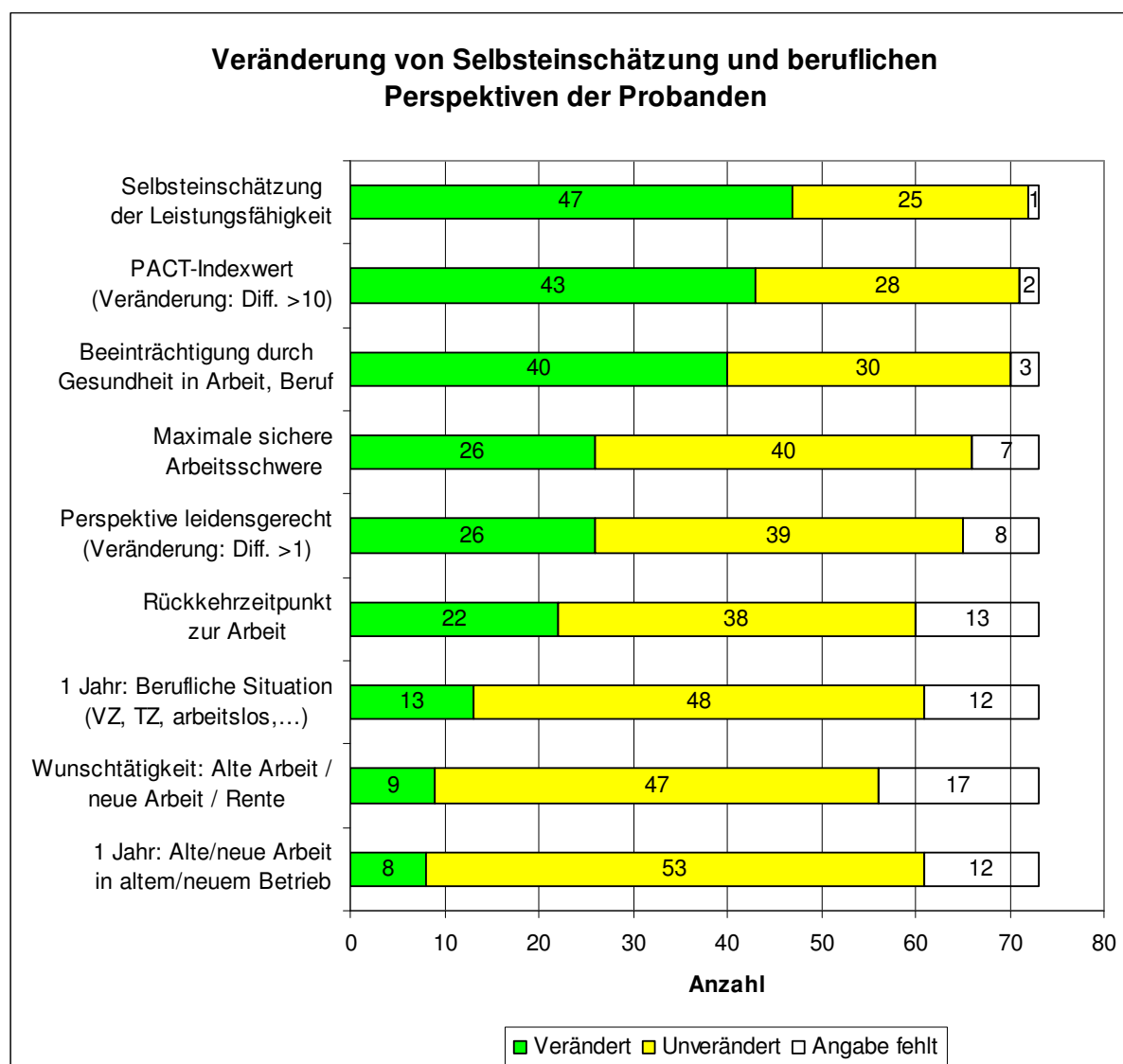


Abbildung 30: Veränderung von Selbsteinschätzung und beruflichen Perspektiven der Probanden

Als Kernaussagen eines Assessments der funktionellen Leistungsfähigkeit werden in der Regel die maximale sichere Arbeitsschwere und eine mögliche Rückkehr in das Erwerbsleben angesehen. Von den Probanden, von denen beide Angaben vorliegen, veränderte in beiden Aspekten jeweils etwa 1/3 ihre Einschätzung, etwa 2/3 blieben bei ihrem ursprünglichen Urteil (Abbildung 31).

¹⁵ Dabei wurden Veränderungen der Selbsteinschätzung in jede Richtung als Veränderung gewertet.

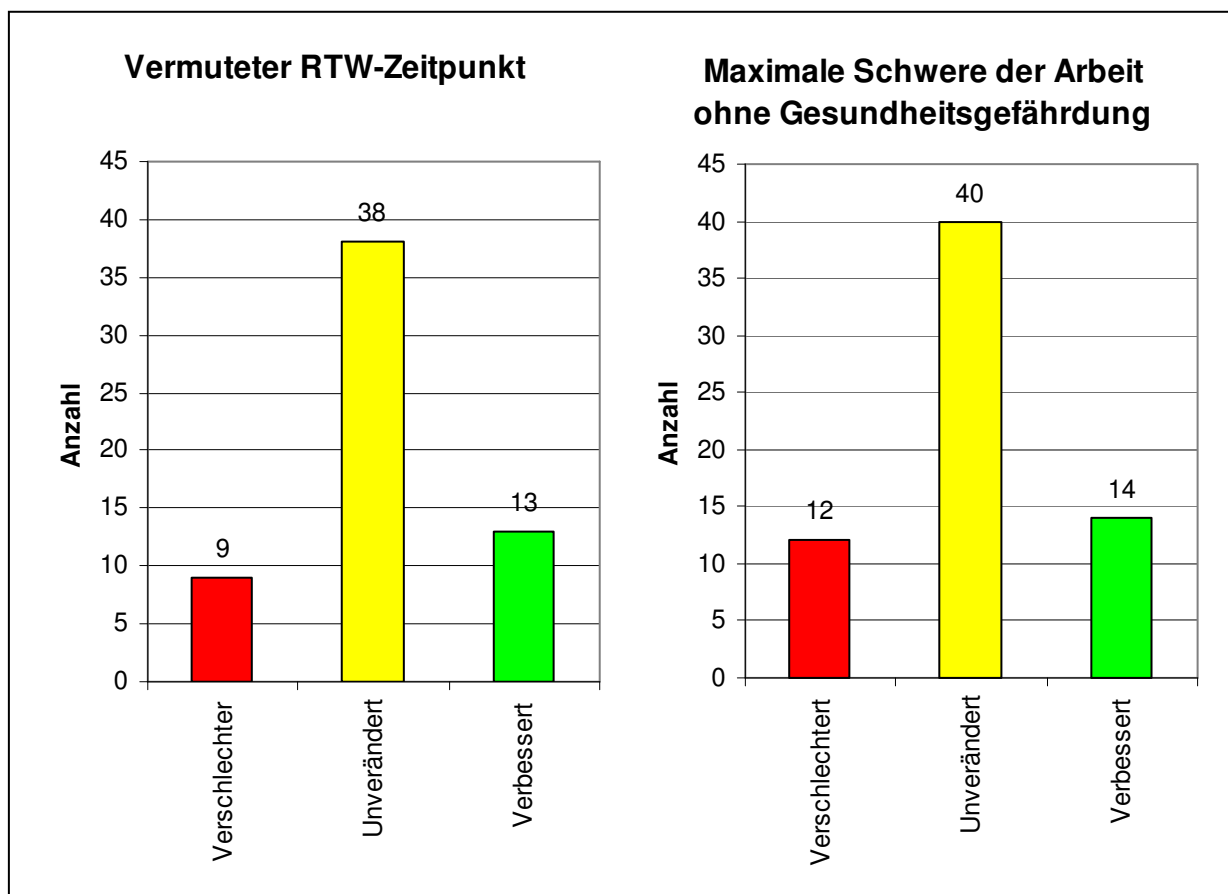


Abbildung 31: Veränderungen der Selbsteinschätzung bezüglich der maximalen Schwere der Arbeit ohne Gesundheitsgefährdung und des vermuteten Rückkehrzeitpunktes in das Arbeitsleben

Auch hier zeigt sich die Notwendigkeit einer differenzierten Betrachtung, denn die mittlere maximale Schwere der Arbeit ohne Gesundheitsgefährdung wurde von den Patienten zu beiden Zeitpunkten ungefähr gleich angegeben (U-Test nach Wilcoxon: $Z=-0,316$; $p=.752$), während die Kreuztabelle (Tabelle 18) Wechsler in beide Richtungen unterscheidet und einen signifikanten Unterschied zwischen den Messzeitpunkten anzeigt ($\text{Chi}^2=63,5$; $\text{df}=16$; $p=.000^{**}$).

Das Gleiche gilt für den vermuteten Zeitpunkt der Rückkehr in das Arbeitsleben. Auch hier konnten wir mit einem U-Test nach Wilcoxon keinen signifikanten Unterschied zwischen den Messzeitpunkten nachweisen ($Z=-0,675$; $p=.500$), obwohl die Kategorien signifikant unterschiedlich besetzt waren ($\text{Chi}^2=129,1$; $\text{df}=16$; $p=.000^{**}$; Tabelle 19). Besonders erwähnt werden sollen die Probanden, die ihrer Einschätzung nach nie wieder gesund genug werden, um wieder arbeiten zu können. Vor dem EFL-Test traf dieses auf 5, anschließend nur auf 2 Probanden zu, dieser Unterschied ist nach einem exakten Test nach Fisher jedoch nicht signifikant ($p=.161$).

		P2: Maximale Arbeitsschwere					Gesamt
		Sitzend (Heben ≤ 5 kg)	Leicht (Heben ≤ 10 kg)	Mittel (Heben ≤ 15 kg)	Schwer (Heben ≤ 40 kg)	Sehr schwer (Heben > 40 kg)	
P1: Maximale Arbeitsschwere	Sitzend (Heben ≤ 5 kg)	4	3	1	0	0	8
	Leicht (Heben ≤ 10 kg)	1	8	3	1	0	13
	Mittel (Heben ≤ 15 kg)	1	2	21	4	0	28
	Schwer (Heben ≤ 40 kg)	0	1	6	6	2	15
	Sehr schwer (Heben > 40 kg)	0	0	0	1	1	2
	Gesamt	6	14	31	12	3	66

Tabelle 18: Kreuztabelle: Maximale Arbeitsschwere ohne Gesundheitsgefährdung vor und nach dem EFL-Test (Selbsteinschätzung der Probanden)
($\chi^2=63,5$; $df=16$; $p=.000^{**}$)

		P2: Vermuteter RTW-Zeitpunkt						Gesamt	
		Sofort	< 1 Monat	< 3 Monate	< 6 Monate	< 12 Monate	> 12 Monate		Gar nicht
P1: Vermuteter RTW-Zeitpunkt	Sofort	10	2	0	0	0	0	0	12
	< 1 Monat	2	12	2	1	0	1	0	18
	< 3 Monate	1	3	9	2	0	0	0	15
	< 6 Monate	0	0	1	3	0	0	1	5
	< 12 Monate	0	0	1	0	3	0	0	4
	> 12 Monate	0	0	0	0	1	0	0	1
	Gar nicht	1	0	0	0	1	2	1	5
	Gesamt	14	17	13	6	5	3	2	60

Tabelle 19: Kreuztabelle: Vermuteter Zeitpunkt der Rückkehr ins Arbeitsleben vor und nach dem EFL-Test (Selbsteinschätzung der Probanden)
($\chi^2=129,1$; $df=16$; $p=.000^{**}$)

8.3 Diskussion

Die dargestellten Ergebnisse zeigen eindrucksvoll, dass die Probanden die EFL durchaus für einen Zugewinn an Informationen über ihr eigenes Leistungsvermögen nutzen können und, wie auch in vielen Gesprächen deutlich geworden ist, sowohl positive als auch negative Überraschungen erleben.

Auf die konkreten beruflichen Perspektiven schlagen sich die Erkenntnisse jedoch nur selten nieder. Dabei muss man bedenken, dass nicht nur die Kenntnis der eigenen Leistungsfähigkeit für die beruflichen Perspektiven eine Rolle spielt. Mindestens ebenso wichtig ist die Verfügbarkeit von Möglichkeiten. Die Bedingungen im Betrieb und auf dem allgemeinen Arbeitsmarkt machen es vielen Patienten schwer, eine andere - leistungsgerechte - Tätigkeit zu finden, und auch Maßnahmen zur Teilhabe am Arbeitsleben, vor allem Umschulungs- und Weiterbildungsangebote, werden nur begrenzt genehmigt. Die von Bürger et al. (2001) berichtete Verschlechterung der subjektiven beruflichen Perspektiven orthopädischer Reha-Patienten während der Maßnahme fand jedoch in unserer Studie keine Bestätigung, obwohl die EFL-Probanden sicherlich zu dem Klientel mit den unsichersten beruflichen Prognosen zählen.

Die dargestellten Ergebnisse lassen uns trotz des geringen Einflusses der EFL-Erfahrung auf berufliche Pläne der Probanden zu der Überzeugung kommen, dass viele Patienten durchaus von der EFL profitieren können. Positive Erfahrungen mit der eigenen Leistungsfähigkeit verbessern oft auch die Leistungsbereitschaft in Alltag und Sport. Weiterreichende Effekte auch auf der psychischen Ebene, vor allem in Bezug auf Selbstwertgefühl und Stimmung, sind ausgesprochen wahrscheinlich. Möglicherweise kann auf diese Weise sogar der von Beutel et al. (1998) vorgestellte Teufelskreis aus 1. der Beeinträchtigung der Leistungs-, Arbeits- und Beziehungsfähigkeit, 2. beruflicher Enttäuschung oder Zurückweisung, längerer Krankheit oder Arbeitslosigkeit und 3. einer Beeinträchtigung des Selbstvertrauens in die eigene Leistungsfähigkeit und der Leistungsmotivation (vgl. Kapitel 2.2) aufgebrochen werden.

Doch auch Korrekturen in die andere Richtung sind durchaus wünschenswert, wenn auch für den Probanden zunächst unangenehm. Sie können Patienten helfen, sich leistungsgerecht zu verhalten, (erneute) überlastungsbedingte Verletzungen zu vermeiden und, wenn sie es möchten und dürfen, ein gezieltes Training zu beginnen, um ihre Fähigkeiten auszubauen.

Insgesamt kann also nach unseren Ergebnissen die von Erbstößer et al. (2003) ausgesprochene Vermutung, dass der Einsatz von FCE-Systemen auch den Probanden selbst zu einer realistischeren Selbsteinschätzung verhelfen kann, für die EFL bestätigt und dieser Effekt begrüßt werden. Damit scheint ein gezielter Einsatz der EFL nicht nur als diagnostisches Mittel für die Ärzte, sondern auch als Therapeutikum für Patienten durchaus empfehlenswert – mit der Einschränkung des in Frage 3 untersuchten und diskutierten Risikos einer Zunahme von Beschwerden.

Weiterhin ungelöst bleiben zurzeit die unter 8.5 angeführten Fragen nach der Bedeutsamkeit der Lerneffekte und der daraus resultierenden Konsequenzen.

8.4 Zusammenfassung zu Frage 2

In der Analyse von Entwicklungen aufgrund komplexer – ambivalenter? – Erfahrungen können Effekte durch eine reine Betrachtung von Maßen der zentralen Tendenz im Prä-Post-Vergleich verwässert werden. Aus diesem Grund ist es unabdingbar, differenziert nicht nur die Ausmaße, sondern auch die Richtungen von Veränderungen zu berücksichtigen.

Nach einem EFL-Test korrigierten gut 40% der Probanden ihre Selbsteinschätzung ihrer allgemeinen Leistungsfähigkeit nach oben. Knapp 40% blieben bei ihrem ursprünglichen Urteil und etwa 20% hätten sich vorher mehr zugetraut. Dabei stieg der Anteil der Probanden, deren Selbsteinschätzung sich mit dem EFL-Ergebnis deckte, von 30 auf 55% an.

Nach eigener Aussage lernten die Probanden durch die EFL am meisten über ihre Fähigkeiten und Grenzen, am wenigsten sahen sie in dem Testergebnis einen Beleg für eine ernsthaft eingeschränkte berufliche Leistungsfähigkeit.

Diese Erkenntnisse hatten jedoch nur wenig Einfluss auf ihre beruflichen Perspektiven: Je konkreter die beruflichen Pläne zum Thema wurden, umso weniger Veränderungen zeigten sich bei den Probanden. Dieser Befund lässt sich leicht dadurch erklären, dass die beruflichen Perspektiven nicht nur von der subjektiven funktionellen Leistungsfähigkeit abhängen, sondern in mindestens dem gleichen Maße von den wahrgenommenen Rahmenbedingungen im Betrieb und auf dem Arbeitsmarkt, die in der Regel erheblich stabiler sind.

Als erfreulich werten wir, dass die Probanden durch die EFL wertvolle Informationen über sich gewannen, die sie zwar möglicherweise nicht für ihre Planung ihrer beruflichen Zukunft, auf jeden Fall aber für ihr Verhalten im Alltag nutzen können. Ebenfalls positiv sehen wir, dass unsere Probanden im Gegensatz zu den Rehabilitanden in der Studie von Bürger et al. (2001) durch das strukturierte Ausprobieren ihrer Fähigkeiten tendenziell eher optimistischer als pessimistischer geworden sind.

8.5 Grenzen der Studie für die Beantwortung dieser Frage

Der Beitrag der vorliegenden Studie für die Beantwortung der Frage nach dem Nutzen von EFL-Tests als Erkenntnisquelle für die Probanden liegt in erster Linie darin, ihn explizit thematisiert und im Sinne einer Pilotstudie grundsätzlich belegt zu haben. An die von uns durchgeführten Analysen können bei Bedarf noch weitere Auswertungen angeschlossen werden. Außerdem machen die positiven Ergebnisse Mut zu Studien, die diesen therapeutischen Aspekt von FCEs stärker ins Zentrum rücken als es im Rahmen unseres Projektes möglich war, und als es nach unserem Wissen bisher geschehen ist. So stellen sich beispielsweise die Fragen:

- Welche Probanden korrigieren ihre ursprüngliche Einschätzung ihrer Leistungsfähigkeit in welche Richtung? Wer lernt was?
- Inwieweit sind die Ergebnisse durch Faktoren der sozialen Erwünschtheit beeinflusst worden?
- In welchem Maß ist die „korrigierte“ Einschätzung der Probanden tatsächlich valider als die ursprüngliche? An welchem Kriterium sollte man sie messen?
- Wie lange hält die Korrektur der Einschätzung an? Durch welche weiteren Erfahrungen wird sie gestärkt oder geschwächt?
- Welche Konsequenzen haben diese Lernerfahrungen? Was machen die Probanden hinterher anders?

Werden diese und weitere Fragen in qualitativ hochwertigen Studien behandelt, besteht eine gute Chance, diesen bisher nur am Rande beachteten Aspekt von EFL-Tests stärker in das Bewusstsein der Anwender zu bringen. Daraus könnte nicht nur eine Veränderung der Verordnungspraxis resultieren, sondern auch eine verbesserte Information und Motivation der Patienten im Vorfeld erwachsen.

8.6 Fazit zu Frage 2

Die EFL führt bei vielen Patienten zu einer erwünschten Korrektur ihrer Selbsteinschätzung, was sich jedoch aufgrund der Relevanz weiterer Rahmenbedingungen kaum in den konkreten beruflichen Perspektiven niederschlägt.

9. Frage 3: Führt die EFL zu einer erheblichen Zunahme der Beschwerden?

Einige Hinweise aus der Literatur und auch Erfahrungen aus der Praxis deuten darauf hin, dass die Anforderung der EFL an die Probanden, immer wieder an ihre Leistungsgrenzen zu gehen, vor allem bei dekonditionierten Patienten zu einer vorübergehenden Zunahme der Beschwerden führen kann. Uns ist allerdings keine Veröffentlichung bekannt, die diese Nebenwirkung gezielt untersucht hätte. Auch in der Fachöffentlichkeit wird dieses Problem kaum öffentlich diskutiert und im Zweifelsfall sogar eher dementiert als ernsthaft aufgenommen.

In unserer Studie wollten wir die erhobenen Daten auch zu einem Versuch nutzen, diesen Widerspruch in den Haltungen und Aussagen lösen. Zusätzlich sollte ermittelt werden, welche Patienten besonders in Gefahr sind, eine erhebliche Zunahme ihrer Beschwerden zu erleben.

9.1 Probanden, die den EFL-Test aufgrund einer erheblichen Zunahme von Beschwerden abgebrochen haben

In der Literatur haben wir keinen expliziten Bericht über EFL-Tests gefunden, die aufgrund einer erheblichen Zunahme von Beschwerden abgebrochen werden mussten. Lediglich Gross und Battié (2002) berichten, dass von 28 Probanden 3 am zweiten Testtag nicht erschienen sind und sich 3 weitere aufgrund von Beschwerden nicht in der Lage fühlten, mit Gegenständen zu hantieren. Sie betonen, dass keiner von ihnen den ersten EFL-Tag als Ursache der Beschwerdezunahme angeführt hat.

In unserer Untersuchung waren es 3 Probanden, die nach dem ersten Testtag so starke Beschwerden entwickelten, dass der zweite Testtag nicht durchgeführt werden konnte. Sie alle führten die Zunahme der Beschwerden auf die EFL zurück. Vor einer quantitativen Analyse der verbleibenden 73 Probanden mit vollständigem EFL-Test sollen diese 3 als kurze Kasuistiken vorgestellt und ihre Verläufe bewertet werden.

Vp 0102 (Männlich, 40 Jahre)

Aufnahmebefund: Der Patient kam aufgrund langjähriger Lumbalgien mit Schmerzausstrahlung in das linke Bein in die Reha. Bekannt war ein älterer Bandscheibenvorfall L4/L5 links. Vor der EFL bestanden bereits seit etwa 2 Jahren stärkere Beschwerden ohne zwingende OP-Indikation. Lasègue und Bragard waren

unauffällig, es lagen somit keine radikulären Symptome vor. Allerdings bestand eine Haltungsinsuffizienz mit Hohlrundrücken.

EFL: Der erste Testtag verlief problemlos. Am zweiten Testtag berichtete der Patient bei Hebebelastungen Druck in der Lendenwirbelsäule und entwickelte im Verlauf der folgenden Tage stärkste Lumbalgien mit weitgehender Bewegungsunfähigkeit und einem Druckgefühl in Höhe der Oberschenkelvorderseite links. Im Rahmen einer darauf eingeleiteten weiterführenden Diagnostik mit Magnetresonanztomographie wurden ein akuter Bandscheibenvorfall L5/S1 sowie eine Radikulitis mit Schwellung der Nervenwurzeln L4 und L5 intraforaminal beschrieben.

Entlassungsbericht: Im Entlassungsbericht wurden noch einmal die bekannten chronischen belastungsabhängigen Lumbalgien bei einer Facettensymptomatik und einer Haltungsinsuffizienz als Folge eines intraforaminalen Bandscheibenvorfalles L4/L5 links 2 Jahre zuvor aufgenommen. Darüber hinaus wurde eine akute Radikulitis mit Schwellung der Nervenwurzeln L4 und L5 intraforaminal und ein Bandscheibenvorfall L5/S1 mit akuter schmerzhafter Bewegungseinschränkung dokumentiert. Der Patient wurde nach 18 Tagen vorzeitig entlassen, da akutmedizinischer Behandlungsbedarf bestand.

Vp 1114 (Männlich, 44 Jahre)

Aufnahmebefund: Der Patient litt zur Aufnahme bereits seit Jahren unter Schmerzen in der Lendenwirbelsäule mit pseudoradikulären Ausstrahlungen in die unteren Extremitäten beidseits. Ein MRT 7 Jahre vor der EFL zeigte eine deutliche Bandscheibendegeneration und osteochondrotische Höhenminderung L4/L5 mit bilateraler Bandscheiben-Protrusion. Weiterhin schilderte der Patient Beschwerden im Nacken- und Schulterbereich im Sinne einer Cervicobrachialgie mit Ausstrahlung in beide Hände. Bei einem Reha-Aufenthalt in einer anderen Einrichtung 3 Monate zuvor erlitt er eine Leistenhernie rechts. Die darauf folgende Operation war gut verlaufen, der Patient war beschwerdefrei. Der aktuelle orthopädische Befund umfasste im Bereich der Halswirbelsäule einen Druckschmerz über C3 bis C5 bei beidseits deutlicher Verspannung des M. trapezius sowie einen Klopf- und Druckschmerz über den Facettengelenken L3-L5. Radikuläre Reizungen fehlten.

EFL: Aufgrund der Leistenhernie waren keine Hebe- und Tragebelastungen möglich. Der Patient gab an, wegen starker Schmerzen seit einem Tag überhaupt nicht geschlafen zu haben, so dass der zweite Testtag nicht mehr durchgeführt wurde.

Entlassungsbericht: Der Patient wurde mit leichten lokalen Restbeschwerden im HWS- und LWS-Bereich arbeitsfähig entlassen. In der sozialmedizinischen Einschätzung wurden mittelschwere körperliche Tätigkeiten unter Wechselbelastungen ohne Zwangshaltungen der Wirbelsäule und Überkopfarbeit für

möglich gehalten. Von den Beschwerden nach der EFL wurde nichts mehr berichtet, sie hatten sich nach ein paar Tagen adäquater Behandlung wieder gegeben.

Vp 1120 (Weiblich, 38 Jahre)

Aufnahmebefund: Die Patientin kam aufgrund der Folgen eines Arbeitsunfalls, bei dem sie sich den rechten Arm, die Schulter und den Nacken verletzt hatte. Das MRT zeigte einen mediorechtslateralen Bandscheibenvorfall mit relativer Einengung des Spinalkanals ohne Myelonkompression oder Myelopathie. Eine Tangierung der Wurzel C6 war möglich. Die Patientin berichtete Schmerzen im rechten M. trapezius und im rechten Arm sowie Parästhesien aller Finger rechts und zum Teil einer Kraftminderung der rechten Hand. Darüber hinaus wurden depressive Episoden diagnostiziert. Der orthopädische Befund bestätigte eine diskret eingeschränkte Beweglichkeit. Die Reklination war endgradig schmerzhaft mit Auslösen von Brachialgien bei gleichzeitiger Rechtsrotation. Druckschmerz konnte occipital rechts und paravertebral rechts etwa in Höhe von C2 und im M. trapezius rechts ausgelöst werden.

EFL: Der erste Testtag verlief auch bei dieser Patientin problemlos. Abends traten heftige Rückenschmerzen ein, vor allem im rechten paravertebralen Teil der Brustwirbelsäule und in der Lendenwirbelsäule. Die Patientin klagte auch über nächtliche starke Schmerzen und Schlaflosigkeit, so dass der zweite Testtag nicht durchgeführt werden konnte.

Entlassungsbericht: Die Halswirbelsäule zeigte sich frei in der Funktion, es waren keine Bewegungseinschränkungen bei Rotation, Seitneigung, Re- und Inklination mehr festzustellen. Die Patientin berichtete noch gelegentliche Kribbelparästhesien in der rechten Hand. Auch die Lendenwirbelsäule war frei beweglich, paravertebral rechts bestand noch ein Druckschmerz. Die Extremitäten waren frei von Kraftminderungen, die Muskeleigenreflexe unauffällig. Sozialmedizinisch wurde eine leichte bis mittelschwere Arbeit vorrangig in Wechselhaltung ohne Überkopfarbeit und regelmäßige Rumpfzwangshaltungen vollschichtig für möglich gehalten, die Patientin jedoch arbeitsunfähig entlassen. Auch bei dieser Patientin waren die Beschwerden nach der EFL unter entsprechender Behandlung nach kurzer Zeit wieder verschwunden und spielten zur Zeit der Entlassung keine Rolle mehr.

Bewertung der medizinischen Drop-Outs

In der einzelfallbezogenen Darstellung der medizinischen Drop-Outs zeigt sich, dass durchaus auch mit einer stärkeren Zunahme von Beschwerden gerechnet werden muss, die eine Fortführung des Tests unmöglich macht. Allerdings waren in zwei von

unseren drei Fällen die Auswirkungen der Anstrengung eher kurzfristig und gingen unter einer adäquaten Behandlung relativ schnell wieder zurück. Nur in einem Fall waren die Folgen in Form einer akuten Radikulitis und einem Bandscheibenvorfall schwerer, wobei in diesem Fall die EFL als Gelegenheitsursache zu werten ist. Da die durchführenden Therapeuten – anders als beispielsweise bei ERGOS – in der EFL sehr genau auf eine ergonomische Ausführung achten, ist zu vermuten, dass der Patient auch bei anderen Anstrengungen im Alltag einen Bandscheibenvorfall hätte erleiden können. Bleibt die Frage, ob diesem Patienten ein EFL-Test hätte verordnet werden dürfen. Die Richtlinien für Indikationen und Kontraindikationen setzen einen medizinisch stabilen Zustand bei Abwesenheit einer akuten Wurzelreizsymptomatik voraus (siehe Kap. 2.3) – dieser lag bei einer Beschwerdedauer von über 2 Jahren auf etwa dem gleichen Niveau bei negativen Lasuège und Bragard-Zeichen mit Sicherheit vor, so dass an dieser Stelle korrekt verordnet worden ist.

Dennoch ist zu konstatieren, dass gerade bei Bandscheibenprotrusionen und Bandscheibenvorfällen die Indikation zum Test aufgrund der durchaus gegebenen stärkeren Belastungen, insbesondere der Press- und Druckbelastungen, nur sehr bewusst gestellt werden sollte. Hier bedarf es einer besonders eingehenden ärztlichen Untersuchung vor den Tests. Auch sollte der Patient vor dem Test immer über die Möglichkeit einer Verschlechterung von Beschwerden aufgeklärt werden. Für den Einsatz in der orthopädischen Reha empfehlen wir zusätzlich eine Durchführung der EFL etwa in der Mitte des Aufenthaltes. Auf diese Weise kann der verordnende Arzt zunächst Rückmeldungen über den Patienten in unterschiedlichen Belastungskontexten im Rahmen der Therapie sammeln und besser einschätzen, ob der Patient den Anforderungen einer EFL gewachsen sein wird. Auf der anderen Seite bleibt im Falle einer erheblichen Beschwerdezunahme noch genügend Behandlungszeit, um diesen optimal begegnen zu können. Gegebenenfalls sollten diese Patienten vorrangig verlängert werden.

Für eine Durchführung von Signifikanztests ist die Gruppe zu klein, doch ein Augenschein-Vergleich mit den 73 Probanden, die die EFL zu Ende geführt haben, lässt die Voraussetzungen der medizinischen Drop-Outs nicht erheblich schlechter erscheinen. Längere Arbeitsunfähigkeit ist ebenso wenig ein spezifisches Problem der medizinischen Drop-Outs wie der Bezug von Rente. Auch die meisten Erwartungen an die EFL liegen im vergleichbaren Rahmen, genau wie die Bereitschaft, sich anzustrengen. Auffällig sind lediglich die mäßigen Fähigkeiten der Drop-outs in Bezug auf ihre Bewegungskoordination und ihre Kenntnisse über ergonomische Arbeitstechniken, ebenso wie die pessimistische Erwartung, dass die EFL ihnen keine uneingeschränkte Leistungsfähigkeit bestätigen wird (was sich allerdings ja auch bewahrheitet hat). Allerdings hat sich wie in Kapitel 9.3 dargestellt, keine dieser Variablen als wichtiger Prädiktor für eine starke Zunahme der Beschwerden etablieren können, so dass auch die eher auffälligen Werte nicht wirklich als Erklärung herangezogen werden können.

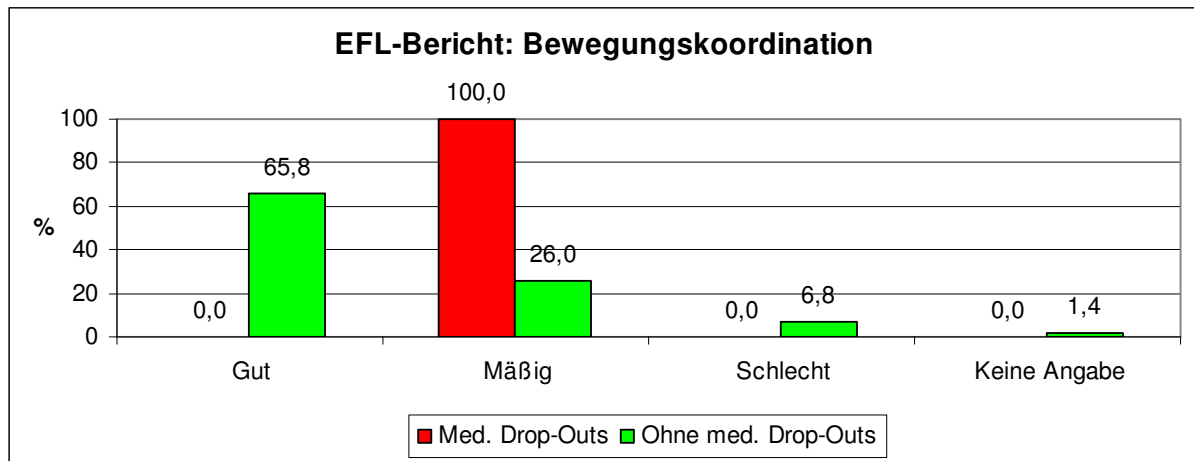


Abbildung 32: Bewegungskoordination - Vergleich der medizinischen Drop-Outs mit den Probanden, die die EFL zu Ende geführt haben

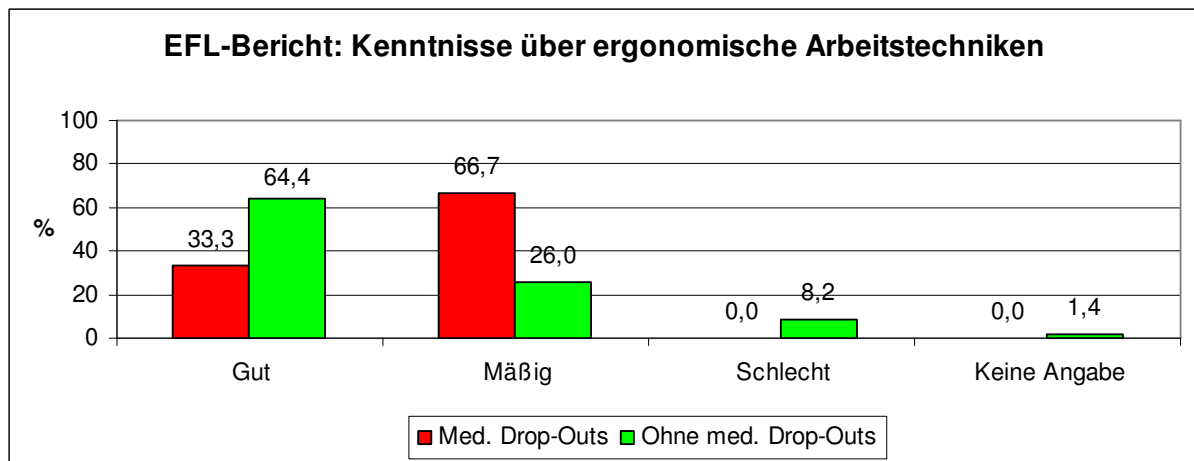


Abbildung 33: Kenntnisse über ergonomische Arbeitstechniken - Vergleich der medizinischen Drop-Outs mit den Probanden, die die EFL zu Ende geführt haben

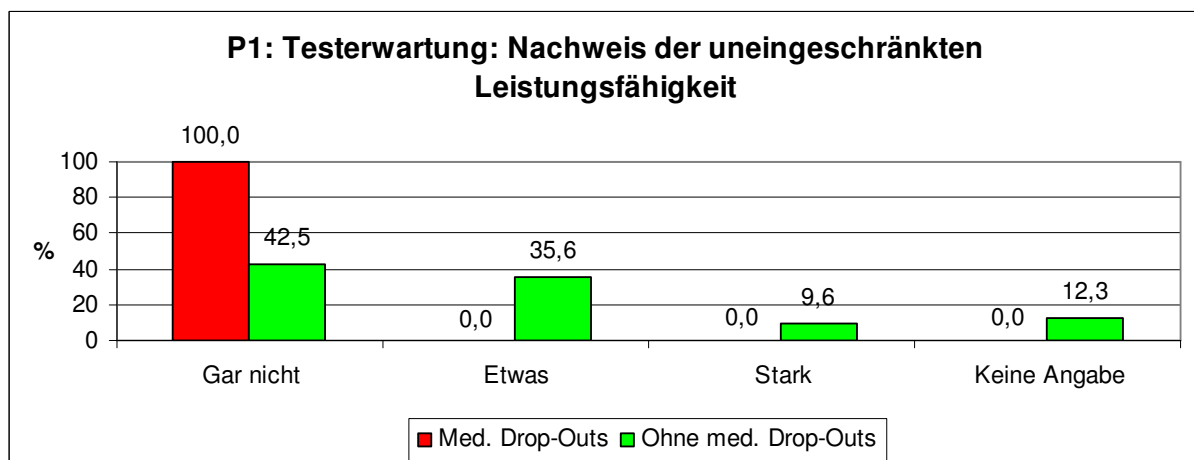


Abbildung 34: EFL-Testerwartung: Nachweis der uneingeschränkten Leistungsfähigkeit - Vergleich der medizinischen Dropouts mit den Probanden, die die EFL zu Ende geführt haben

Grundsätzlich kann man sagen, dass die permanente Anwesenheit eines ausgebildeten Therapeuten im 1:1-Kontakt während der EFL eine erhebliche zusätzliche Sicherheit für den Patienten gegenüber weniger stark beaufsichtigten Assessment-Verfahren darstellt. Dieses gilt jedoch nur, wenn die Therapeuten gut ausgebildet sind und den Probanden aufmerksam beobachten. Bei Risikopatienten sollte die Indikation scharf gestellt und darauf geachtet werden, dass die Belastung im Test in etwa der Belastung im Alltag entspricht.

9.2 Quantitative Auswertung der 73 verbleibenden Probanden – Design und Instrumente

Die Kerninstrumente für die Ermittlung der Beschwerdezunahme bei den Probanden, die den EFL-Test bis zum Ende durchgeführt haben, waren die Angaben der Probanden auf den numerischen Analogskalen, auf denen sie jeweils zu Beginn und zum Ende der beiden Testtage ihre aktuellen Schmerzen einschätzen.

Für die Prädiktorenanalyse wurden verschiedene Angaben der Probanden zu ihren Schmerzen in den letzten 7 Tagen, ihrer gesundheitlichen Verfassung allgemein, ihrer Stimmung, ihrer Einstellung zur Arbeit und ihren selbsteingeschätzten Fähigkeiten aus den Patientenfragebögen herangezogen und durch spezifische Angaben aus dem EFL-Bericht ergänzt. Eine Übersicht über die eingesetzten Instrumente gibt die Abbildung 35 .

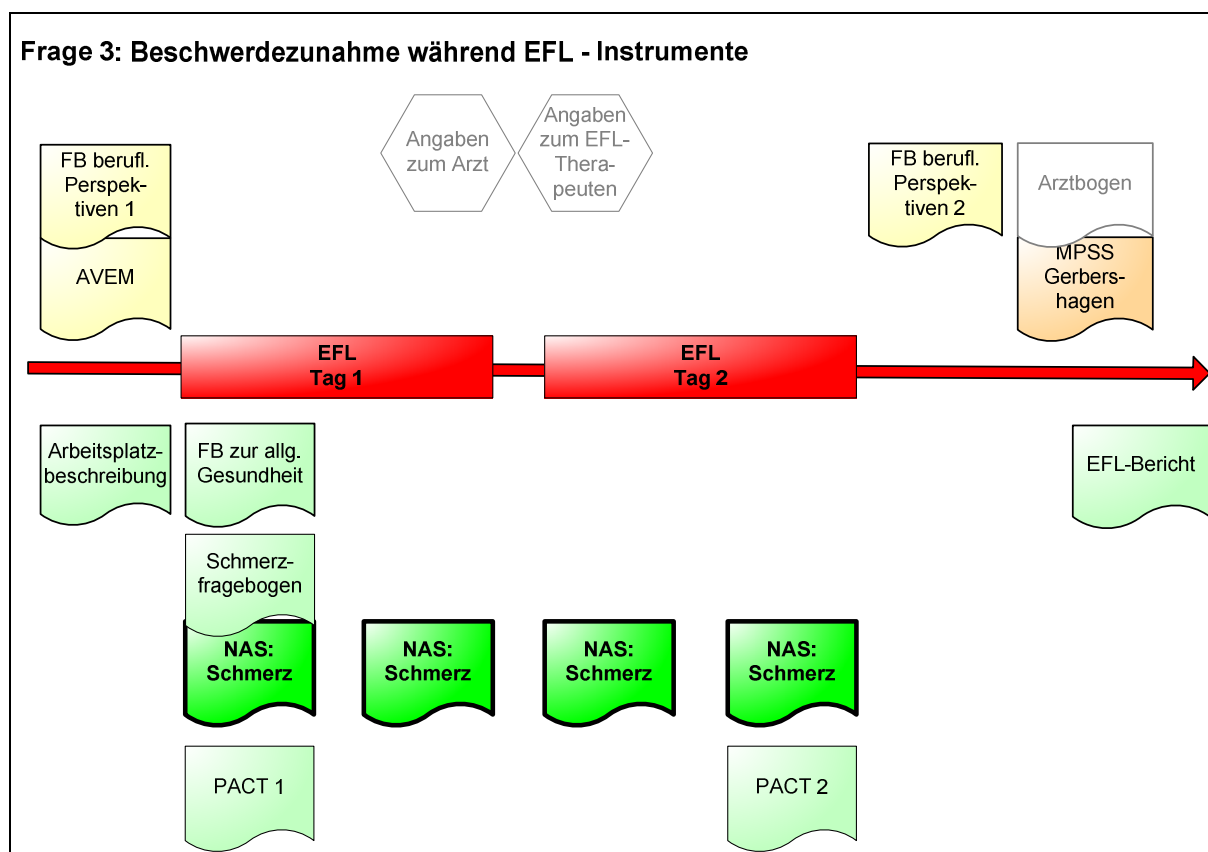


Abbildung 35: Frage 3: Beschwerdezunahme während EFL – Instrumente
 Die Kerninstrumente sind satt, die Instrumente für die Prädiktorenanalyse sind blass und die für diese Frage nicht verwendeten Instrumente grau dargestellt.

9.3 Ergebnisse

Schmerzentwicklung während der EFL

Tatsächlich ließen sich die unstandardisierten Aussagen der Patienten, die wir in zahlreichen Gesprächen auf den Fluren von ihnen gehört haben, in den Daten bestätigen: Während der EFL erlebten viele Probanden eine erhebliche Schmerzzunahme. Abbildung 36 zeigt die durchschnittliche Schmerzentwicklung während des EFL-Tests und gibt zum Vergleich den Mittelwert der von den Patienten auf die gleiche Weise berichteten schlimmsten Schmerzen in den letzten 7 Tagen wieder. Danach stiegen die Schmerzen von 3,20 (SD=1,94) auf einer numerischen Analogskala mit den Polen 0 (kein Schmerz) und 10 (unerträglicher Schmerz) signifikant auf 4,75 (SD=2,47) am Ende des ersten Testtages. Bis zum zweiten Testtag gab es eine leichte Reduktion auf 4,58 (SD=2,63) und einen erneuten Anstieg auf 5,09 (SD=2,65) – ein Wert, der dem durchschnittlichen schlimmsten Schmerz in den letzten 7 Tagen von 5,88 schon sehr nahe kommt. Post-hoc-Tests einer einfaktoriellen Varianzanalyse mit Messwiederholung zeigen, dass die

Unterschiede zwischen dem Beginn des ersten Testtages und allen weiteren Messzeitpunkten signifikant sind, während sich die Werte der letzten 3 Messungen untereinander nicht signifikant unterscheiden¹⁶.

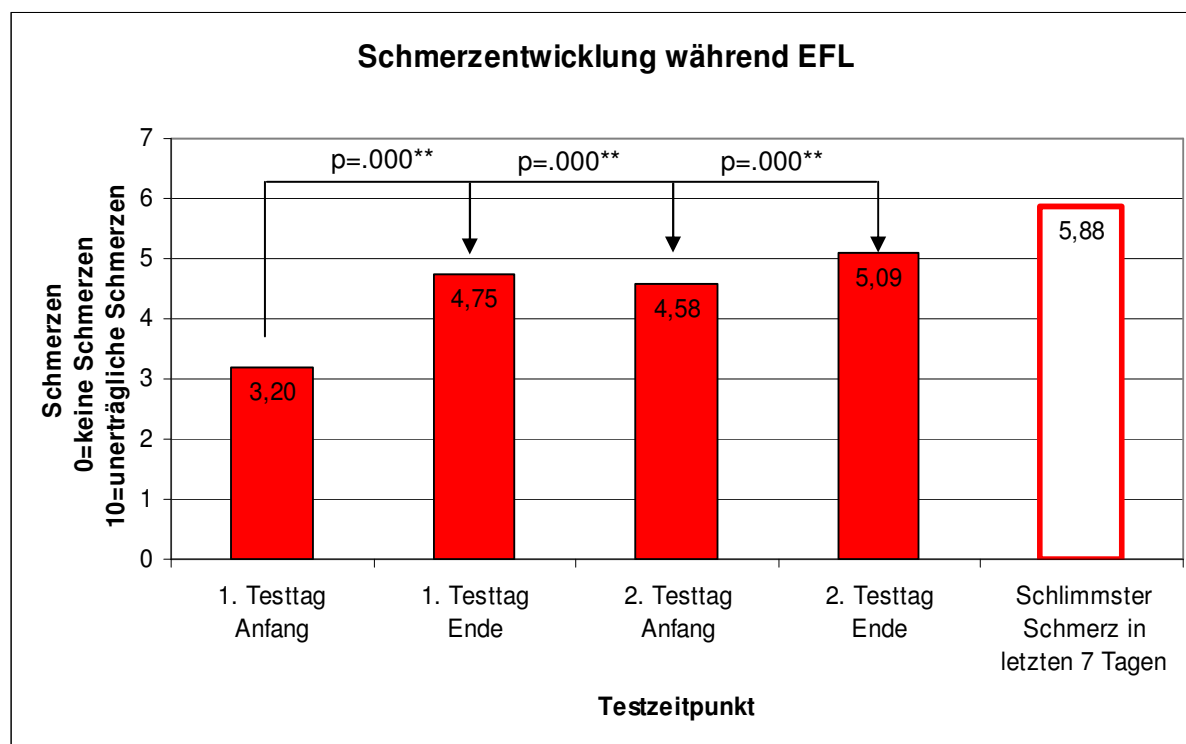


Abbildung 36: Schmerzentwicklung während EFL

Signifikante Unterschiede bestehen zwischen dem Anfang des ersten Testtages und allen weiteren Zeitpunkten (jeweils $p=.000^{**}$).

Bei einer so gravierenden mittleren Zunahme der Beschwerden ist es wünschenswert, zu wissen, ob es typische Risikopatienten gibt, bei denen man über eine Verordnung einer EFL besonders gut nachdenken sollte, oder die zumindest eine besondere Vor- und Nachbereitung benötigen.

Wir haben eine Vielfalt der erhobenen Variablen auf ihren prädiktiven Wert hinsichtlich der Schmerzzunahme überprüft und einige statistisch signifikante Prädiktoren gefunden. Um den Rahmen nicht zu sprengen, möchten wir nur diese an dieser Stelle genauer beschreiben und die exakten Angaben zu den Variablen ohne statistisch signifikanten prädiktiven Wert außen vor lassen. Je nach Datenniveau und –struktur wurden Mittelwertsvergleiche angestellt oder Korrelationen berechnet. Die angegebenen Mittelwerte stellen immer die mittlere Schmerzzunahme der entsprechenden Subgruppe dar, operationalisiert als mittlere Differenz der Schmerzstärke zum Ende des zweiten Testtages minus der Schmerzstärke zu Beginn des ersten Testtages. Die Korrelationen geben die Zusammenhänge

¹⁶ Das Signifikanzniveau wurde nach Bonferroni für einen sechsfachen Paarvergleich auf $.05/6=.008$ adjustiert.

zwischen der Ausprägung der jeweiligen Variablen und der Schmerzzunahme an, dargestellt werden in Abhängigkeit von Skalenniveau und Verteilung der Werte die Produkt-Moment-Korrelation nach Pearson (r) oder Kendalls Tau (τ).

Prädiktoren: Soziodemographische Variablen

- **Ort**

($MW_{\text{Bad Rothenfelde}}=2,402$; $MW_{\text{Bad Pyrmont}}=1,234$; $T=2,433$; $p=.017^*$)

Nicht signifikant prädiktiv waren *Geschlecht*, *Alter*, *Lebensgemeinschaft/Ehe*, die *Anzahl weiterer Personen im Haushalt*, die *Anzahl von Personen unter 18 Jahren im Haushalt*, die *Anzahl der Kinder*, die *Anzahl der unterstützungspflichtigen Kinder*, der *Schulabschluss* und die *Berufsausbildung*.

Prädiktoren: Beruf

- **Berufliche Situation:** keine signifikanten Prädiktoren

Überprüft wurden der *berufliche Status*, *psychische Belastungen am Arbeitsplatz*, der *Arbeitsumfang*, die *Dauer der Firmenzugehörigkeit*, die *Stellung im Betrieb*, das *Vorliegen eines gültigen Arbeitsvertrages*, die *Schwere der letzten Arbeit nach REFA* sowie der *Bezug von Rente* oder das *Laufen eines Rentenantrags*.

- **Psychische Aspekte der Arbeit:**

- *Meine Arbeit macht mir Freude.*

($\tau=-.252$, $p=.013^*$)

- *Ich fühle mich durch meine Arbeit überlastet.*

($\tau=.198$; $p=.050^*$)

- *Zufriedenheit mit beruflichen Perspektiven*

($\tau=-.262$; $p=.013^*$)

Nicht signifikant prädiktiv waren die Items *Bei uns ist es selbstverständlich, dass man sich unter Arbeitskollegen hilft* und *Ich kann meine Talente und Fähigkeiten voll einbringen* sowie die *Zufriedenheit mit den Beziehungen zu Arbeitskollegen und Vorgesetzten*, mit dem *Betriebsklima*, mit der *Ausstattung des Arbeitsplatzes*, mit dem *Lohn oder Gehalt*, mit der *Arbeit an sich* und mit der *Arbeitsstelle allgemein*.

- **Berufliche Perspektive (vor EFL erhoben):**

- *Vermutete berufliche Situation in einem Jahr*

(Univariate Varianzanalyse: $F=2,417$; $p=.038^*$)

Da aufgrund der Besetzung einiger Gruppen mit $n=1$ keine Post-hoc-Tests zur genaueren Lokalisation der Unterschiede möglich waren, wurde eine zweite Analyse durchgeführt, in die nur die Gruppen *voll erwerbstätig* ($n=40$), *teilzeit erwerbstätig* ($n=9$) und *Ausbildung/Umschulung* ($n=8$) einbezogen wurden. Das Ergebnis dieser zweiten univariaten Varianzanalyse bestätigt signifikante Unterschiede zwischen den Gruppen ($F=5,898$; $p=.005^{**}$ mit $MW_{\text{Voll erwerbstätig}}=1,28$; $MW_{\text{Teilzeit erwerbstätig}}=3,61$; und $MW_{\text{Ausbildung/Umschulung}}=2,63$). Zur Lokalisation der Unterschiede wurden Post-hoc-Tests nach Bonferroni durchgeführt¹⁷. Sie zeigten, dass ein signifikanter Unterschied zwischen den Probanden mit der Perspektive, in einem Jahr vollzeitlich erwerbstätig zu sein (mittlere Schmerzzunahme um 1,28 Punkte), und denen mit der Perspektive, teilzeitlich erwerbstätig zu sein (mittlere Schmerzzunahme um 3,61 Punkte), bestand ($p=.007^*$). Die Probanden, die vermuteten, in einem Jahr in Ausbildung oder Umschulung zu sein, lagen mit einer mittleren Schmerzzunahme von 2,63 Punkten dazwischen und unterschieden sich von keiner der anderen signifikant.

Nicht signifikant prädiktiv waren die *Wunschstätigkeit nach der Reha*, der *vermutete Zeitpunkt der Rückkehr ins Erwerbsleben*, die *Aussicht, in dem alten oder einem neuen Betrieb die alte oder eine neue Tätigkeit auszuführen*, und die subjektive *Wahrscheinlichkeit, dass der Tätigkeit, die in einem Jahr ausgeführt wird, ohne gesundheitliche Bedenken nachgegangen werden kann*.

Prädiktoren: Gesundheit

- **Gesundheit allgemein:**

- *Keine sonstigen Erkrankungen*

($MW_{\text{trifft nicht zu}}=2,469$; $MW_{\text{trifft zu}}=1,474$; $T=2,005$; $p=.049^*$)

- *Weitere Erkrankungen*

($MW_{\text{trifft nicht zu}}=1,583$; $MW_{\text{trifft zu}}=2,682$; $T=-2,066$; $p=.043^*$)

- *Beeinträchtigung durch Krankheit*

($T=.260$; $p=.007^{**}$)

¹⁷ Das Signifikanzniveau wurde nach Bonferroni für einen dreifachen Paarvergleich auf $.05/3=.017$ adjustiert.

Nicht signifikant waren das Vorliegen von *Erkrankungen des Herz-Kreislauf-Systems*, von *Erkrankungen der Atmungsorgane* sowie die Einschätzung von *Gesundheit allgemein*.

- **Risikofaktoren:** keine signifikanten Prädiktoren

Überprüft wurden die Risikofaktoren *Alter über 40 Jahre*, *Gewicht mehr als 5 kg über der Norm*, *BMI*, *Trainingsmangel*, *Herzprobleme*, *Herzerkrankung von engen Verwandten vor dem 60. Lebensjahr*, *Einnehmen von Medikamenten wegen Herzproblemen*, *Rauchen – mehr als 10 Zigaretten pro Tag*, *Diabetes* und *Epilepsie*.

- **Lokalisation des arbeitsbezogenen Problems:** keine signifikanten Prädiktoren

Überprüft wurden das Vorliegen von *arbeitsbezogenen Problemen in den oberen Extremitäten*, *im Rücken*, *in den unteren Extremitäten*, das Vorliegen von *Kraft- oder Ausdauerschwäche*, *psychosozialen Problemen* und die *Abwesenheit von arbeitsbezogenen Problemen*.

- **Arbeitsunfähigkeit:** keine signifikanten Prädiktoren

Überprüft wurden das Vorliegen einer *Arbeitsunfähigkeit* zur Zeit des EFL-Tests sowie die *Dauer der AU* vor dem Test.

- **Psychische Situation:**

- Stimmung: *Glaube nicht, dass es noch einmal besser wird*

($T=.235$; $p=.013^*$)

- Kraftquellen: *Hobbies, die mir Spaß machen*

($T=.193$; $p=.044^*$)

Nicht signifikant prädiktiv waren die anderen Items zu Stimmung (*Stimmung allgemein*, Komponenten schlechter Stimmung: *bin niedergeschlagen und traurig*; *kann mich zu nichts aufraffen*; *fühle mich erschöpft*; *bin unzufrieden mit meiner Situation*; *leide darunter, dass es mir wegen der Fehler anderer so schlecht geht*; *werfe mir vor, selbst Schuld zu sein*; *bin ängstlich*; *habe Angst, dass mein Körper versagen könnte*; *bin reizbar*; *ärgere mich schon über Kleinigkeiten*; *fühle mich ausgebrannt*; *fühle mich allein*), die anderen Items zu Kraftquellen (*finanzielle Freiräume*; *ein Beruf, der mir gefällt*; *besondere Fähigkeiten*; *Sonstiges*) und die Quellen verlässlicher Unterstützung (*Familie*; *Freunde*; *Bekannte*, *Nachbarn*; *Vereinskollegen*, *Clubkameraden*;

Arbeitskollegen; Sonstige). Auch die Ausprägungen in den grob formulierten Items zu Kontrollüberzeugungen hatten keinen signifikanten prädiktiven Wert.

Prädiktoren: Schmerz

- **Schmerz:**

- *Schlimmster Schmerz in den letzten 7 Tagen*
($r=.304$; $p=.009^{**}$)
- *7-Tage-Schmerz-Indexwert¹⁸*
($r=.292$; $p=.012^*$)
- *Tage mit sehr starkem Schmerz*
($T=.285$; $p=.003^{**}$)
- *Tage mit Schmerzmedikamenten*
($T=.228$; $p=.018^*$)
- *Tage mit schmerzbedingten Schlafstörungen*
($T=.247$; $p=.010^{**}$)

Nicht signifikant prädiktiv waren der *Schmerz zu Beginn des ersten Testtages* und die *Anzahl der Tage mit Schmerz in den letzten 7 Tagen*.

- **Schmerzverhalten während EFL:**

- *Entlasten der Schmerzzone*
($MW_{\text{nein}}=1,353$; $MW_{\text{ja}}=2,359$; $T=-2,085$; $p=.041^*$)
- *Verminderter Armschwung beim Gehen*
($MW_{\text{nein}}=2,144$; $MW_{\text{ja}}=0,821$; $T=2,169$; $p=.033^*$)
- *Reiben/Halten*
($MW_{\text{nein}}=1,338$; $MW_{\text{ja}}=2,458$; $T=-2,345$; $p=.022^*$)
- *Schmerzmimik*
($MW_{\text{nein}}=1,438$; $MW_{\text{ja}}=2,760$; $T=-2,654$; $p=.010^{**}$)
- *Seufzen*
($MW_{\text{nein}}=1,691$; $MW_{\text{ja}}=3,150$; $T=-2,085$; $p=.041^*$)

¹⁸ Der Indexwert wird errechnet als Summe der Anzahl Tage mit Schmerzen + Anzahl Tage mit außerordentlichen Schmerzen + Anzahl Tage mit Schmerzmedikamenten + Anzahl Tage mit schmerzbedingten Schlafstörungen und nimmt Werte zwischen 0 und maximal $4 \times 7 = 28$ Punkten an.

Nicht signifikant prädiktiv waren *Ausweichbewegungen, Muskelverspannung, Hinken, übervorsichtige Bewegungen, Starre/abnorme Haltung, sehr häufige verbale Schmerzäußerung* sowie ein *unauffälliges Schmerzverhalten*.

Prädiktoren: Leistungsbereitschaft

- **AVEM:**

- *Subjektive Bedeutsamkeit der Arbeit* (Rohwert)

($r = -.186$; $p = .041^*$)

Nicht signifikant prädiktiv waren alle anderen Skalen und (Risiko-)Muster des AVEM (*Beruflicher Ehrgeiz, Verausgabungsbereitschaft, Perfektionsstreben, Distanzierungsfähigkeit, Resignationstendenz bei Misserfolg, Offensive Problembewältigung, Innere Ruhe und Ausgeglichenheit, Erfolgserleben im Beruf, Lebenszufriedenheit, Erleben sozialer Unterstützung, Risikomuster A Überhöhtes Engagement, Risikomuster B Resignation, Muster S Schonung, Muster G Gesundheitsförderliches Verhalten gegenüber der Arbeit*) sowie die *Selbstlimitierung während des EFL-Tests infolge von Schmerz*.

- **Leistungsbereitschaft während der EFL:**

- *Zuverlässigkeit der Leistungsbereitschaft während der EFL*

(Einfaktorielle Varianzanalyse: $F = 4,044$; $p = .022^*$ mit $MW_{\text{nicht zuverlässig}} = 5,000$; $MW_{\text{fraglich}} = 2,633$; $MW_{\text{zuverlässig}} = 1,580$. Paarweise Vergleiche nach Bonferroni unter einer Alpha-Adjustierung für einen dreifachen Paarvergleich auf $.05/3 = .016$ ergaben keine signifikanten Unterschiede zwischen einzelnen Subgruppen; ein Scheffé-Test ordnete in einer Analyse homogener Subgruppen die Probanden mit fraglicher Leistungsbereitschaft sowohl denen mit zuverlässiger als auch denen mit nicht zuverlässiger Bereitschaft zu.)

- *Konsistenz: nicht signifikant prädiktiv*

- *Hebeleistungen am 2. Testtag reproduziert?*

(Einfaktorielle Varianzanalyse: $F = 3,173$; $p = .048^*$ mit $MW_{\text{gemindert}} = 3,714$; $MW_{\text{reproduziert}} = 1,647$; $MW_{\text{gesteigert}} = 1,929$. Paarweise Vergleiche nach Bonferroni unter einer Alpha-Adjustierung für einen dreifachen Paarvergleich auf $.05/3 = .016$ ergaben keine signifikanten Unterschiede zwischen einzelnen Subgruppen; ein Scheffé-Test ordnete in einer Analyse homogener Subgruppen alle Probanden einer einzigen homogenen Untergruppe zu.)

Prädiktoren: Test-Erwartungen und -Befürchtungen

- Erwartung: *Allgemeiner Nachweis, dass meine Leistungsfähigkeit eingeschränkt ist*

($\tau=.212$; $p=.042^*$)

Nicht signifikant prädiktiv waren alle anderen Erwartungen (Erwartungen, *genaue Kenntnisse über seine Fähigkeiten und Grenzen zu bekommen*; Erwartungen der Nachweise, dass die *alte Tätigkeit noch ausgeübt werden kann*; dass die *alte Tätigkeit nicht mehr ausgeübt werden kann und eine neue Stelle nötig ist*; dass *gar keine berufliche Tätigkeit mehr möglich und eine Rente nötig ist*; dass die *Leistungsfähigkeit nicht eingeschränkt ist*; Erwartung der *Orientierung bei der Arbeitsplatzsuche*) sowie alle Befürchtungen in Anbetracht des EFL-Tests (Befürchtungen des Probanden, *vom Therapeuten nicht ernst genommen zu werden*; *sich zu blamieren oder zu versagen*; Befürchtung, dass die *Beschwerden durch die Belastung schlimmer werden* oder *sich der Proband gefährlich verletzen könnte*; Befürchtungen, dass herauskommt, dass der Proband *wieder an seine alte Arbeitsstelle zurück muss*; dass er *seinen Job nicht mehr machen darf, keine Chancen auf Leistungen zur Teilhabe am Arbeitsleben oder auf eine Rente* hat).

Prädiktoren: Leistungsniveau

- *Maximale Arbeitsschwere nach REFA laut EFL-Ergebnis*

(Doppelt ausgewertet:

1. $\tau=-.340$; $p=.000^{**}$;

2. Einfaktorielle Varianzanalyse: $F=5,978$; $p=.001^{**}$ mit $MW_{\text{leicht}}=2,750$; $MW_{\text{leicht bis mittelschwer}}=2,583$; $MW_{\text{mittelschwer}}=2,729$; $MW_{\text{schwer}}=0,758$.

Paarweise Vergleiche nach Bonferroni unter einer Alpha-Adjustierung für einen sechsfachen Paarvergleich auf $.05/6=.008$ lokalisieren mit $p=.002^*$ einen signifikanten Unterschied zwischen den Probanden, die laut EFL-Ergebnis mittelschwer und denen, die schwer arbeiten können.)

Nicht signifikant prädiktiv waren die maximale Arbeitsschwere laut ärztlicher Einschätzung, die maximale Arbeitsschwere laut Einschätzung des Probanden vor der EFL sowie sein Indexwert im PACT 1 und dessen Übereinstimmung mit dem EFL-Ergebnis. Ebenso wenig Aussagekraft für die Schmerzzunahme hatten die Kenntnisse des Probanden über ergonomische Arbeitstechniken und seine Bewegungskoordination.

9.4 Diskussion

Tatsächlich lässt sich eine zumindest kurzfristige Schmerzzunahme während der EFL nachweisen, doch sind ernsthaftere Komplikationen seltene Ausnahmen, für die eine korrekt durchgeführte EFL eher eine Gelegenheitsursache darzustellen scheint. Mit einem Personal-Patienten-Schlüssel von 1:1 ist die EFL konzeptuell den eher sicheren Assessments zur Messung der Leistungsfähigkeit zuzurechnen, wenn die durchführenden Therapeuten gut ausgebildet sind und die Probanden aufmerksam beobachten. Dennoch bedarf es zu dieser Testung einer sehr bewussten Indikationsstellung und umfassenden Aufklärung des Patienten.

Viele unserer Probanden waren schon recht lange arbeitsunfähig und gaben selbst Trainingsmangel als Risikofaktor an. Bei diesen Patienten wäre nach Hart et al. (1993) eine besondere Zunahme von Schmerzen aufgrund der für dekontionierte Patienten ungewohnten Anstrengung zu erwarten gewesen. Diese Annahme scheint plausibel, ließ sich jedoch empirisch nicht bestätigen, denn weder die Dauer der Arbeitsunfähigkeit noch der von den Probanden selbst angegebene Trainingsmangel erwiesen sich als Prädiktoren für eine stärkere Schmerzzunahme. Ein Grund dafür könnte darin liegen, dass bei der EFL als kinesiotherapeutischem Verfahren alle Probanden, unabhängig von ihrem Trainingszustand, an ihre individuellen Grenzen gebracht werden. So arbeiten dekontionierte und trainierte Probanden zwar auf unterschiedlichem Niveau – belastend ist die EFL aber für beide Gruppen.

Für besonders bemerkenswert halten wir, dass die berufliche Situation und die Rahmenbedingungen der Erwerbsarbeit für die Schmerzzunahme ebenso wenig prädiktiven Wert hatten wie die meisten objektivierbaren Gesundheitsvariablen. Von entscheidender Bedeutung hingegen waren die eher „weichen“ Indikatoren, die eng an das Erleben und die persönlichen Bewertungen der Probanden gekoppelt sind. Freude an der Arbeit, eine gute berufliche Perspektive und eine hohe subjektive Bedeutsamkeit der Arbeit scheinen sich protektiv ausgewirkt zu haben, während eine negative Einstellung zur Arbeit und das Verfolgen interessanter Hobbies mit einer stärkeren Zunahme von Schmerzen einhergingen. Auch eine starke wahrgenommene Beeinträchtigung in Arbeit und Beruf durch den Gesundheitszustand sowie eine geringe Hoffnung auf Besserung waren Kennzeichen von Risikopatienten. Insgesamt scheinen also die Patienten mit kritischem beruflichen Hintergrund mit einer besonders starken Beschwerdezunahme zu reagieren. Dazu passt, dass in der Regel genau diese Gruppe in der orthopädischen Reha als eher schwierige Gruppe mit unspezifischer Schmerzbelastung und Chronifizierungstendenz bekannt ist. Tatsächlich ist der erlebte starke Schmerz in den letzten 7 Tagen ein guter Prädiktor auch für eine Schmerzzunahme während der EFL, ebenso wie ein appellatives Schmerzverhalten während des Tests, ohne dass im gleichen Maß eine Selbstlimitierung auftritt.

Vermutlich drückt sich hierin die Zwickmühle aus, in der sich diese Patienten befinden: Einerseits ist ihnen wichtig, eine gewisse Leistungsfähigkeit zu zeigen, andererseits scheinen sie zu befürchten, mit einem zu guten Ergebnis an ihren ungeliebten Arbeitsplatz zurückgeschickt zu werden.

Solche Risikopatienten sollten möglichst bereits im Vorfeld identifiziert und besonders gut über die EFL und ihre möglichen Ergebnisse informieren werden. Diese umfassen sowohl die Empfehlung der Rückkehr an den alten Arbeitsplatz wie auch einer beruflichen Veränderung, eventuell unter Nutzung von Maßnahmen zur Teilhabe am Arbeitsleben, aber auch einer Berentung. Auf diese Weise ist es vermutlich möglich, die Compliance dieser Patienten zu erhöhen und starken Aggravationstendenzen vorzubeugen. Darüber hinaus sollten für die EFL genau die gleichen Regeln gelten, die auch in der Schmerztherapie bedeutsam sind: Die von den Patienten geschilderten Beschwerden sind auf jeden Fall ernst zu nehmen, und die Probanden sollten im Falle einer erheblichen Zunahme der Beschwerden schnellstmöglich eine optimale Versorgung bekommen, auch wenn die Schmerzen körperlich nicht begründet erscheinen.

Es wäre jedoch ein Fehler, eine Schmerzzunahme einzig auf eine unbefriedigende berufliche Situation zurückzuführen, denn auch motivierte, beruflich zufriedene Patienten geben durchaus eine Zunahme von Schmerzen an, wenn diese auch geringer ausfällt. Vermutlich spielt doch die ungewohnte Belastung über einen relativ langen Zeitraum die größte Rolle in der Entwicklung der Beschwerden innerhalb der beiden Testtage. Auch der Befund, dass das Ausmaß der Selbstlimitierung die Beschwerdezunahme nicht vorhersagen kann, spricht gegen eine rein motivationale Verursachung. Vielmehr scheint die angenommene Unterscheidung von Schmerz und Funktion bei manchen Patienten besser, bei anderen schlechter zuzutreffen. So werden vermutlich gerade in Tests wie der EFL unterschiedliche Stile des Umgangs mit Schmerz sichtbar, die sich auf unterschiedliche Bedeutungszuschreibungen gründen und eine unterschiedlich starke appellative Funktion mit sich bringen.

Diskussionswürdig ist auch das Ergebnis, dass die Patienten, die laut EFL schwer arbeiten können, eine geringere Schmerzzunahme haben als die Patienten, denen nur eine Eignung für mittelschwere Arbeiten bescheinigt wird. Vermutlich haben sich diese Probanden aufgrund eines relativ geringen allgemeinen Schmerzniveaus wenig geschont. Nur so können sie ausreichend gut trainiert sein, um eine so hohe Anforderung überhaupt erfüllen zu können. Tatsächlich lässt sich nachweisen, dass eine hohe Eingruppierung nach REFA durch die EFL auch mit geringeren Schmerzen in den letzten 7 Tagen einhergeht. Reagiert aber jemand grundsätzlich auf körperliche Belastungen kaum mit Schmerz und ist dazu noch gut trainiert, hat er gute Chancen, auch die Anforderungen der EFL mit nur gering erhöhten Beschwerden zu bestehen.

9.5 Zusammenfassung zu Frage 3

Das Problem der Schmerzzunahme bei EFL-Tests ist vielen Praktikern durchaus bekannt, wird aber kaum offen diskutiert. Umso bedeutsamer erscheint uns der empirische Nachweis, dass bei einem Einsatz der EFL in der stationären orthopädischen Reha durchaus bei einer ganzen Reihe von Patienten mit einer nicht unerheblichen Schmerzzunahme gerechnet werden muss. Die mittlere aktuelle Schmerzstärke unserer Probanden stieg während des Tests von 3,20 auf 5,09 auf einer numerischen Analogskala mit den Polen 0 (keine Schmerzen) und 10 (unerträgliche Schmerzen) signifikant an. In einigen Fällen führten die Beschwerden zu einem vorzeitigen Testabbruch, ernsthafte längerfristige Komplikationen scheinen jedoch Einzelfällen vorbehalten zu sein, in denen die EFL vermutlich eher als Gelegenheitsursache zu werten ist. Typische Merkmale eines Probanden mit erhöhtem Risiko, eine erhebliche Zunahme seiner Beschwerden zu erleben, fasst die Abbildung 37 zusammen. Auffällig ist der Schwerpunkt auf dem negativen Erleben der Arbeit, verbunden mit erheblichen Schmerzen in den letzten 7 Tagen und einem appellativen Schmerzverhalten während der EFL.

Im Klinikalltag sollten gerade solche Risikopatienten intensiv auf die EFL vorbereitet werden und eine optimale Nachbetreuung erfahren. Dazu gehören neben einer differenzierten Voruntersuchung die Mitbehandlung beruflicher Probleme, der Einsatz verantwortungsbewusster, erfahrener EFL-Therapeuten, eine gute Kommunikation zwischen dem verordnenden Arzt und dem Testleiter, eine ausgiebige Information des Patienten über die EFL und ihre möglichen Aussagen sowie ein Schwerpunkt der Behandlung auf detonisierenden Maßnahmen nach der EFL.

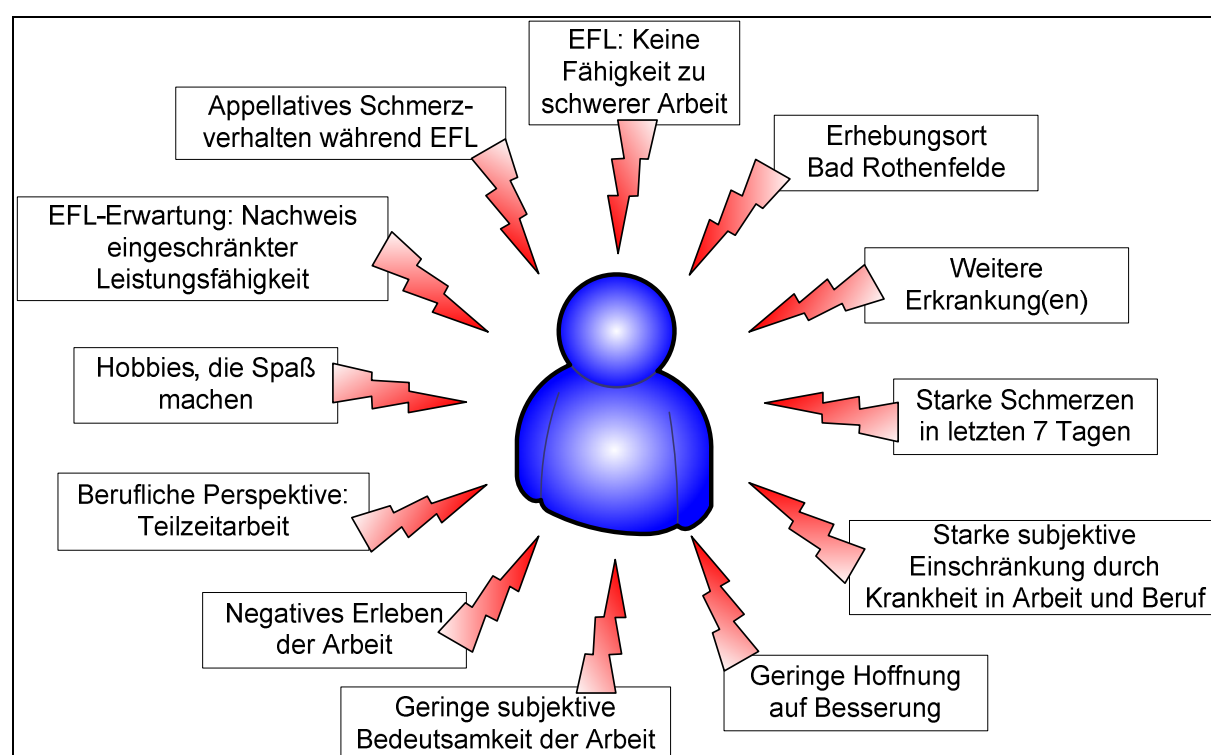


Abbildung 37: Profil eines Risikopatienten für erhebliche Schmerzzunahme während der EFL

9.6 Grenzen der Studie für die Beantwortung dieser Frage

Soweit uns bekannt ist, wurde in unserer Untersuchung zum ersten Mal die Frage der Beschwerdezunahme bei EFL-Tests explizit thematisiert und systematisch untersucht. Da sich dieses Problem jedoch erst während der Datenerhebung als relevant gezeigt hat, konnte es in die Projektplanung noch nicht differenziert einbezogen werden. Aus diesem Grund konnten wir auf der Basis des Studiendesigns leider nur mit den Standardinstrumenten eine kurzfristige Erhebung der Schmerzen nach der EFL durchführen. Interessant wäre gewesen, die Entwicklung der Beschwerden einige Tage zu verfolgen. Dann wäre nicht nur eine bessere Aussage über die durchschnittliche Dauer der Beschwerdezunahme möglich gewesen, sondern man hätte auch schauen können, inwieweit die Behandlung angepasst wurde, und welchen Erfolg diese Reaktionen von Seiten der Ärzte und Therapeuten gehabt haben.

Eine weitere Schwäche sehen wir darin, dass tatsächlich nur die Entwicklung der Schmerzstärke betrachtet wurde. Zwischen verschiedenen Schmerzarten, zum Beispiel Muskel-, Gelenk- und sonstigen Schmerzen, ist dabei nicht unterschieden worden. Auch eine Zunahme sonstiger Beschwerden wie Parästhesien, Taubheits- oder Steifheitsgefühlen in den Gelenken wurde nicht erhoben, ohne dass in diesen Bereichen ungünstige Entwicklungen ausgeschlossen werden könnten.

9.7 Fazit zu Frage 3:

Eine im Durchschnitt erhebliche Schmerzzunahme während der EFL ist empirisch nachweisbar. Diese hängt vermutlich mit der appellativen Funktion von Schmerz und Schmerzverhalten zusammen und sollte ernst genommen werden. Es scheint wichtig zu sein, gerade Risikopatienten, die befürchten, aufgrund einer zu guten EFL-Leistung wieder an einen ungeliebten Arbeitsplatz zurückgeschickt zu werden, besonders intensiv auf die EFL vorzubereiten. Darüber hinaus sollten, wie in der Schmerztherapie üblich, die angegebenen Beschwerden ernst genommen und optimal behandelt werden, damit einer weiteren Chronifizierung und Verängstigung der Patienten vorgebeugt werden kann.

10. Allgemeine Diskussion der Studie

Um die Darstellung des Projektes möglichst übersichtlich zu halten, haben wir die Ergebnisse zu den einzelnen Fragen, denen wir in der vorliegenden Studie nachgegangen sind, direkt im Anschluss diskutiert. Aus diesem Grund möchten wir uns an dieser Stelle auf eine zusammenfassende Diskussion der gesamten Studie beschränken.

Der Überblick über den bisherigen Forschungsstand zur Testgüte der EFL zeigt eine deutliche Diskrepanz zwischen den hohen Erwartungen an das Verfahren und der bisher relativ begrenzten Evidenzlage. In einem solchen Fall stellt sich die Frage nach dem Testnutzen mit besonderer Brisanz.

Strong et al. (2004) berichten, dass zwei beziehungsweise acht Monate nach einer EFL diese von etwa 50% der Patienten und von 89% der Auftraggeber als nützlich angesehen wurde. Im Unterschied zu ihnen haben wir versucht, einen Nutzen eher in Bezug auf definierte Fragestellungen als in Bezug auf persönliche Ziele zu erfassen und sind zu stimmigen Ergebnissen gekommen. Insgesamt scheint die EFL danach sowohl zur sozialmedizinischen Einschätzung durch die Ärzte als auch zur Selbsteinschätzung der Patienten einen guten Beitrag leisten zu können, wenn gleichzeitig ihre Grenzen beachtet werden. Fatalerweise haben diese den gleichen Ursprung wie auch die Stärken des Verfahrens. Dazu gehört vor allem die begrenzte Testdauer der EFL, die sie bezahlbar macht, ihren Ergebnissen jedoch den Charakter einer Momentaufnahme verleiht. Das zweite Merkmal, auf das Lob und Kritik gleichermaßen zutreffen, ist die hohe Standardisierung. Durch sie werden interindividuelle Vergleiche möglich, ihre Anwendung auf hochindividuelle Bedingungen erscheint jedoch fraglich. Eine Folge ist eine gewisse Alltagsferne, die unbedingt durch einen erfahrenen Arzt relativiert werden muss. Auf dieser Basis können sowohl Arzt als auch Patient erheblich von einer EFL profitieren, wenn die EFL gut vorbereitet und nicht nur verantwortungsvoll durchgeführt, sondern auch interpretiert wird.

Als besonders wertvoll in dieser Studie möchten wir die parallele aber unabhängige Beachtung aller drei Ebenen der Beurteilung der Leistungsfähigkeit der Probanden, nämlich des EFL-Ergebnisses, der ärztlichen Einschätzung und der Selbsteinschätzung der Patienten, nennen. Um die Unabhängigkeit zu gewährleisten, war ein erheblicher organisatorischer Aufwand nötig, der sich jedoch in interessanten empirischen Befunden ausgezahlt hat. Ebenso war die Erhebung der Daten an zwei verschiedenen Orten förderlich, da sie zum einen die Aussagekraft der Ergebnisse erhöht, zum anderen interessante Klinikvergleiche ermöglicht.

Schwachstellen des Projektes sind kaum für alle drei Fragestellungen einheitlich zu lokalisieren und sind deshalb bereits auch im Anschluss an die jeweilige Diskussion der Ergebnisse aufgezeigt worden. Hervorgehoben werden soll hier nur noch einmal das Fehlen katamnestischer Nachuntersuchungen, ohne die es nicht möglich ist, zu entscheiden, welche der drei vorliegenden Einschätzungen der Leistungsfähigkeit der Realität am nächsten kommt und damit die höchste prognostische Validität hat. Doch auch die begrenzten Kerninstrumente für die Frage nach der Zunahme von Beschwerden ist im Nachhinein bedauerlich – könnte aber in einer weiteren Studie umso besser ausgeglichen werden.

Erfreulich ist, dass sich die Beendigung der Datenerhebung bei 76 anstelle der angestrebten 82 Probanden aufgrund der überraschend großen Effekte nicht nachteilig ausgewirkt hat, so dass das Projekt erfolgreich abgeschlossen werden konnte.

11. Zusammenfassung der Studie

In den letzten Jahren sind FCE-Systeme wie die EFL nach Isernhagen als Instrumente zur Unterstützung der sozialmedizinischen Einschätzung von orthopädischen und zum Teil auch neurologischen Reha-Patienten zunehmend stärker in den Mittelpunkt des Interesses gerückt. Sie werden als objektive, reliable und valide Instrumente zur Messung der funktionellen Fähigkeiten der Probanden angesehen und zur Unterstützung entsprechender Beurteilungen eingesetzt.

Die in der Literatur geführte Diskussion über die Testgüte der EFL zeigt eine begrenzte Datenlage, die auf eher mäßige Belege für Objektivität, Reliabilität und Validität hindeutet. Diese Befunde zeigen eine gewisse Diskrepanz zu dem hohen Stellenwert, der den Testergebnissen in der Praxis oft zugestanden wird.

In unserer Untersuchung wollten wir vor allem Klarheit über den Nutzen der EFL gewinnen und haben gefragt, inwiefern die EFL

- den Ärzten für ihre sozialmedizinische Einschätzung Nutzen bringt,
- den Patienten nutzt, indem sie ihnen ermöglicht, ihre eigene Leistungsfähigkeit realistischer einzuschätzen und möglicherweise ihre beruflichen Perspektiven daran anzupassen,
- intangible Kosten mit sich bringt, indem sie zu einer erheblichen Zunahme von Beschwerden führt.

Diese Fragen sollten auf der Basis der Daten von 76 stationären orthopädischen Reha-Patienten, die in der Klinik Münsterland oder im Reha-Zentrum Bad Pyrmont einen EFL-Test durchlaufen haben, beantwortet werden.

Tatsächlich zeigten sich zwischen den verblindeten ärztlichen Einschätzungen der Leistungsfähigkeit der Patienten und den EFL-Ergebnissen überraschend große Unterschiede. Die allgemeine Leistungsfähigkeit der Patienten wurde durch die Ärzte höher, die einzelnen Fähigkeiten hingegen geringer eingeschätzt als es das EFL-Ergebnis nahelegte. Umfangreiche Prädiktorenanalysen auf der Basis der Auseinandersetzung mit der Diskussion um die Validität von EFL-Ergebnissen ließen uns schlussfolgern, dass diese Unterschiede zustande kommen, weil

- die EFL im Gegensatz zu den Ärzten keine weiteren Belastungen berücksichtigt.
- die EFL im Gegensatz zu den Ärzten keine Entwicklungen und Prognosen berücksichtigt.
- neben der reinen funktionellen Leistungsfähigkeit weitere patientenbezogene Variablen eine wichtige Rolle spielen.

Aufgrund fehlender Katamnesen nicht bestätigen aber auch nicht widerlegen konnten wir den Anspruch der EFL auf stärkere Objektivität im Vergleich zum ärztlichen Urteil.

Zur Überprüfung der Kritik, dass die Extrapolation der Testergebnisse auf volle Arbeitstage kritisch sei, haben wir keine Daten sammeln können, halten diese Zweifel aber für ausgesprochen berechtigt und schließen uns der Kritik an.

⇒ **Unser Fazit:**

Vor allem bei schwer einzuschätzenden Patienten kann die EFL den Ärzten durchaus wertvolle Zusatzinformationen liefern. Diese müssen jedoch, damit sie zu langfristig validen Beurteilungen beitragen, auf die individuelle Realität der Patienten abgestimmt werden. Sie können die Einschätzung eines erfahrenen Arztes also ergänzen, aber keinesfalls ersetzen!

Neben den Ärzten profitieren auch die Probanden von einer EFL. Das strukturierte Ausprobieren der eigenen Fähigkeiten veranlasste etwa 60% der Probanden dazu, ihre Selbsteinschätzung zu revidieren. Allerdings hatten die gewonnenen Erkenntnisse nur wenig Einfluss auf die beruflichen Perspektiven der Probanden: Je konkreter die beruflichen Pläne zum Thema wurden, umso weniger Veränderungen zeigten sich. Dieser Befund lässt sich leicht darüber erklären, dass die beruflichen Perspektiven nicht nur von der subjektiven funktionellen Leistungsfähigkeit abhängen, sondern in mindestens dem gleichen Maße von den wahrgenommenen Rahmenbedingungen im Betrieb und auf dem Arbeitsmarkt, die in der Regel erheblich stabiler sind. Trotzdem halten wir die Lernerfahrung der Probanden für sehr nützlich, da sie ihnen helfen kann, im Alltag besser einzuschätzen, was sie sich zutrauen dürfen und wo ihre Grenzen liegen. Wichtig ist jedoch dabei, dass auch den Probanden klar ist, dass die Testergebnisse eine Momentaufnahme sind und sich vieles auch wieder ändern kann.

⇒ **Unser Fazit:**

Die EFL führt bei vielen Patienten zu einer erwünschten Korrektur ihrer Selbsteinschätzung, was sich jedoch aufgrund der Relevanz weiterer Rahmenbedingungen kaum in den konkreten beruflichen Perspektiven niederschlägt.

Das Problem der Schmerzzunahme bei EFL-Tests ist vielen Praktikern durchaus bekannt, wird aber kaum offen diskutiert. In unserer Untersuchung mussten drei Patienten die EFL aufgrund einer erheblichen Zunahme von Beschwerden vorzeitig abbrechen, zwei von ihnen erholten sich unter adäquater Behandlung jedoch schnell wieder vollständig. Auch die übrigen 73 Probanden wiesen eine im Mittel signifikante Zunahme der Schmerzen auf. Typische Merkmale eines Patienten mit hohem Risiko einer starken Schmerzzunahme sind ein negatives Erleben der Arbeit, verbunden mit

starken Schmerzen in den letzten 7 Tagen vor der EFL und einem appellativen Schmerzverhalten während des Tests.

Im Klinikalltag sollten gerade solche Risikopatienten intensiv auf die EFL vorbereitet und ihnen eine optimale Nachbetreuung ermöglicht werden, damit die Compliance maximiert und das Risiko stärkerer Beschwerden minimiert werden kann.

⇒ **Unser Fazit:**

Eine im Durchschnitt erhebliche Schmerzzunahme während der EFL ist empirisch nachweisbar. Diese hängt vermutlich mit der appellativen Funktion von Schmerz und Schmerzverhalten zusammen und sollte ernst genommen werden. Es scheint wichtig zu sein, gerade Risikopatienten, die befürchten, aufgrund einer zu guten EFL-Leistung wieder an einen ungeliebten Arbeitsplatz zurückgeschickt zu werden, besonders intensiv auf die EFL vorzubereiten. Darüber hinaus sollten, wie in der Schmerztherapie üblich, die angegebenen Beschwerden ernst genommen und optimal behandelt werden, damit einer weiteren Chronifizierung und Verängstigung der Patienten vorgebeugt werden kann.

12. Bedeutung der Ergebnisse für die Praxis - Fazit

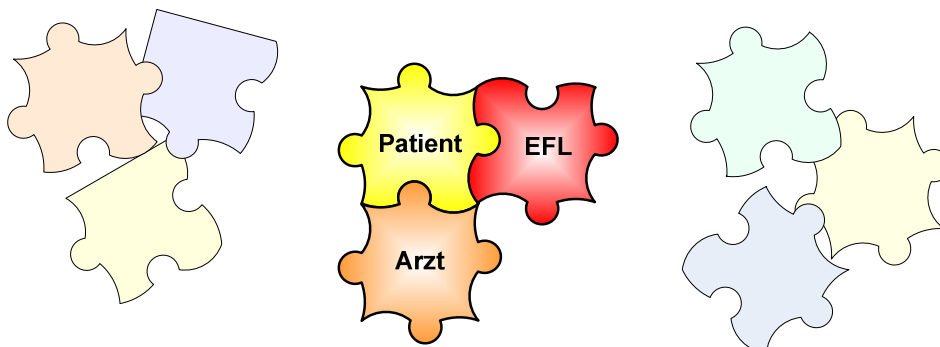
Wir sehen eine EFL für die sozialmedizinische Einschätzung als wertvoll an, wenn

- es schwierig ist, die Leistungsfähigkeit eines Patienten einzuschätzen.
- die Patienten, vor allem Risikopatienten mit appellativem Schmerzverhalten und geringer Arbeitszufriedenheit sowie resignativen Tendenzen, gut auf die EFL vorbereitet werden,
- sich Auftraggeber und Anwender der inhaltlichen Grenzen des Verfahrens bewusst sind (zusammengefasst ist die EFL ist eher eine **Functional Performance Evaluation** als eine **Functional Capacity Evaluation**) und
- die EFL in der sozialmedizinischen Beurteilung als **ein** Baustein unter mehreren angesehen wird.

Darüber hinaus kann sie den Patienten wertvolle Informationen für die Einschätzung der eigenen Leistungsfähigkeit bieten, wenn

- diese sich der Möglichkeiten und Grenzen des Verfahrens, vor allem seines Schnappschusscharakters, bewusst sind,
- sie sich aber trotzdem darauf einlassen und sich trauen, ihre Fähigkeiten wirklich auszuprobieren,
- und sie sicher sein können, dass eine zeitweilige Zunahme von Beschwerden aufgrund der starken Belastung professionell gehandhabt wird.

Die Verantwortung dafür, dass diese Voraussetzungen gegeben sind, liegt in erster Linie bei den Ärzten, die EFLs verordnen, über sie informieren, die Ergebnisse interpretieren und nutzen. Wenn sie diese Verantwortung ernst nehmen und einlösen, kann die EFL trotz ihrer methodischen und inhaltlichen Grenzen für alle Beteiligten ein wertvolles Teil im Puzzle der orthopädischen Rehabilitation darstellen.



13. Literatur

- Abdel-Moty, E., Fishbain, D.A., Khalil, T.M., Sadek, S., Cutler, R., Rosomoff, R.S. & Rosomoff, H.L. (1993). Functional Capacity and Residual Functional Capacity and Their Utility in Measuring Work Capacity. *The Clinical Journal of Pain*, 9, S. 168-173.
- Beutel, M., Kayser, E., Vorndran, A., Farley, A. & Bleichner, F. (1998). Die integrierte Belastungserprobung in der medizinischen Rehabilitation – Erfahrungen und Perspektiven am Beispiel der psychosomatischen Rehabilitation. *Rehabilitation*, 37, S. 85-92.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2. Aufl.). Springer, Berlin.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Springer, Heidelberg.
- Brouwer, S., Reneman, M.F., Dijkstra, P.U., Groothoff, J.W., Schellekens, J.M. & Goeken, L.N. (2003). Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. *Journal of Occupational Rehabilitation*, 13, 4, S. 2207-2218.
- Brouwer, S., Dijkstra, P.U., Stewart, R.E., Göeken, L.N.H., Groothoff, J.W. & Geertzen, J.H.B. (2005). Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disability and Rehabilitation*, 27, 17, S. 999-1005.
- Bürger, W., Dietsche, M., Morfeld, M. & Koch, U. (2001). Multiperspektivische Einschätzungen zur Wahrscheinlichkeit der Wiedereingliederung von Patienten ins Erwerbsleben nach orthopädischer Rehabilitation – Ergebnisse und prognostische Relevanz. *Rehabilitation*, 40, S. 217-225.
- Deutsche Rentenversicherung (2006). Website: http://www.deutsche-rentenversicherung.de/nn_15758/SharedDocs/de/Navigation/Rehabilitation_node.html_nnn=true (Stand: 21.06.2006; Zugriff: 12.11.2007)
- Deutsches Institut für Medizinische Dokumentation und Information, DIMDI (2005). Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) (Stand Oktober 2005). http://www.dimdi.de/dynamic/de/klassi/downloadcenter/icf/endaussung/icf_endfassung-2005-10-01.pdf (Zugriff: 12.11.2007)
- Die Kölnische Rück (1999). Assessmentsysteme für die medizinische Leistungsbeurteilung/Begutachtung. PS aktuell 1/99, S. 5-7.

- EFL-Akademie (2001). Objektive Beurteilungskriterien am Beispiel Heben
Tailenhöhe-Kopfhöhe. Auszug aus Kursunterlagen; Ausdruck vom
03.04.2001.
- EFL-Akademie (2007). Website: http://www.efl-akademie.de/efl_akademie.htm
(Stand März 2007; Zugriff: 23.10.2007).
- Erbstößer, S., Nellesen, G. & Schuntermann, M. (2003). Abschlussbericht der FCE-
Studie: FCE-Systeme zur Beurteilung der arbeitsbezogenen
Leistungsfähigkeit – Bestandsaufnahme und Experteneinschätzung. Website:
[http://www.deutsche-
rentenversicherung.de/nn_7112/SharedDocs/de/Inhalt/04_Formulare Publi-
kationen/03_publicationen/Publicationen/Fachzeitschrift_DRV/DRV_Schrift-
en/44_FCE_Studie.html](http://www.deutsche-
rentenversicherung.de/nn_7112/SharedDocs/de/Inhalt/04_Formulare_Publi-
kationen/03_publicationen/Publicationen/Fachzeitschrift_DRV/DRV_Schrift-
en/44_FCE_Studie.html) (Zugriff: 14.09.2004).
- Fishbain, D.A. (1999). Validity of the dictionary of occupational titles residual
functional capacity battery. *Clinical Journal of Pain*, 15, 2, S. 102-110.
- Gatty, C.M. (2002). FCEs: to use or not to use? *Rehabilitation Management*, 15, 8, S.
16-17.
- Gerbershagen, H.U., Korb, J., Nagel, B. & Nilges, P. (keine Jahresangabe). Das
Mainzer Stadiensystem der Schmerzchronifizierung. Mainz Pain Staging
System (MPSS). Homepage des DRK-Schmerzzentrums Mainz.
http://www.schmerz-zentrum.de/downloads/pdf/mpss_deu.pdf (Zugriff:
17.07.2004)
- Greitemann, B. (2004). Sozialmedizinische Leistungsbeurteilung unter
Berücksichtigung von Verfahren zur Erfassung funktioneller
Leistungsfähigkeiten und zu Profilvergleichsverfahren. Vortrag auf der Tagung
des Instituts für Qualität in Prävention und Rehabilitation (iqpr) „Dialog
zwischen Richtern und Sachverständigen zum Thema: Qualitätsstandards der
Begutachtung – Möglichkeiten und Nutzen des Einsatzes von
Profilvergleichsverfahren“ im Dezember 2004 in Kassel
- Gross, D.P. & Battié, M.C. (2002). Reliability of Safe Maximum Lifting Determinations
of a Functional Capacity Evaluation. *Physical Therapy*, 82, 4, S. 364-371.
- Gross, D.P. & Battie, M.C. (2003). Construct validity of a kinesiophysical functional
capacity evaluation administered within a worker's compensation environment.
Journal of Occupational Rehabilitation, 13, 4, S. 287-295.
- Gross, D.P. & Battié, M.C. (2004). The Prognostic Value of Functional Capacity
Evaluation in Patients With Chronic Low Back Pain. Part 2: Sustained
Recovery. *Spine*, 29, 8, S. 920-924.
- Gross, D.P., Battié, M.C. & Cassidy, J.D. (2004). The Prognostic Value of Functional
Capacity Evaluation in Patients With Chronic Low Back Pain. Part 1: Timely
Return to Work. *Spine*, 29, 8, S. 914-919.

- Gross, D.P. & Battié, M.C. (2005). Factors Influencing Results of Functional Capacity Evaluations in Workers' Compensation Claimants With Low Back Pain. *Physical Therapy*, 85, 4, S. 315-322.
- Hart, D.L., Isernhagen, S.J. & Matheson, L.N. (1993). Guidelines for Functional Capacity Evaluation of People With Medical Conditions. *Journal of Orthopaedic and Sports Physical Therapy*, 18, 6, S. 682-686.
- Innes, E. & Straker, L. (1999a). Reliability of work-related assessments. *Work*, 13, S. 107-124.
- Innes, E. & Straker, L. (1999b). Validity of work-related assessments. *Work*, 13, S. 125-152.
- Interessengemeinschaft Physiotherapie Rehabilitation (IGPTR) (keine Jahresangabe). Evaluation der funktionellen Leistungsfähigkeit (EFL) nach Isernhagen. Quelle: http://www.igptr.ch/pdf/ass_ba_efl.pdf (Zugriff: 30.08.2005).
- Isernhagen, S.J., Hart, D.L. & Matheson, L.M. (1999). Reliability of independent observer judgement of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *Work*, 12, S. 145-150.
- Kaiser, H., Kersting, M., Schian, H.-M., Jacobs, A. & Kasprowski, D. (2000). Der Stellenwert des EFL-Verfahrens nach Susan Isernhagen in der medizinischen und beruflichen Rehabilitation. *Rehabilitation*, 39, S. 297-306.
- King, P.M., Tuckwell, N. & Barrett, T.E. (1998): A Critical Review of Functional Capacity Evaluations. *Physical Therapy*, 78, 8, S. 852-866.
- Kopp, H.G., Oliveri, M. & Thali, A. (1997/98). Erfassung und Umgang mit Symptomausweitung. *Suva Medizinische Mitteilungen*, Winter 1997/98, S. 56-78.
- Kraus, J. (1997). The independent medical examination and the functional capacity evaluation. *Occupational medicine*. 12, 3, S. 525-556.
- Kürten, L. (2004). Chronischer Schmerz. Ergebnisse der Forschung verbessern die Versorgung der Patienten. Broschüre des Bundesministeriums für Bildung und Forschung (BMBF). http://www.bmbf.de/pub/chronischer_schmerz.pdf
- Lackner, J.M., Carosella, A.M. & Feuerstein, M. (1996). Pain Expectancies, Pain, and Functional Self-Efficacy Expectancies as Determinants of Disability in Patients With Chronic Low Back Disorders. *Journal of Consulting and Clinical Psychology*, 64, 1, S. 212-220.
- Lemstra, M., Olszynski, W.P. & Enright, W. (2004). The sensitivity and specificity of functional capacity evaluations in determining maximal effort: a randomized trial. *Spine*, 29, 9, S. 953-959.
- Lienert, G. & Raatz, U. (1994). Testaufbau und Analyse. Beltz PVU, Weinheim.

- Lux, A. (2007). Mail vom 14.11.2007.
- Matheson, L.N. & Matheson, M.L. (1989). Performance and Capacity Testing (PACT). Hrsg: Performance and Capacity Testing (PACT) ®, Kalifornien
- Matheson, L.N., Isernhagen, S.J. & Hart, D.L. (2002). Relationships Among Lifting Ability, Grip Force, And Return to Work. *Physical Therapy*, 82, 3, S. 249-256.
- Nellesen, G., Froböse, I., Schüle, K. & Schian, H.M. (2002). Leistungsbeurteilungen in Bezug auf den 8-Stunden-Arbeitstag – zwischen Anspruch und Wirklichkeit. Tagungsband zum 11. Rehabilitationswissenschaftlichen Kolloquium „Teilhabe durch Rehabilitation“ vom 04.03. bis 06.03.2002 in München. DRV-Schriften Band 33, S. 75-77.
- Nellesen, G. & Schüle, K. (2000). Zur Problematik des Transfers der US-amerikanischen FCE-Verfahren (Functional Capacity Evaluation) auf deutsche Verhältnisse. Tagungsband zum 9. Rehabilitationswissenschaftlichen Kolloquium „Individualität und Rehaprozess“ vom 13.03. bis 15.03.2000 in Würzburg. DRV-Schriften Band 20, S. 179-180.
- Oesch, P. (2000). Die Evaluation der funktionellen Leistungsfähigkeit: Outcome bei verschiedenen Nationalitäten. *Schweizerische Ärztezeitung*, 81, 47, S. 2658-2659.
- Oliveri, M., Denier-Bont, F. & Hallmark Itty, M.L. (1996). Evaluation der funktionellen Leistungsfähigkeit (EFL) nach Susan Isernhagen. *Suva Medizinische Mitteilungen* Nr. 69, Dezember 1996, Suva.
- Reneman, M.F., Jaegers, S.M., Westmaas, M. & Goeken, L.N. (2002). The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work*, 18, 1, S. 23-27.
- Reneman, M.F., Jorritsma, W., Dijkstra, S.J. & Dijkstra, P.U. (2003). Relationship between kinesiophobia and performance in a functional capacity evaluation. *Journal of Occupational Rehabilitation*, 13, 4, S. 277-285.
- Reneman, M.F., Jorritsma, W., Schellekens, J.M. & Goeken, L.N. (2002). Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. *Journal of Occupational Rehabilitation*, 12, 3, S. 119-129.
- Rivier, G. & Seewer, M. (2002). Evaluation der funktionellen Leistungsfähigkeit. *Suva – Medizinische Mitteilungen*, Nr. 73, S. 33-41.
- Schweizerische Arbeitsgemeinschaft Rehabilitation (SAR) (1997). Tests an den zwei Tagen. Auszug aus dem EFL-Manual für Kursunterlagen; Ausdruck vom 24.10.1997.
- Schweizerische Arbeitsgemeinschaft für Rehabilitation (SAR) (1998). Selbsteinschätzung der körperlichen Fähigkeiten (PACT). Luzern.

- Schweizerische Arbeitsgemeinschaft Rehabilitation (SAR) (1999). Objektive Beurteilungskriterien am Beispiel „Heben Boden zu Taillenhöhe“. Auszug aus dem EFL-Manual für Kursunterlagen; Ausdruck vom 16.02.1999.
- Saunders, R.L., Beissner, K.L. & McManis, B.G. (1997). Estimates of Weight That Subjects Can Lift Frequently in Functional Capacity Evaluations. *Physical Therapy*, 77, 12, S. 1717-1728.
- Schaarschmidt, U. & Fischer, A. (2003). Arbeitsbezogene Verhaltens- und Erlebensmuster AVEM. 2. Auflage. Swets Test Services, Frankfurt am Main.
- SOMEKO (2004). Abschlussbericht der Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung – SOMEKO. VDR-Schriften Band 53. Hrsg.: Verband Deutscher Rentenversicherungsträger.
- Strong, S., Baptiste, S., Clarke, J., Cole, D. & Costa, M. (2004). Use of functional capacity evaluations in workplaces and the compensation system: A report on workers' and report users' perceptions. *Work*, 23, S. 67-77.
- World Health Organisation (2006). BMI Classification.
http://www.who.int/bmi/index.jsp?introPage=intro_3.html (Zugriff: 26.11.2007)
- Zwerenz, R., Knickenberg, R.J., Schattenburg, L. & Beutel, M.E. (2004). Berufliche Belastungen und Ressourcen berufstätiger Patienten der psychosomatischen Rehabilitation im Vergleich zur Allgemeinbevölkerung. *Rehabilitation*, 43, S. 10-16.

14. Eigene Veröffentlichungen und Präsentationen der Studie

Veröffentlichungen:

- Büschel, C., Greitemann, B. & Schaidhammer-Placke, M. (2007). Worüber keiner spricht – Beschwerdezunahme bei EFL-Tests. *Orthopädische Praxis* 43, 9, S. 499-503.
- Büschel, C., Schaidhammer-Placke, M., Greitemann, B. (2007). Welche Rolle kann ein technisches Verfahren wie die Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL) in der sozialmedizinischen Leistungsbeurteilung spielen? *Medizinisch-Orthopädische Technik* 127, 4, S. 21-44. Ausgezeichnet mit dem MOT-Preis 2007.
- Büschel, C., Greitemann, B. & Schaidhammer, M. (2008). Stellenwert der Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL) in der sozialmedizinischen Begutachtung des Leistungsvermögens. Teil 1: Möglichkeiten und Grenzen des Verfahrens nach dem aktuellen Stand der Wissenschaft. *Der medizinische Sachverständige*, 104, 5, S. 195-200.
- Büschel, C., Greitemann, B. & Schaidhammer, M. (2008). Stellenwert der Evaluation der funktionellen Leistungsfähigkeit nach Isernhagen (EFL) in der sozialmedizinischen Begutachtung des Leistungsvermögens. Teil 2: Eigene Ergebnisse zu Nutzen und Risiken des Verfahrens für Gutachter und Probanden. *Der medizinische Sachverständige*, 104, 6, S. 211-219.

Vorträge und Poster

- Büschel, C., Greitemann, B. & Schaidhammer-Placke, M. (2006). Stellenwert der EFL nach Isernhagen in der sozialmedizinischen Beurteilung stationärer orthopädischer Reha-Patienten. Vortrag auf dem 15. Rehabilitationswissenschaftlichen Kolloquium „Rehabilitation und Arbeitswelt – Herausforderungen und Strategien“ vom 13. bis zum 15.03.2006 in Bayreuth. Abstract: *DRV Schriften Band 64*, S. 43-45.
- Büschel, C., Greitemann, B., Schaidhammer-Placke, M. (2007). Worüber keiner spricht – Beschwerdezunahme bei EFL-Tests. Vortrag auf der 55. Jahrestagung der Vereinigung Süddeutscher Orthopäden e.V. vom 26. bis zum 29.04.2007, Baden-Baden.
- Büschel, C., Schaidhammer, M. & Greitemann, B. (2007). Sind die Testbedingungen der EFL nach Isernhagen zu alltagsfern? Poster auf dem 16. Rehabilitationswissenschaftlichen Kolloquium „Gesund älter werden – mit

Prävention und Rehabilitation“ vom 26. bis zum 28. 03.2007 in Berlin.
Abstract: DRV Schriften Band 72, S. 263-264.

Büschel, C., Schaidhammer-Placke, M. & Greitemann, B. (2007). Auswirkungen der EFL nach Isernhagen auf Selbsteinschätzung und berufliche Perspektiven der Patienten. Vortrag auf dem 16. Rehabilitationswissenschaftlichen Kolloquium „Gesund älter werden – mit Prävention und Rehabilitation“ vom 26. bis zum 28.03.2007 in Berlin.

Abstract: DRV Schriften Band 72, S. 376-377.

Büschel, C., Schaidhammer-Placke, M. & Greitemann, B. (2007). Die EFL nach Isernhagen – Diagnostikum und Intervention in der stationären orthopädischen Rehabilitation. Vortrag auf der Tagung „Krankheitsbezogene Forschung in der medizinischen Rehabilitation“ des Rehabilitationszentrums Seehof vom 22. bis 23.06.2007 in Teltow/Berlin.