APPENDIX B Knee Methodological Quality

Red type = Author's comments and methodological quality ratings Strikethrough indicates that the grading criteria are not applicable

Battaglia/Am J Sports Med/2007	
L Baffaglia / Am L Shorfs Med / / IIII /	
Dattagna/Mili Sports Mea/2007	

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality

C. Hypothesis Testing/Construct Validity N=102

Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of	NA	NA
percentage of	missing items	missing items		
missing items	described	NOT described		
given?				
none				
Was there a	Described how	Not described	Not clear how	
description of	missing items	but it can be	missing items were	
how missing	were handled	deduced how	handled	
items were		missing items	Tiuriuru	
handled? none		were handled		
Was the sample	Adequate sample	Good sample size	Moderate sample	Small sample
size included in	size (≥100 per	(50-99 per	size (30-49 per	size (<30 per
the analysis	analysis)	analysis)	analysis)	analysis)
adequate?	•			
_				
Were hypotheses	hypotheses	Minimal number	Hypotheses vague	Unclear what
regarding	formulated a	of hypotheses	or not formulated	was expected
correlations or	priori	formulate a	but possible to	
mean differences		priori	deduce what was	
formulated a			expected	
priori (i.e. before				
data collection)?				
Was the expected	Expected	Expected		
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
included in the		stated –		
hypotheses?	Evnocted	Evnocted		
Was the expected	Expected	Expected		

absolute or relative magnitude of correlations or mean differences included in the hypotheses?	magnitude of the correlations or differences stated	magnitude of the correlations or differences NOT stated		NO 1
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Bjorklund/Knee Surg Sports Traumatol/2006

STEP1: Evaluated measurement properties in the article

DI DI TI DVATAGE A MEGISTI CINCI	ordin 1. Dranautea measarement properties in the article					
Check if present	Property	Location				
✓	Reliability	Box A				
	Agreement/Measurement Error	Box B				
1	Hypothesis Testing / Construct	Box C				

validity	
Criterion/Predictive Validity	Box D

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability

A. Remadinty	I	Ι	T	T _
Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing	Percentage of	Percentage of		
subjects given?	missing	missing		
none	subjects	subjects NOT		
	described	described		
Was there a description of	Described how	Not described	Not clear how	
how missing subjects were	missing	but it can be	missing	
handled? none	subjects were	deduced how	subjects were	
	handled	missing	handled	
		subjects were		
		handled		
Was the sample size included	Adequate	Good sample	Moderate	Small sample
in the analysis adequate?	sample size	size (50-99)	sample size	size (<30)
	(≥100)		(30-49)	
Were at least two	At least two			Only one
measurements available?	measurements			measurement
Intra and inter intra only				
Were the administrations	Independent	Assumable	Doubtful	Measurements
independent? Intra	measurements	that the	whether the	NOT
		measurement	measurements	independent
		s were	were	
		independent	independent	
Was the time interval stated?	Time interval		Time interval	
Intra	stated		NOT stated	
Were patients stable in the	Patients were	Assumable	Unclear if	Patients were
interim period on the	stable	that patients	patients were	NOT stable
construct to be measured?	(evidence	were stable	stable	
Intra	provided)			
Was the time interval	Time interval		Doubtful	Time interval
appropriate?	appropriate		whether time	NOT
Intra			interval was	appropriate
			appropriate	
Were the test conditions	Test	Assumable	Unclear if test	Test
similar for both	conditions	that test	conditions	conditions
measurements? e.g. type of	were similar	conditions	were similar	were NOT
administration, environment,	(evidence	were similar		similar
instructions	provided)			
Were there any important	No other		Other minor	Other
flaws in the design or	important		methodologica	important
methods of the study?	methodologica		l flaws in the	methodologica
Intra	l flaws in the		design or	l flaws in the
	design or		execution of	design or
	execution of		the study	execution of
	the study			the study
for continuous scores: Was an	ICC calculated	ICC calculated	Pearson or	No ICC or
intraclass correlation	and model or	but model or	Spearman	Pearson or
coefficient (ICC) calculated?	formula of the	formula of the	correlation	Spearman
	ICC is	ICC not	coefficient	correlations

	described	described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
discriminantfor ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

C. Hypothesis Testing / Construct Validity

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled? none	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or	Expected direction of the correlations or	Expected direction of the correlations or		

mean differences included in the hypotheses?	differences stated	differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Bjorklund/Knee Surg Sports Traumatol/2009

Check if present	Property	Location	
	Reliability	Box A	
	Agreement/Measurement Error	Box B	
1	Hypothesis Testing / Construct validity	Box C	
	Criterion/Predictive Validity	Box D	
✓	Responsiveness	Box E	
	Predictive/Criterion validity	Box F	

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity

di nij potnesis res				
Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled? none	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an	Adequate description of the	Adequate description of	Poor description of the constructs	NO description of the constructs

- 1				
adequate	constructs	most of the	measured by the	measured by the
description	measured by the	constructs	comparator	comparator
provided of the	comparator	measured by the	instrument(s)	instrument(s)
comparator	instrument(s)	comparator		
instrument(s)?		instrument(s)		
for convergent	Adequate	Adequate	Some information	No information
validity: Were the	measurement	measurement	on measurement	on the
measurement	properties of the	properties of the	properties (or a	measurement
properties of the	comparator	comparator	reference to a study	properties of the
comparator	instrument(s) in a	instrument(s)	on measurement	comparator
instrument(s)	population	but not sure if	properties) of the	instrument(s)
adequately	similar to the	these apply to	comparator	NOTE:
described?	study population	the study	instrument(s) in	First time tested
	IKDC, SF-36	population	any study	in this study
			population	
Were there any	No other		Other minor	Other important
important flaws	important		methodological	methodological
in the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
	execution of the		data presented on a	study
	study		comparison with	
	•		an instrument that	
			measures another	
			construct)	
Were design and	Statistical	Assumable that	Statistical methods	Statistical
statistical	methods applied	statistical	applied NOT	methods applied
methods	appropriate	methods were	optimal	NOT appropriate
adequate for the		appropriate, e.g.		'' '
hypotheses to be		Pearson		
tested?		correlations		
		applied, but		
		distribution of		
		scores or mean		
		(SD) not		
		presented		
E. Responsiveness	<u>. </u>	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1	1

L. Responsiveness				
Design requirements	Excellent	Good	Fair	Poor
Was the	Percentage of	Percentage of		
percentage of	missing items	missing items		
missing items	described	NOT described		
given?	described	nor described		
Was there a	Described how	Not described	Not clear how	
description of how	missing items	but it can be	missing items were	
missing items	were handled	deduced how	handled	
were handled?		missing items		
were nandicu:		were handled		
Was the sample	Adequate sample	Good sample	Moderate sample	Small sample size
size included in	size (≥100)	size (50-99)	size (30-49)	(<30)
the analysis				
adequate?				
Was a longitudinal	Longitudinal			No longitudinal
design with at	design used			design used

least two				
measurement				
used?				
Was the time	Time interval			Time interval
interval stated?	adequately			NOT described
	described			
If anything	Anything that	Assumable what	Unclear or NOT	If anything
occurred in the	occurred during	occurred during	described what	occurred in the
interim period	the interim	the interim	occurred during	interim period
(e.g. intervention,	period (e.g.	period	the interim period	(e.g. intervention,
other relevant	treatment)			other relevant
events), was it	adequately			events), was it
adequately	described			adequately
described?				described?
Was a proportion	Part of the	NO evidence	Unclear if part of	Patients were
of the patients	patients were	provided, but	the patients were	NOT changed
changed (i.e.	changed	assumable that	changed	
improvement or	(evidence	part of the		
deterioration)?	provided)	patients were		
		changed		
For constructs for	which a gold standa	rd was not availal	ble:	
More have ath age	H-mathagas		II	Unclear what was
Were hypotheses	Hypotheses formulated a		Hypotheses vague or not formulated	
about changes in scores formulated				expected
	priori		but possible to deduce what was	
a priori (i.e. before				
data collection)?	F	Para a stand	expected	
Was the expected	Expected	Expected direction of the		
direction of	direction of the			
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT stated		
		i stateu		
of the change				
scores of HR-PRO				
scores of HR-PRO instruments				
scores of HR-PRO instruments included in these				
scores of HR-PRO instruments included in these hypotheses?				
scores of HR-PRO instruments included in these hypotheses? Were the expected	Expected	Expected		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or	magnitude of the	magnitude of		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude	magnitude of the correlations or	magnitude of the correlations		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or	magnitude of the	magnitude of the correlations or differences		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences	magnitude of the correlations or	magnitude of the correlations		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change	magnitude of the correlations or	magnitude of the correlations or differences		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO	magnitude of the correlations or	magnitude of the correlations or differences		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments	magnitude of the correlations or	magnitude of the correlations or differences		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these	magnitude of the correlations or	magnitude of the correlations or differences		
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	magnitude of the correlations or differences stated	magnitude of the correlations or differences	Door description of	NO description of
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	magnitude of the correlations or differences stated	magnitude of the correlations or differences	Poor description of	NO description of
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses? Was an adequate description	magnitude of the correlations or differences stated Adequate description of the	magnitude of the correlations or differences	the constructs	the constructs
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses? Was an adequate description provided of the	Adequate description of the constructs	magnitude of the correlations or differences	the constructs measured by the	the constructs measured by the
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses? Was an adequate description provided of the comparator	Adequate description of the constructs measured by the	magnitude of the correlations or differences	the constructs measured by the comparator	the constructs measured by the comparator
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses? Was an adequate description provided of the	Adequate description of the constructs	magnitude of the correlations or differences	the constructs measured by the	the constructs measured by the
scores of HR-PRO instruments included in these hypotheses? Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses? Was an adequate description provided of the comparator	Adequate description of the constructs measured by the comparator	magnitude of the correlations or differences	the constructs measured by the comparator	the constructs measured by the comparator

comparator instrument(s) in appulation described? Were there any important study? Were design and statistical methods of the study? Were design and statistical methods adequate for the hypotheses to be bested? For constructs for which a gold standard was available: Can the criterion considered as a reasonable gold standard? Were there any important flavs in the design or methods of the study. Were there any important flavs in the design or methods of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there any important flavs in the design or execution of the study. Were there an	properties of the	properties of the	properties of	properties (or a	measurement
instrument(s) and adequate(s) in apopulation similar to the study population similar to the study population study population of the study population study population or methods of the design or methods of the study? Were design and statistical methods adequate for change be considered an adequate gold standard? Can the criterion for change be considered an adequate gold standard? Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study in the design of execution of the study. Were design and statistical methods applied appropriate with an instrument that measures another constructs for which a gold standard was available: Can the criterion sor dequate gold standard? Carter on used can be considered an adequate gold standard? (evidence provided) Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. So that is titled and adequate gold standard. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. So the study is the design or execution of the study. Were there any important flaws in the design or execution of the study. So the study is the study is the study is the study is the study. Were there any important flaws in the design or execution of the study. So the study is the study is the study is the study is the study. So the study is the stud					
adequately described? Statistical methods of the study properties of the study population Statistical methods applied appropriate hypotheses to be tested?	_	_	_		
described? similar to the study population Were there any important flaws in the design or methods of the study Were design and statistical methods adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for change be considered an areasonable gold standard? Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were design and statistical methods adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion standard? Can the criterion sed considered as a reasonable gold standard? Were there any important flaws in the design or methods of the study. Were there any important flaws in the design or methods of the study. Were there any important flaws in the design or execution of the study. For constructs for which a gold standard' was available: Criterion used can be considered an adequate gold standard' standard' (evidence provided) standard' were there any important flaws in the design or methods of the study. For constructs for which a gold standard' Correlations or Area under the study. For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? For dichotomous scores: Were correcalculated? For dichotomous scores or the area under the Receiver Operator Curve (ROC) curve calculated. Sensitivity and specificity was specificity with a sensitivity and specificity was specificity was sensitivity and specificity was specificity was sensitivity and specificity was sensitive			` ,	-	_
Were there any important flaws in the design or execution of the study where there any important flaws in the design or execution of the study where t					moer amene(o)
Were there any important flaws in the design or execution of the study Were design and sadequate for the hypotheses to be tested? For constructs for which a gold standard was available that reasonable gold standard? Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Statistical methods applied appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate	aeserisear				
Were there any important flaws in the design or methods of the study? Were design and statistical methods applied appropriate hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for change be considered an a reasonable gold standard? Were there any important flaws in the design or execution of the study to standard? Were there any important flaws in the design or execution of the study to standard? Were there any important flaws in the design or execution of the study to standard? Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Correlations the design or execution of the study. Sometimes are an adequate gold standard. No other important methodological flaws in the design or execution of the study. Correlations or Aucrolations or Aucro		beauty population	_	_	
Were there any important flaws in the design or execution of the study Were design and statistical methods applied appropriate Were design and statistical methods applied appropriate Tor constructs for which a gold standard was available: Can the criterion for change be considered an adequate for change be considered an provided) Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Statistical methods applied appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied NOT appropriate Unclear whether the criterion used can be considered an adequate 'gold standard			population	` ,	
Were there any important flaws in the design or methodological flaws in the design or execution of the study					
important flaws in the design or methodological flaws in the design or execution of the study? Were design and statistical methods adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for change be considered as a reasonable gold standard? Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Were there any important flaws in the design or execution of the study. Correlations or Area under the ROC Curve (AUC) calculated. For continuous scores: Were correlations or Area under the Roc Curve (AUC) calculated. Sensitivity and specificity each specificity and specificity and specificity and specificity and specificity calculated.	Were there any	No other			Other important
methodological flaws in the design or execution of the study? Were design and statistical methods applied adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for can be considered an adequate gold standard' vere there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Were there any important flaws in the design or execution of the study Statistical methods applied appropriate No evidence provided, but can be considered an adequate 'gold standard' Statistical methods applied nethods applied NOT optimal Were there any important flaws in the design or execution of the study Were there any important methodological flaws in the design or execution of the study Statistical methods applied approvided, but can be considered an adequate 'gold standard' Unclear whether the criterion used can be considered an adequate 'gold standard' Statistical methods applied nethods applied NOT optimal Word which a gold standard was available: Criterion used can be considered an adequate 'gold standard' Statistical methods applied nethods applied NOT optimal Statistical methods applied approvided, but can be considered an adequate 'gold standard' Statistical methods applied nethods applied NOT optimal Word which appropriate No evidence provided, but can be considered an adequate 'gold standard' Statistical methods applied and applied nethods applied not an adequate 'gold standard' Statistical methods applied an adequate 'gold standard' Criterion used can be considered an adequate 'gold standard' Statistical methods applied an adequate 'gold standard' Statistical methods applied an adequate 'gold standard' Correlations or AUC NOT calculated Statistical me					
methods of the study? Statistical methods adequate for the hypotheses to be tested? Standard?		-			O .
Study? design or execution of the study study (e.g. only data presented on a comparison with an instrument that measures another constructs (or which a gold standard was available: Can the criterion for change be considered as a reasonable gold standard' (evidence provided) Standard' (evidence provided) Standard' Statistical methods applied and adequate 'gold standard' Standard' Statistical methods applied and adequate 'gold standard' Statistical methods applied and adequate 'gold standard' Statistical methods applied hNOT appropriate Statistical methods applied approvided, but considered an adequate 'gold standard' Statistical methods applied hNOT appropriate Statistical methods applied approvided, but can be considered an adequate 'gold standard' Statistical methods applied hNOT appropriate Statistical methods applied hNOT appropriate Statistical methods applied approvided, but can be considered an adequate 'gold standard' Statistical methods applied hNOT appropriate Statistical methods applied approvided, but can be considered an adequate 'gold standard' Statistical methods applied hat the criterion used can be considered an adequate 'gold standard' Statistical methodological flaws in the design or execution of the study Statistical method approvided Statistical methodological flaws in the design or execution of the study Statistical methodological flaws in the design or execution of the study Statistical	_	_		_	design or
Study Statistical methods applied appropriate Statistical methods applied NOT optimal NOT appropriate Appropriate NOT ap	study?	design or		study (e.g. only	-
Study Study Study Statistical methods and adequate for the hypotheses to be tested? Statistical methods adequate for the hypotheses to be tested? Statistical methods applied appropriate Statistical methods applied not optimal Statistical methods applied not optimal Statistical methods applied NOT appropriate Statistical methods applied NOT appropriate NOT appropriate Statistical methods applied NOT appropriate NOT appropriate NOT appropriate Statistical methods applied NOT appropriate NOT appropriate NOT appropriate Statistical methods applied NOT appropriate NOT appropriate NOT appropriate Statistical methods applied noT with applied noT appropriate Statistical methods applied noT appropriate Statisti	J	_			study
Were design and statistical methods applied adequate for the hypotheses to be tested? Statistical methods applied appropriate Statistical methods applied appropriate Statistical methods applied appropriate Statistical methods applied appropriate Statistical methods applied NOT optimal NOT appropriate		study		_	
Were design and statistical methods applied adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for change be considered an adequate 'gold standard' Were there any important flaws in the design or methods of the study? Were there any important flaws in the design or methods of the study? For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? For dichotomous scales: Were sensitivity and specificity Were design and statistical methods applied appropriate statistical methods applied applied appropriate subject options applied NOT appropriate NOT app				-	
Were there any important flaws in the design or methods of the study? No other important flaws in the design or methods of the study? For continuous scales: Were correlations between change scensitivity and specificity Sensitivity and specificity Sensitivity and specificity Statistical methods applied appropriate Statistical methods applied appropriate Statistical methods applied appropriate Mo evidence optimal Mot appropriate Mot appro					
statistical methods adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for change be considered an adequate 'gold standard' Were there any important flaws in the design or methodological flaws in the study? For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? For constructs for which a gold standard was available: Criterion used can be provided, but assumable that adequate 'gold standard' (evidence onsidered an adequate 'gold standard' (evidence onsidered an adequate 'gold standard') Were there any important flaws in the design or methodological flaws in the design or execution of the study? Correlations or Area under the Roc Curve (ROC) curve calculated? For dichotomous seales: Were sensitivity and specificity Sensitivity and specificity To the minor methodological flaws in the design or execution of the study Sensitivity and specificity To the ninor methodological flaws in the design or execution of the study Sensitivity and specificity Sensitivity and specificity To the ninor methodological flaws in the design or execution of the study Sensitivity and specificity Sensitivity and specificity Sensitivity and specificity To the ninor methodological flaws in the design or execution of the study Sensitivity and specificity Sensitivity and specificity Sensitivity and specificity Sensitivity and specificity				construct)	
statistical methods adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion considered as a reasonable gold standard' Were there any important flaws in the design or methodological flaws in the study? For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? For constructs for which a gold standard was available: Criterion used can be considered an assumable that can be considered an adequate 'gold standard' (evidence provided) standard' (evidence considered an adequate 'gold standard' No other important methodological flaws in the design or execution of the study? Correlations or Area under the Roc Curve (AUC) calculated For dichotomous scales: Were sensitivity and specificity Sensitivity and specificity Criterion used can be considered an adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Sensitivity and specificity Sensitivity and specificity calculated Triterion used can be considered an adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Correlations or Area under the Roc Curve (AUC) calculated Sensitivity and specificity Correlations or Auc NoT calculated Sensitivity and specificity NoT calculated	Were design and	Statistical		Statistical methods	Statistical
adequate for the hypotheses to be tested? For constructs for which a gold standard was available: Can the criterion for change be considered an reasonable gold standard' Were there any important flaws in the design or methodological flaws in the study? For continuous scores: Were correlations or Area under the ROC Curve (ROC) curve calculated? For constructs for which a gold standard was available: Criterion used can be provided, but assumable that can be considered an adequate 'gold standard' Unclear whether the criterion used can NOT be considered an adequate 'gold standard' Unclear whether the criterion used can NOT be considered an adequate 'gold standard' Were there any important flaws in the design or execution of the study? Were there any important flaws in the design or execution of the study? Correlations or Area under the Roc Curve (ROC) curve calculated? For dichotomous scales: Were sensitivity and specificity calculated Sensitivity and specificity calculated Ano vidence provided, but can be considered an adequate 'gold standard' Unclear whether the criterion used can be considered an adequate 'gold standard' Unclear whether the criterion used can be considered an adequate 'gold standard' Standard' Other minor methodological flaws in the design or execution of the study Study Correlations or Area under the ROC Curve (AUC) calculated Sensitivity and specificity and specificity calculated	statistical methods	methods applied		applied NOT	methods applied
For constructs for which a gold standard was available: Can the criterion for change be considered an reasonable gold standard? Were there any important flaws in the design or methods of the study? For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? For constructs for which a gold standard was available: Criterion used can be provided, but assumable that adequate 'gold standard' Unclear whether the criterion used can be considered an adequate 'gold standard' Unclear whether the criterion used can be considered an adequate 'gold standard' Considered an adequate 'gold standard' No other important methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study For continuous scores: Were correlations Exercive Operator Curve (AUC) calculated Sensitivity and specificity Sensitivity and specificity scales: Were sensitivity and specificity scales: Were can be considered an adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Correlations or Auc NOT calculated Sensitivity and specificity or calculated	adequate for the	appropriate		optimal	NOT appropriate
For constructs for which a gold standard was available: Can the criterion for change be considered as a reasonable gold standard? Were there any important flaws in the design or methods of the study? For continuous scores: Were correlations or Area under the Roc Curve (ROC) curve calculated? For dichotomous scales: Were sensitivity and specificity Sandard was available: No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard' Unclear whether the criterion used can bor can be considered an adequate 'gold standard' Standard' Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Correlations or Area under the Roc Curve (AUC) calculated Sensitivity and specificity calculated Sensitivity and specificity Criterion used can NOT be considered an adequate 'gold standard' Standard' Other minor methodological flaws in the design or execution of the study Correlations or AUC NOT calculated Sensitivity and specificity calculated Sensitivity and specificity calculated	hypotheses to be			_	
Can the criterion for change be considered as a reasonable gold standard? Were there any important flaws in the design or methods of the study? For continuous scores: Were correlations between change scores, or the and under the Receiver Operator Curve (ROC) curve calculated? Criterion used can be considered an adequate 'gold standard' No other important methodological flaws in the design or execution of the study Correlations or Area under the Receiver Operator Curve (ROC) curve calculated? Sensitivity and specificity Criterion used can NOT be considered an adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Correlations or Area under the Roc Curve (AUC) calculated Sensitivity and specificity calculated Sensitivity and specificity calculated	tested?				
for change be considered as a reasonable gold standard? Were there any important flaws in the design or methodological flaws in the design or execution of the study? For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? For dichotomous scales: Were sensitivity and specificity For considered an adequate 'gold standard' No other important methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Correlations or Area under the ROC Curve (AUC) calculated For dichotomous scales: Were sensitivity and specificity adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Correlations or Area under the Roc Curve (AUC) calculated Sensitivity and specificity adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Sensitivity and specificity calculated			T .	T	
considered as a reasonable gold standard? Sericitivity and specificity assumable that the criterion used can be considered an adequate 'gold standard' (covidence provided) Standard' (evidence provided) Standard' (evidence provided) Standard' (evidence provided) Standard' No other important methodological flaws in the design or execution of the study Standard' Other minor methodological flaws in the design or execution of the study For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? Sensitivity and specificity Sensitivity and specificity can be considered an adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Sensitivity and specificity assumable that the criterion used can be considered an adequate 'gold standard' Standard' Other minor methodological flaws in the design or execution of the study Scoresitivity and specificity ealeulated Sensitivity and specificity calculated Sensitivity and specificity calculated					
reasonable gold standard? Standard Stan	_		_		
standard? Standard' (evidence provided) Standard' Standard'					
Cevidence provided Considered an adequate 'gold standard'	_				
Were there any important flaws in the design or methodological flaws in the design or execution of the study? for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? Sensitivity and specificity provided) adequate 'gold standard' Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Correlations or Area under the ROC Curve (AUC) calculated Sensitivity and specificity ealculated Sensitivity and specificity ealculated Sensitivity and specificity ealculated	standard?			standard'	standard'
Were there any important flaws in the design or methodological flaws in the design or methodological flaws in the design or execution of the study? for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity for dichotomous scales: Were calculated for dichotomous specificity standard' Other minor methodological flaws in the design or execution of the study Other minor methodological flaws in the design or execution of the study Correlations or Area under the ROC Curve (AUC) calculated Sensitivity and specificity calculated Sensitivity and specificity calculated		*			
Were there any important flaws in the design or methodological flaws in the design or methodological flaws in the design or execution of the study? For continuous scores: Were correlations or hetwork excores, or the area under the Receiver Operator Curve (ROC) curve calculated? For dichotomous scales: Were sensitivity and specificity For dichotomous specificity For dic		provided)			
important flaws in the design or methodological flaws in the design or methods of the study? for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity for dichotomous scales: Were sensitivity and specificity for dichotomous scales: Were sensitivity and specificity important flaws in the design or execution of the study flaws in the design or execution of the study calculated flaws in the design or execution of the study calculated flaws in the design or execution of the study calculated flaws in the design or execution of the study calculated flaws in the design or execution of the study calculated flaws in the design or execution of the study calculated study Correlations or AUC NOT calculated Sensitivity and specificity value specificity value specificity calculated	Were there any	No other	standard	Other minor	Other important
the design or methodological flaws in the design or execution of the study? for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity for dichotomous scales: Were sensitivity and specificity methodological flaws in the design or execution of the study for execution of the study Correlations or Auc NOT calculated Correlations or Auc NOT calculated Sensitivity and specificity calculated Sensitivity and specificity calculated					
methods of the study? for continuous correlations or Area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity for design or execution of the study for execution of the study Correlations or Area under the AUC NOT calculated Correlations or Auch NOT calculated Sensitivity and specificity calculated Sensitivity and specificity calculated	_	_			
study? design or execution of the study for continuous scores: Were correlations or Area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity specificity design or execution of the study execution of the study Correlations or AUC NOT calculated AUC NOT calculated Sensitivity and specificity calculated Sensitivity and specificity calculated		_			
execution of the study for continuous correlations or Area under the correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity execution of the study Correlations or AUC NOT calculated Calculated Sensitivity and specificity calculated Sensitivity and specificity calculated					
for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity specificity correlations or AUC NOT calculated calculated Sensitivity and specificity calculated Sensitivity and specificity calculated	study.			Study	
for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity specificity Correlations or AUC NOT calculated Sensitivity and specificity calculated Sometimes of Auc NoT calculated Sensitivity and specificity calculated Sensitivity and specificity calculated					Study
scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity calculated Auc NOT calculated Sensitivity calculated Sensitivity and specificity calculated Sensitivity and specificity calculated	for continuous				Correlations or
between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity sensitivity and specificity calculated calculated specificity calculated calculated calculated	scores: Were	Area under the			
between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity sensitivity and specificity calculated calculated specificity calculated calculated calculated	correlations				
scores, or the area under the Receiver Operator Curve (ROC) curve calculated? for dichotomous Sensitivity and scales: Were sensitivity and specificity calculated specificity sensitivity and specificity specificity specificity specificity specificity calculated					
under the Receiver Operator Curve (ROC) curve calculated? for dichotomous scales: Were sensitivity and specificity sensitivity and specificity specificity specificity specificity specificity specificity					
Curve (ROC) curve calculated? for dichotomous seales: Were specificity sensitivity and specificity sensitivity and specificity specificity specificity specificity specificity specificity					
Curve (ROC) curve calculated? for dichotomous seales: Were specificity sensitivity and specificity sensitivity and specificity specificity specificity specificity specificity specificity	Receiver Operator				
for dichotomous Sensitivity and scales: Were sensitivity and specificity Sensitivity and specificity Sensitivity and specificity sensitivity and specificity calculated calculated	Curve (ROC) curve				
scales: Were specificity sensitivity and specificity specificity sensitivity and specificity specifici	calculated?				
scales: Were specificity sensitivity and calculated specificity calculated		Sensitivity and			Sensitivity and
specificity					
	•	calculated			calculated
(changed versus					
tenangeu versus	(changed versus				

not changed)		
determined?		

Carter/Br J Sports Med/ 1997

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
/	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
✓	Responsiveness	Box E
	Predictive validity/Criterion/Predictive Validity	Box F

Step 2. Determining if a study meets the standards for good methodological quality

C. Hypothesis Testing / Construct Validity

Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of	NA	NA
percentage of	missing items	missing items		
missing items	described	NOT described		
given?				
none				
Was there a	Described how	Not described	Not clear how	
description of	missing items	but it can be	missing items were	
how missing	were handled	deduced how	handled	
items were		missing items		
handled? none		were handled		
Was the sample	Adequate sample	Good sample size	Moderate sample	Small sample
size included in	size (≥100 per	(50-99 per	size (30-49 per	size (<30 per
the analysis	analysis)	analysis)	analysis)	analysis)
adequate?		50 + 23 controls		
Were hypotheses	hypotheses	Minimal number	Hypotheses vague	Unclear what
regarding	formulated a	of hypotheses	or not formulated	was expected
correlations or	priori	formulate a	but possible to	
mean differences		priori	deduce what was	
formulated a			expected	
priori (i.e. before				
data collection)?				
Was the expected	Expected	Expected		
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
included in the		stated –		
hypotheses?	Evmostod	Ermostad		
Was the expected absolute or	Expected	Expected magnitude of the		
absolute of	magnitude of the	magnitude of the		

relative magnitude of correlations or mean differences included in the hypotheses?	correlations or differences stated	correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population IKDC, SF-36	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

E. Responsiveness

2. Hesponsi. eness				
Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of		
percentage of	missing items	missing items		
missing items	described	NOT described		
given? none				
Was there a	Described how	Not described	Not clear how	
description of how	missing items	but it can be	missing items were	

missing items were handled?	were handled	deduced how missing items were handled	handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Was a longitudinal design with at least two measurement used?	Longitudinal design used			No longitudinal design used
Was the time interval stated?	Time interval adequately described			Time interval NOT described
If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?
Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided) which a gold standa	NO evidence provided, but assumable that part of the patients were changed rd was not availal	Unclear if part of the patients were changed	Patients were NOT changed
	Willen a gola stalla			
Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
Were the expected absolute or relative magnitude of correlations or mean differences of the change	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		

hypotheses?				
Was an adequate	Adequate		Poor description of	NO description of
description	description of the		the constructs	the constructs
provided of the	constructs		measured by the	measured by the
comparator	measured by the		comparator	comparator
instrument(s)?	comparator		instrument(s)	instrument(s)
moti amentioj.	instrument(s)		moti ament(s)	motrument(s)
Were the	Adequate	Adequate	Some information	NO information
measurement	measurement	measurement	on measurement	on the
properties of the	properties of the	properties of	properties (or a	measurement
comparator	comparator	the comparator	reference to a	properties of the
instrument(s)	instrument(s) in	instrument(s)	study on	comparator
adequately	a population	but not sure if	measurement	instrument(s)
described?	similar to the	these apply to	properties) of the	moti umencioj
acseribea.	study population	the study	comparator	
	study population	population	instrument(s) in	
		population	any study	
			population	
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
study.	execution of the		data presented on	study
	study		a comparison with	Study
	Study		an instrument that	
			measures another	
			construct)	
Were design and	Statistical		Statistical methods	Statistical
statistical methods	methods applied		applied NOT	methods applied
adequate for the	appropriate		optimal	NOT appropriate
hypotheses to be	арргорише		op viii ui	nor appropriate
tested?				
	which a gold standa	rd was available:	I.	
Can the criterion	Criterion used	No evidence	Unclear whether	Criterion used
for change be	can be	provided, but	the criterion used	can NOT be
considered as a	considered an	assumable that	can be considered	considered an
reasonable gold	adequate 'gold	the criterion	an adequate 'gold	adequate 'gold
standard?	standard'	used can be	standard'	standard'
o tarratar a r	(evidence	considered an	Junuara	o darradi d
	provided)	adequate 'gold		
	provided	standard'		
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study	execution of the
 	execution of the			study
	study			
for continuous	Correlations or			Correlations or
scores: Were	Area under the			AUC NOT
correlations	ROC Curve (AUC)			calculated
between change	calculated			
scores, or the area				
stores, or the area	l	<u> </u>	I	l

under the Receiver Operator Curve (ROC) curve calculated?			
for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated		Sensitivity and specificity NOT calculated

Crossley/J Orthop Res/ 2007
STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
✓	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
✓	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability n=10

A. Kellability II-10		1	1	T
Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing	Percentage of	Percentage of		
subjects given?	missing	missing		
none	subjects	subjects NOT		
	described	described		
Was there a description of	Described how	Not described	Not clear how	
how missing subjects were	missing	but it can be	missing	
handled? none	subjects were	deduced how	subjects were	
	handled	missing	handled	
		subjects were		
		handled		
Was the sample size included	Adequate	Good sample	Moderate	Small sample
in the analysis adequate?	sample size	size (50-99)	sample size	size (<30)
	(≥100)		(30-49)	
Were at least two	At least two			Only one
measurements available?	measurements			measurement
Intra and inter intra only				
Were the administrations	Independent	Assumable	Doubtful	Measurements
independent? Intra	measurements	that the	whether the	NOT
		measurement	measurements	independent
		s were	were	
		independent	independent	
Was the time interval stated?	Time interval		Time interval	
Intra	stated		NOT stated	
Were patients stable in the	Patients were	Assumable	Unclear if	Patients were
interim period on the	stable	that patients	patients were	NOT stable
construct to be measured?	(evidence	were stable	stable	
Intra	provided)			
Was the time interval	Time interval		Doubtful	Time interval

appropriate? Intra	appropriate- 7-10 days		whether time interval was appropriate	NOT appropriate
Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
Were there any important flaws in the design or methods of the study? Intra	No other important methodologica I flaws in the design or execution of the study		Other minor methodologica I flaws in the design or execution of the study	Other important methodologica I flaws in the design or execution of the study
for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

D. Criterion/Predictive Validity (Predictive in this case n= 58 but only 27 used in analysis

D. Criterion/11 curetive variately (11 curetive in this case in 30 but only 27 used in analysis				
Design	Excellent	Good	Fair	Poor
requirements				
Was the percentage of missing items	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
given? none	described	uescribeu		
Was there a	Described how	Not described but it	Not clear how	
description of	missing items	can be deduced how	missing items	
how missing	were handled	missing items were	were handled	

items were handled? none	Adamsta sampla	handled- did not use controls in the final linear regression model	Madayata gamala	Con all games also
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30) N=27
Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study	Were there any important flaws in the design or methods of the study?
for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated			Correlations or AUC NOT calculated
for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Eastlack/MSSE/ 1999

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
✓	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity n=45

Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of	NA	NA
percentage of	missing items	missing items		
missing items	described	NOT described		

	T	Г	T	T
given?				
none				
TAY .1	D 11 11	XI . 1	NI 1 1	
Was there a	Described how	Not described	Not clear how	
description of	missing items	but it can be	missing items were	
how missing	were handled	deduced how	handled	
items were		missing items		
handled? none	A.1	were handled	36.1	0 11 1
Was the sample	Adequate sample	Good sample size	Moderate sample	Small sample
size included in	size (≥100 per	(50-99 per	size (30-49 per	size (<30 per
the analysis	analysis)	analysis)	analysis)	analysis)
adequate?				
Were hypotheses	hypotheses	Minimal number	Hypotheses vague	Unclear what
regarding	formulated a	of hypotheses	or not formulated	was expected
correlations or	priori	formulate a	but possible to	was expected
mean differences	priori	priori	deduce what was	
formulated a		priori	expected	
priori (i.e. before			слрессеи	
data collection)?				
Was the expected	Expected	Expected		
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
included in the	amerenees stated	stated -		
hypotheses?				
Was the expected	Expected	Expected		
absolute or	magnitude of the	magnitude of the		
relative	correlations or	correlations or		
magnitude of	differences stated	differences NOT		
correlations or		stated		
mean differences				
included in the				
hypotheses?				
for convergent	Adequate	Adequate	Poor description of	NO description
validity: Was an	description of the	description of	the constructs	of the constructs
adequate	constructs	most of the	measured by the	measured by the
description	measured by the	constructs	comparator	comparator
provided of the	comparator	measured by the	instrument(s)	instrument(s)
comparator	instrument(s)	comparator		
instrument(s)?		instrument(s)		
for convergent	Adequate	Adequate	Some information	No information
validity: Were the	measurement	measurement	on measurement	on the
measurement	properties of the	properties of the	properties (or a	measurement
properties of the	comparator	comparator	reference to a study	properties of the
comparator	instrument(s) in a	instrument(s)	on measurement	comparator
instrument(s)	population	but not sure if	properties) of the	instrument(s)
adequately	similar to the	these apply to	comparator	NOTE:
described?	study population	the study	instrument(s) in	First time tested
		population	any study	in this study
Mono there are	No othor		population	Othonimasstart
Were there any	No other		Other minor	Other important
important flaws	important		methodological	methodological

in the design or methods of the study?	methodological flaws in the design or execution of the study		flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	flaws in the design or execution of the study
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

D. Criterion/Predictive Validity (Predictive in this case)

Design	Excellent	Good	Fair	Poor
requirements Was the percentage of missing items given? none	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled? none	Described how missing items were handled	Not described but it can be deduced how missing items were handled- did not use controls in the final linear regression model	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study	Were there any important flaws in the design or methods of the study?
for continuous scores: Were	Correlations or AUC calculated –			Correlations or AUC NOT

correlations, or	Multiple		calculated
the area under	regression used		
the receiver	for predictive		
operating curve	validity		
calculated?	•		
for dichotomous	Sensitivity and		Sensitivity and
scores: Were	specificity		specificity NOT
sensitivity and	calculated		calculated
specificity			
determined?			

Gauffin/Int J Sports Med/1990

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
•	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity n=30? Never said how many in ref group and no indication whether all were used in analyses

indication whether all were used in analyses					
Design requirements	Excellent	Good	Fair	Poor	
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA	
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled		
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)	
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected	

Was the expected	Expected	Expected		
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
included in the		stated –		
hypotheses?				
Was the expected	Expected	Expected		
absolute or	magnitude of the	magnitude of the		
relative	correlations or	correlations or		
magnitude of	differences stated	differences NOT		
correlations or		stated		
mean differences				
included in the				
hypotheses?				
for convergent	Adequate	Adequate	Poor description of	NO description
validity: Was an	description of the	description of	the constructs	of the constructs
adequate	constructs	most of the	measured by the	measured by the
description	measured by the	constructs	comparator	comparator
provided of the	comparator	measured by the	instrument(s)	instrument(s)
comparator	instrument(s)	comparator		
instrument(s)?	Adaguata	instrument(s)	Some information	No information
for convergent	Adequate	Adequate	on measurement	on the
validity: Were the measurement	measurement properties of the	measurement properties of the	properties (or a	measurement
properties of the	comparator	comparator	reference to a study	properties of the
comparator	instrument(s) in a	instrument(s)	on measurement	comparator
instrument(s)	population	but not sure if	properties) of the	instrument(s)
adequately	similar to the	these apply to	comparator	mstrument(s)
described?	study population	the study	instrument(s) in	
	population	population	any study	
		r-r-	population	
Were there any	No other		Other minor	Other important
important flaws	important		methodological	methodological
in the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
	execution of the		data presented on a	study
	study		comparison with	
			an instrument that	
			measures another	
YAY 1	Cr. rr rr 1	A 11 .1 .	construct)	Cr. r. r. 1
Were design and	Statistical	Assumable that	Statistical methods	Statistical
statistical	methods applied	statistical	applied NOT	methods applied
methods	appropriate	methods were	optimal	NOT appropriate
adequate for the		appropriate, e.g. Pearson		
hypotheses to be tested?		correlations		
iesieu:		applied, but		
		distribution of		
		scores or mean		
		(SD) not		
		presented		
	1	<u> </u>	ı	<u>. </u>

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
✓	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality E. Responsiveness n=27

E. Responsiveness	n=27			
Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of		
percentage of	missing items	missing items		
missing items	described	NOT described		
given?				
Was there a	Described how	Not described	Not clear how	
description of how	missing items	but it can be	missing items were	
missing items	were handled	deduced how	handled	
were handled?		missing items		
		were handled		
Was the sample	Adequate sample	Good sample	Moderate sample	Small sample size
size included in	size (≥100)	size (50-99)	size (30-49)	(<30)
the analysis				
adequate?				
Was a longitudinal	Longitudinal			No longitudinal
design with at	design used			design used
least two				
measurement				
used?				
Was the time	Time interval			Time interval
interval stated?	adequately			NOT described
	described			
If anything	Anything that	Assumable what	Unclear or NOT	If anything
occurred in the	occurred during	occurred during	described what	occurred in the
interim period	the interim	the interim	occurred during	interim period
(e.g. intervention,	period (e.g.	period	the interim period	(e.g. intervention,
other relevant	treatment)			other relevant
events), was it	adequately			events), was it
adequately	described			adequately
described?				described?
Was a proportion	Part of the	NO evidence	Unclear if part of	Patients were
of the patients	patients were	provided, but	the patients were	NOT changed
changed (i.e.	changed	assumable that	changed	
improvement or	(evidence	part of the		
deterioration)?	provided)	patients were		
		changed		
For constructs for	which a gold standa	ırd was not availal	ole:	
Were hypotheses	Hypotheses		Hypotheses vague	Unclear what was
about changes in	formulated a		or not formulated	expected
scores formulated	priori		but possible to	_

	T	T	T = = =	T
a priori (i.e. before			deduce what was	
data collection)?			expected	
Was the expected	Expected	Expected		
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
of the change		stated		
scores of HR-PRO				
instruments				
included in these				
hypotheses?				
Were the expected	Expected	Expected		
absolute or	magnitude of the	magnitude of		
relative <i>magnitude</i>	correlations or	the correlations		
of correlations or	differences stated	or differences		
mean differences	uniterences stated	NOT stated		
		NOT Stated		
of the change scores of HR-PRO				
instruments				
included in these				
hypotheses?				
Was an adequate	Adequate		Poor description of	NO description of
description	description of the		the constructs	the constructs
provided of the	constructs		measured by the	measured by the
comparator	measured by the		comparator	comparator
instrument(s)?	comparator		instrument(s)	instrument(s)
	instrument(s)			
Were the	Adequate	Adequate	Some information	NO information
measurement	measurement	measurement	on measurement	on the
properties of the	properties of the	properties of	properties (or a	measurement
comparator	comparator	the comparator	reference to a	properties of the
instrument(s)	instrument(s) in	instrument(s)	study on	comparator
adequately	a population	but not sure if	measurement	instrument(s)
described?	similar to the	these apply to	properties) of the	
	study population	the study	comparator	
		population	instrument(s) in	
		1 1	any study	
			population	
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
study:	execution of the		data presented on	study
I				Juay
	study		a comparison with	
			a comparison with an instrument that	
			a comparison with an instrument that measures another	
Word design and	study		a comparison with an instrument that measures another construct)	Statistical
Were design and	study Statistical		a comparison with an instrument that measures another construct) Statistical methods	Statistical methods applied
statistical methods	Statistical methods applied		a comparison with an instrument that measures another construct) Statistical methods applied NOT	methods applied
statistical methods adequate for the	study Statistical		a comparison with an instrument that measures another construct) Statistical methods	
statistical methods adequate for the hypotheses to be	Statistical methods applied		a comparison with an instrument that measures another construct) Statistical methods applied NOT	methods applied
statistical methods adequate for the hypotheses to be tested?	Statistical methods applied		a comparison with an instrument that measures another construct) Statistical methods applied NOT	methods applied

Can the criterion	Criterion used	No evidence	Unclear whether	Criterion used
for change be	can be	provided, but	the criterion used	can NOT be
considered as a	considered an	assumable that	can be considered	considered an
reasonable gold	adequate 'gold	the criterion	an adequate 'gold	adequate 'gold
standard?	standard'	used can be	standard'	standard'
	(evidence	considered an	o darrada d	o varraur a
	provided)	adequate 'gold		
	Province	standard'		
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study	execution of the
	execution of the			study
	study			
for continuous	Correlations or			Correlations or
scores: Were	Area under the			AUC NOT
correlations	ROC Curve (AUC)			calculated
between change	calculated			
scores, or the area				
under the				
Receiver Operator				
Curve (ROC) curve				
calculated?				
for dichotomous	Sensitivity and			Sensitivity and
scales: Were	specificity			specificity NOT
sensitivity and	calculated			calculated
specificity				
(changed versus				
not changed)				
determined?				

Hurd/Am J Sports Med/2008

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
✓	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality D. Criterion/Predictive Validity (Predictive in this case) N= 345

D. 0110011011/1100	D. Criterion, i redictive variates (i redictive in this case) N = 315					
Design	Excellent	Good	Fair	Poor		
requirements						
Was the	Percentage of	Percentage of	NA	NA		
percentage of	missing items	missing items NOT				
missing items	described	described				
given? none						
Was there a	Described how	Not described but it	Not clear how			
description of	missing items were	can be deduced how	missing items			

1	1 11 1	,	1 11 1	
how missing	handled	missing items were	were handled	
items were		handled- did not		
handled? none		use controls in the		
		final linear		
*** .1		regression model	27 2	a 11 1
Was the sample	Adequate sample	Good sample size	Moderate sample	Small sample
size included in	size (≥100)	(50-99)	size (30-49)	size (<30)
the analysis				
adequate?				
Can the criterion	Criterion used can	No evidence	Unclear whether	Criterion used
used or	be considered an	provided, but	the criterion used	can NOT be
employed be	adequate 'gold	assumable that the	can be considered	considered an
considered as a	standard'	criterion used can	an adequate 'gold	adequate 'gold
reasonable 'gold	(evidence	be considered an	standard'	standard'
standard'?	provided)	adequate 'gold		
		standard'		
Were there any	No other important	Other minor	Other important	Were there
important flaws	methodological	methodological	methodological	any important
in the design or	flaws in the design	flaws in the design	flaws in the	flaws in the
methods of the	or execution of the	or execution of the	design or	design or
study?	study	study	execution of the	methods of the
			study	study?
for continuous	Correlations or			Correlations
scores: Were	AUC calculated -			or AUC NOT
correlations, or	since it was			calculated
the area under	predictive validity,			
the receiver	a hierarchical			
operating curve	regression method			
calculated?	was used			
for dichotomous	Sensitivity and			Sensitivity and
scores: Were	specificity			specificity
sensitivity and	calculated			NOT
specificity				calculated
determined?				

Koutras/Int J Sports Med/2009

STEP1: Evaluated measurement properties in the article

31 Li 1. Evaluateu incasul cincii	31 Li 1. L'aiuateu measurement properties in the article					
Check if present	Property	Location				
	Reliability	Box A				
	Agreement/Measurement Error	Box B				
	Hypothesis Testing / Construct validity	Box C				
	Criterion/Predictive Validity	Box D				
✓	Responsiveness	Box E				

Step~2.~Determining~if~a~study~meets~the~standards~for~good~methodological~quality

E. Responsiveness n=20

Design	Excellent	Good	Fair	Poor	
requirements					
Was the	Percentage of	Percentage of			
percentage of	missing items	missing items			

· · · · ·		NOTE 1 1 1		<u> </u>
missing items given?	described	NOT described		
Was there a	Described how	Not described	Not clear how	
description of how	missing items	but it can be	missing items were	
missing items	were handled	deduced how	handled	
were handled?		missing items		
		were handled		
Was the sample	Adequate sample	Good sample	Moderate sample	Small sample size
size included in	size (≥100)	size (50-99)	size (30-49)	(<30)
the analysis				
adequate?				
Was a longitudinal	Longitudinal			No longitudinal
design with at	design used			design used
least two				
measurement				
used?				
Was the time	Time interval			Time interval
interval stated?	adequately			NOT described
moor var statear.	described			1.01 400011004
If anything	Anything that	Assumable what	Unclear or NOT	If anything
occurred in the	occurred during	occurred during	described what	occurred in the
interim period	the interim	the interim	occurred during	interim period
(e.g. intervention,	period (e.g.	period	the interim period	(e.g. intervention,
other relevant	treatment)	periou	the internii periou	other relevant
events), was it	adequately			events), was it
	described			
adequately described?	described			adequately described?
	Dont - Calo	NO and days as	II	
Was a proportion	Part of the	NO evidence	Unclear if part of	Patients were
of the patients	patients were	provided, but	the patients were	NOT changed
changed (i.e.	changed	assumable that	changed	
improvement or	(evidence	part of the		
deterioration)?	provided)	patients were		
For constructs for	 which a gold standa	changed	alo:	
roi constitucts for	willen a golu Stanua	ii u was iiot avaiiai	oie:	
Were hypotheses	Hypotheses		Hypotheses vague	Unclear what was
about changes in	formulated a		or not formulated	expected
scores formulated	priori		but possible to	
a priori (i.e. before			deduce what was	
data collection)?			expected	
Was the expected	Expected	Expected	-	
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
of the change		stated		
scores of HR-PRO				
instruments				
included in these				
hypotheses?				
Were the expected	Expected	Expected		
absolute or	magnitude of the	magnitude of		
relative magnitude	correlations or	the correlations		
of correlations or	differences stated	or differences		
mean differences		NOT stated		
	I .	1.01 Stated	<u> </u>	<u> </u>

	T	T		T
of the change				
scores of HR-PRO				
instruments				
included in these				
hypotheses?				
Was an adequate	Adequate		Poor description of	NO description of
description	description of the		the constructs	the constructs
provided of the	constructs		measured by the	measured by the
comparator	measured by the		comparator	comparator
instrument(s)?	comparator		instrument(s)	instrument(s)
	instrument(s)			
Were the	Adequate	Adequate	Some information	NO information
measurement	measurement	measurement	on measurement	on the
properties of the	properties of the	properties of	properties (or a	measurement
comparator	comparator	the comparator	reference to a	properties of the
instrument(s)	instrument(s) in	instrument(s)	study on	comparator
adequately	a population	but not sure if	measurement	instrument(s)
described?	similar to the	these apply to	properties) of the	
	study population	the study	comparator	
		population	instrument(s) in	
			any study	
			population	
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
	execution of the		data presented on	study
	study		a comparison with	-
			an instrument that	
			measures another	
			construct)	
Were design and	Statistical		Statistical methods	Statistical
statistical methods	methods applied		applied NOT	methods applied
adequate for the	appropriate		optimal	NOT appropriate
hypotheses to be			•	11 1
tested?				
For constructs for	which a gold standa	rd was available:	1	1
Can the criterion	Criterion used	No evidence	Unclear whether	Criterion used
for change be	can be	provided, but	the criterion used	can NOT be
considered as a	considered an	assumable that	can be considered	considered an
reasonable gold	adequate 'gold	the criterion	an adequate 'gold	adequate 'gold
standard?	standard'	used can be	standard'	standard'
	(evidence	considered an		
	provided)	adequate 'gold		
	F. 5	standard'		
Were there any	No other	- Juliuui u	Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study	execution of the
July!	execution of the		July	study
	study			Study
ļ			<u> </u>	<u> </u>
for continuous	Correlations or			Correlations or

scores: Were	Area under the		AUC NOT calculated
	ROC Curve (AUC)		calculated
between change	calculated		
scores, or the area			
under the			
Receiver Operator			
Curve (ROC) curve			
calculated?			
for dichotomous	Sensitivity and		Sensitivity and
scales: Were	specificity		specificity NOT
sensitivity and	calculated		calculated
specificity			
(changed versus			
not changed)			
determined?			

Myer/JOSPT/2011

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity N= 18

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was	Unclear what was expected

_	T	T		
formulated a			expected	
priori (i.e. before				
data collection)?				
Was the expected	Expected	Expected		
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
included in the		stated –		
hypotheses?				
Was the expected	Expected	Expected		
absolute or	magnitude of the	magnitude of the		
relative	correlations or	correlations or		
magnitude of	differences stated	differences NOT		
correlations or	anner ences stated	stated		
		Stateu		
mean differences				
included in the				
hypotheses?				
for convergent	Adequate	Adequate	Poor description of	NO description
validity: Was an	description of the	description of	the constructs	of the constructs
adequate	constructs	most of the	measured by the	measured by the
description	measured by the	constructs	comparator	comparator
provided of the	comparator	measured by the	instrument(s)	instrument(s)
comparator	instrument(s)	comparator		moti unicite(3)
_	mon ument(s)	_		
instrument(s)?	A 1 .	instrument(s)	C : C ::	N · C · ·
for convergent	Adequate	Adequate	Some information	No information
validity: Were the	measurement	measurement	on measurement	on the
measurement	properties of the	properties of the	properties (or a	measurement
properties of the	comparator	comparator	reference to a study	properties of the
comparator	instrument(s) in a	instrument(s)	on measurement	comparator
instrument(s)	population	but not sure if	properties) of the	instrument(s)
adequately	similar to the	these apply to	comparator	
described?	study population	the study	instrument(s) in	
described.	study population	the study	moti amendo) m	
		nonulation		
		population	any study	
More there are	No othor	population	any study population	Othersimmentant
Were there any	No other	population	any study population Other minor	Other important
important flaws	important	population	any study population Other minor methodological	methodological
		population	any study population Other minor	_
important flaws	important	population	any study population Other minor methodological	methodological flaws in the design or
important flaws in the design or	important methodological	population	any study population Other minor methodological flaws in the design	methodological flaws in the
important flaws in the design or methods of the	important methodological flaws in the	population	any study population Other minor methodological flaws in the design or execution of the study (e.g. only	methodological flaws in the design or
important flaws in the design or methods of the	important methodological flaws in the design or execution of the	population	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a	methodological flaws in the design or execution of the
important flaws in the design or methods of the	important methodological flaws in the design or	population	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with	methodological flaws in the design or execution of the
important flaws in the design or methods of the	important methodological flaws in the design or execution of the	population	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that	methodological flaws in the design or execution of the
important flaws in the design or methods of the	important methodological flaws in the design or execution of the	population	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another	methodological flaws in the design or execution of the
important flaws in the design or methods of the study?	important methodological flaws in the design or execution of the study		any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	methodological flaws in the design or execution of the study
important flaws in the design or methods of the study?	important methodological flaws in the design or execution of the study Statistical	Assumable that	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods	methodological flaws in the design or execution of the study
important flaws in the design or methods of the study? Were design and statistical	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods	important methodological flaws in the design or execution of the study Statistical	Assumable that statistical methods were	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods	methodological flaws in the design or execution of the study
important flaws in the design or methods of the study? Were design and statistical	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical methods were	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods adequate for the	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical methods were appropriate, e.g.	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods adequate for the hypotheses to be	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical methods were appropriate, e.g. Pearson correlations	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods adequate for the hypotheses to be	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods adequate for the hypotheses to be	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied
important flaws in the design or methods of the study? Were design and statistical methods adequate for the hypotheses to be	important methodological flaws in the design or execution of the study Statistical methods applied	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but	any study population Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied NOT	methodological flaws in the design or execution of the study Statistical methods applied

presented	
-----------	--

Nagano/ Open Sports Medicine Journal/2010

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
V	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
V	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability n=14 subjects

A. Renability n=14 subjects		1	1	
Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing	Percentage of	Percentage of		
subjects given?	missing	missing		
Not addressed	subjects	subjects NOT		
	described	described		
Was there a description of	Described how	Not described	Not clear how	
how missing subjects were	missing	but it can be	missing	
handled?	subjects were	deduced how	subjects were	
	handled	missing	handled	
		subjects were		
		handled		
Was the sample size included	Adequate	Good sample	Moderate	Small sample
in the analysis adequate?	sample size	size (50-99)	sample size	size (<30)
	(≥100)		(30-49)	
Were at least two	At least two			Only one
measurements available?	measurements			measurement
Intra and inter intra only				
Were the administrations	Independent	Assumable	Doubtful	Measurements
independent? Intra	measurements	that the	whether the	NOT
		measurement	measurements	independent
		s were	were	
		independent	independent	
Was the time interval stated?	Time interval		Time interval	
Intra	stated		NOT stated	
Were patients stable in the	Patients were	Assumable	Unclear if	Patients were
interim period on the	stable	that patients	patients were	NOT stable
construct to be measured?	(evidence	were stable	stable	
Intra	provided)			
Was the time interval	Time interval		Doubtful	Time interval
appropriate?	appropriate		whether time	NOT
Intra			interval was	appropriate
			appropriate	
Were the test conditions	Test	Assumable	Unclear if test	Test
similar for both	conditions	that test	conditions	conditions
measurements? e.g. type of	were similar	conditions	were similar	were NOT
administration, environment,	(evidence	were similar		similar
instructions	provided)		_	
Were there any important	No other		Other minor	Other

flaws in the design or methods of the study? Intra	important methodologica l flaws in the design or execution of the study		methodologica l flaws in the design or execution of the study	important methodologica l flaws in the design or execution of the study
for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has	No ICC or Pearson or Spearman correlations calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

D. Criterion/Predictive Validity (Predictive in this case) N= 59 subjects; 114 legs

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled- did not use controls in the final linear regression model	Not clear how missing items were handled Not clear if there were missing items	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Can the criterion used or	Criterion used can be considered an	No evidence provided, but	Unclear whether the criterion used	Criterion used can NOT be

employed be considered as a reasonable 'gold standard'?	adequate 'gold standard' (evidence provided)	assumable that the criterion used can be considered an adequate 'gold standard'	can be considered an adequate 'gold standard'	considered an adequate 'gold standard'
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study	Were there any important flaws in the design or methods of the study?
for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated 1. correlation between video peak knee flexion and the anterior reach of the YBT 2. No correlation between YBT and peak knee valgus 3. since it was predictive validity, a stepwise regression method was used			Correlations or AUC NOT calculated
for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Noyes/Am J Sports Med/1991

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
V	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity $\,$ n=67

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA

Was there a description of how missing items were handled? none Was the sample size included in the analysis adequate?	Described how missing items were handled Adequate sample size (≥100 per analysis)	Not described but it can be deduced how missing items were handled Good sample size (50-99 per analysis)	Not clear how missing items were handled Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Were there any important flaws	No other important methodological		Other minor methodological	Other important methodological

in the design or methods of the study?	flaws in the design or execution of the study		flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	flaws in the design or execution of the study
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate- single and then multiple linear regression for limb symmetry as dependent variable and only significant correlation was a low one (.49) with isokinetic testing at 60 degrees/second	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

D. Criterion/Predictive Validity N= 67 subjects

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given? none	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled- did not use controls in the final linear regression model	Not clear how missing items were handled Not clear if there were missing items	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study	Were there any important flaws in the design or methods of the study?

for continuous scores: Were	Correlations or AUC calculated		Correlations or AUC NOT
correlations, or	11000011000		calculated
the area under			
the receiver			
operating curve			
calculated?			
for dichotomous	Sensitivity and		Sensitivity and
scores: Were	specificity		specificity NOT
sensitivity and	calculated		calculated
specificity	Single hop 52;97;		
determined?	LR+ 17.3		
	Timed hop		
	49;92;LR+ 6.125		

Purdam/PT in Sport/2003

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
✓	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
✓	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability n=46 subjects

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing	Percentage of	Percentage of		
subjects given?	missing	missing		
	subjects	subjects NOT		
	described	described		
Was there a description of	Described how	Not described	Not clear how	
how missing subjects were	missing	but it can be	missing	
handled?	subjects were	deduced how	subjects were	
	handled	missing	handled	
		subjects were		
		handled		
Was the sample size included	Adequate	Good sample	Moderate	Small sample
in the analysis adequate?	sample size	size (50-99)	sample size	size (<30)
	(≥100)		(30-49)	
Were at least two	At least two			Only one
measurements available?	measurements			measurement
Intra and inter intra only				
Were the administrations	Independent	Assumable	Doubtful	Measurements
independent? Intra	measurements	that the	whether the	NOT
		measurement	measurements	independent
		s were	were	
		independent	independent	
Was the time interval stated?	Time interval		Time interval	
Intra	stated		NOT stated	
Were patients stable in the	Patients were	Assumable	Unclear if	Patients were
interim period on the	stable	that patients	patients were	NOT stable

construct to be measured? Intra	(evidence provided)	were stable	stable	
Was the time interval appropriate? Intra	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
Were there any important flaws in the design or methods of the study? Intra	No other important methodologica I flaws in the design or execution of the study		Other minor methodologica I flaws in the design or execution of the study	Other important methodologica l flaws in the design or execution of the study
for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated – actually TEM calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described	With TEM- not sure how to answer this	

E. Responsiveness n=56, then 50 eligible, then 15/17 control and 13/15 case subjects completing 3 testing sessions + 18 more added for reliability study

Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of		
percentage of	missing items	missing items		

	T	T	T	T
missing items given?	described	NOT described		
Was there a	Described how	Not described	Not clear how	
description of how	missing items	but it can be	missing items were	
_	were handled	deduced how		
missing items were handled?	were nandled		handled	
were nandled?		missing items		
*** .1		were handled	27 2	g 11 1 1
Was the sample	Adequate sample	Good sample	Moderate sample	Small sample size
size included in	size (≥100)	size (50-99)	size (30-49)	(<30)
the analysis				N=28
adequate?				
Was a longitudinal	Longitudinal			No longitudinal
design with at	design used			design used
least two				
measurement				
used?				
Was the time	Time interval			Time interval
interval stated?	adequately			NOT described
	described			
If anything	Anything that	Assumable what	Unclear or NOT	If anything
occurred in the	occurred during	occurred during	described what	occurred in the
interim period	the interim	the interim	occurred during	interim period
(e.g. intervention,	period (e.g.	period	the interim period	(e.g. intervention,
other relevant	treatment)	periou	the interim period	other relevant
events), was it	adequately			events), was it
adequately	described			adequately
described?	uescribeu			described?
	Part of the	NO evidence	Unclear if part of	Patients were
Was a proportion				
of the patients	patients were	provided, but assumable that	the patients were	NOT changed
changed (i.e.	changed		changed	
improvement or	(evidence	part of the		
deterioration)?	provided)	patients were		
For constructs for	 which a gold standa	changed	<u> </u>	
For constructs for	willen a golu stanua	ii u was not avanai	oie.	
Were hypotheses	Hypotheses		Hypotheses vague	Unclear what was
about changes in	formulated a		or not formulated	expected
scores formulated	priori		but possible to	
a priori (i.e. before			deduce what was	
data collection)?			expected	
Was the expected	Expected	Expected	•	
direction of	direction of the	direction of the		
correlations or	correlations or	correlations or		
mean differences	differences stated	differences NOT		
of the change	3	stated		
scores of HR-PRO				
instruments				
included in these				
hypotheses?				
Were the expected	Expected	Expected		
absolute or	magnitude of the	magnitude of		
relative magnitude	correlations or	the correlations		
of correlations or	correlations or	the correlations	1	
	differences stated	or differences		
mean differences	differences stated	or differences NOT stated		

	T	T	T	T
of the change				
scores of HR-PRO				
instruments				
included in these				
hypotheses?				
Was an adequate	Adequate		Poor description of	NO description of
description	description of the		the constructs	the constructs
provided of the	constructs		measured by the	measured by the
comparator	measured by the		comparator	comparator
instrument(s)?	comparator		instrument(s)	instrument(s)
	instrument(s)			
Were the	Adequate	Adequate	Some information	NO information
measurement	measurement	measurement	on measurement	on the
properties of the	properties of the	properties of	properties (or a	measurement
comparator	comparator	the comparator	reference to a	properties of the
•	instrument(s) in	_		
instrument(s)		instrument(s)	study on	comparator
adequately	a population	Dat not bar on	measurement	instrument(s)
described?	similar to the	these apply to	properties) of the	
	study population	the study	comparator	
		population	instrument(s) in	
			any study	
			population	
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
	execution of the		data presented on	study
	study		a comparison with	
	-		an instrument that	
			measures another	
			construct)	
Were design and	Statistical		Statistical methods	Statistical
statistical methods	methods applied		applied NOT	methods applied
adequate for the	appropriate		optimal	NOT appropriate
hypotheses to be	арргоргасс		optima	1101 appropriate
tested?				
	which a gold stands	rd was available.		
Can the criterion	which a gold standa Criterion used	No evidence	Unclear whether	Criterion used
	can be		the criterion used	can NOT be
for change be		provided, but		
considered as a	considered an	assumable that	can be considered	considered an
reasonable gold	adequate 'gold	the criterion	an adequate 'gold	adequate 'gold
standard?	standard'	used can be	standard'	standard'
	(evidence	considered an		
	provided)	adequate 'gold		
		standard'		
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study	execution of the
-			_	study
	execution of the			study
	study			study

scores: Were	Area under the	pain scale is	AUC NOT
correlations	ROC Curve (AUC)	interval data	calculated
between change	calculated		
scores, or the area			
under the			
Receiver Operator			
Curve (ROC) curve			
calculated?			
for dichotomous	Sensitivity and		Sensitivity and
scales: Were	specificity		specificity NOT
sensitivity and	calculated		calculated
specificity			
(changed versus			
not changed)			
determined?			

Ross/Knee Sport Taumatol/2002

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
✓	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality

A. Reliability n=50 subjects but only 10 for reliability

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing subjects given?	Percentage of missing subjects described	Percentage of missing subjects NOT described		
Was there a description of how missing subjects were handled?	Described how missing subjects were handled	Not described but it can be deduced how missing subjects were handled	Not clear how missing subjects were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Were at least two measurements available? Intra and inter intra only	At least two measurements			Only one measurement
Were the administrations independent? Intra	Independent measurements	Assumable that the measurement s were independent	Doubtful whether the measurements were independent	Measurements NOT independent
Was the time interval stated? Intra	Time interval stated		Time interval NOT stated	
Were patients stable in the interim period on the	Patients were stable	Assumable that patients	Unclear if patients were	Patients were NOT stable

construct to be measured?	(evidence	were stable	stable	
Intra Was the time interval appropriate? Intra	rime interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
Were there any important flaws in the design or methods of the study? Intra	No other important methodologica I flaws in the design or execution of the study		Other minor methodologica I flaws in the design or execution of the study	Other important methodologica I flaws in the design or execution of the study
for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has	No ICC or Pearson or Spearman correlations calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated - actually TEM calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described	With TEM- not sure how to answer this	

C. Hypothesis Testing / Construct Validity

Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of	NA	NA
percentage of	missing items	missing items		
missing items	described	NOT described		

given?				
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-functional tests correlate positively with self-report	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population- KOS, ADLS, SAS used but only the	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study

important flaws in the design or methodological flaws in the design or execution of the study? Were design and statistical methods applied methods adequate for the hypotheses to be tested? important methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Assumable that statistical methods applied appropriate Methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct) Statistical methods applied appropriate Assumable that statistical methods applied NOT optimal NOT appropriate NOT appropriate NOT optimal		rel/valid of KOS cited			
statistical methods applied appropriate methods applied appropriate methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean methods applied statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean	important flaws in the design or methods of the	important methodological flaws in the design or execution of the		methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another	design or execution of the
presented	statistical methods adequate for the hypotheses to be	methods applied	statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not	applied NOT	Statistical methods applied NOT appropriate

Ross/Knee Sport Taumatol/2010

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

C. Hypothesis 1C.	sung / construct vant	11ty 11-10		
Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of	NA	NA
percentage of	missing items	missing items		
missing items	described	NOT described		
given?				
O				
Was there a	Described how	Not described	Not clear how	
description of	missing items were	but it can be	missing items were	
how missing	handled	deduced how	handled	
items were		missing items		
handled?		were handled		
Was the sample	Adequate sample	Good sample	Moderate sample	Small sample
size included in	size (≥100 per	size (50-99 per	size (30-49 per	size (<30 per
the analysis	analysis)	analysis)	analysis)	analysis)
adequate?	J J		J J	, ,
	1		1	

Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated- FABQ correlate positively with self-report	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population- KOS, SAS, ADLS used and the reliability/validity of KOS cited	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an	Other important methodological flaws in the design or execution of the study –

			instrument that measures another construct)- they adapted FABQ to knee so metrics of FABQ not known in knee	
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Svensson/Knee Sport Taumatol/2006

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
✓	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality E. Responsiveness n=59

E. Responsiveness	11-37			
Design	Excellent	Good	Fair	Poor
requirements				
Was the	Percentage of	Percentage of		
percentage of	missing items	missing items		
missing items	described	NOT described		
given?				
Was there a	Described how	Not described	Not clear how	
description of how	missing items	but it can be	missing items were	
missing items	were handled	deduced how	handled	
were handled?		missing items		
		were handled		
Was the sample	Adequate sample	Good sample	Moderate sample	Small sample size
size included in	size (≥100)	size (50-99)	size (30-49)	(<30)
the analysis				
adequate?				
Was a longitudinal	Longitudinal			No longitudinal
design with at	design used			design used
least two				
measurement				
used?				
Was the time	Time interval			Time interval
interval stated?	adequately			NOT described
	described			

If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described? Was a proportion of the patients changed (i.e. improvement or deterioration)?	Anything that occurred during the interim period (e.g. treatment) adequately described Part of the patients were changed (evidence provided)-actually, all were improved at 2	Assumable what occurred during the interim period NO evidence provided, but assumable that part of the patients were changed	Unclear or NOT described what occurred during the interim period Unclear if part of the patients were changed	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described? Patients were NOT changed
For constructs for	year follow-up which a gold standa	rd was not availal	l ole:	
Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of th√e correlations or differences NOT stated		
Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
Were the measurement properties of the comparator instrument(s) adequately	Adequate measurement properties of the comparator instrument(s) in a population	Adequate measurement properties of the comparator instrument(s) but not sure if	Some information on measurement properties (or a reference to a study on measurement	NO information on the measurement properties of the comparator instrument(s)

	T	1	1	1
described?	similar to the	these apply to	properties) of the	
	study population	the study	comparator	
		population	instrument(s) in	
			any study	
			population	
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study (e.g. only	execution of the
	execution of the		data presented on	study
	study		a comparison with	
			an instrument that	
			measures another	
			construct)	
Were design and	Statistical		Statistical methods	Statistical
statistical methods	methods applied		applied NOT	methods applied
adequate for the	appropriate		optimal	NOT appropriate
hypotheses to be				
tested?				
For constructs for	which a gold standa	ırd was available:		
Can the criterion	Criterion used	No evidence	Unclear whether	Criterion used
for change be	can be	provided, but	the criterion used	can NOT be
considered as a	considered an	assumable that	can be considered	considered an
reasonable gold	adequate 'gold	the criterion	an adequate 'gold	adequate 'gold
standard?	standard'	used can be	standard'	standard'
	(evidence	considered an		
	provided)	adequate 'gold		
		standard'		
Were there any	No other		Other minor	Other important
important flaws in	important		methodological	methodological
the design or	methodological		flaws in the design	flaws in the
methods of the	flaws in the		or execution of the	design or
study?	design or		study	execution of the
	execution of the			study
	study			
for continuous	Correlations or			Correlations or
scores: Were	Area under the			AUC NOT
correlations	ROC Curve (AUC)			calculated
between change	calculated			
scores, or the area				
under the				
Receiver Operator				
Curve (ROC) curve				
calculated?				
for dichotomous	Sensitivity and			Sensitivity and
scales: Were	specificity			specificity NOT
sensitivity and	calculated			calculated
specificity				
(changed versus				
not changed)				
determined?				

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

	Excellent	Good	Fair	Poor
Design	Excellent	G000	rdII	roor
requirements Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled? Was the sample size included in the analysis adequate?	Described how missing items were handled Adequate sample size (≥100 per analysis)	Not described but it can be deduced how missing items were handled Good sample size (50-99 per analysis)	Not clear how missing items were handled Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate	Adequate description of the constructs	Adequate description of most of the	Poor description of the constructs measured by the	NO description of the constructs measured by the

description provided of the	measured by the comparator	constructs measured by the	comparator instrument(s)	comparator instrument(s)
comparator instrument(s)?	instrument(s)	comparator instrument(s)		
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population- KOS, SAS, ADLS used and the rel/valid of KOS cited	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Witvrouw/Scand J Med Sci Sports/2002 STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
1	Reliability	Box A
	Agreement/Measurement Error	Box B
*	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability n=25 for reliability

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing	Percentage of	Percentage of		
subjects given?	missing	missing		
none	subjects	subjects NOT		

	described	described		
Was there a description of how missing subjects were handled?	Described how missing subjects were handled	Not described but it can be deduced how missing subjects were handled	Not clear how missing subjects were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Were at least two measurements available? Intra and inter intra only	At least two measurements			Only one measurement
Were the administrations independent? Intra	Independent measurements	Assumable that the measurement s were independent	Doubtful whether the measurements were independent	Measurements NOT independent
Was the time interval stated? Intra	Time interval stated		Time interval NOT stated	
Were patients stable in the interim period on the construct to be measured? Intra	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
Was the time interval appropriate? Intra	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
Were there any important flaws in the design or methods of the study? Intra	No other important methodologica I flaws in the design or execution of the study		Other minor methodologica I flaws in the design or execution of the study	Other important methodologica I flaws in the design or execution of the study
for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic	No ICC or Pearson or Spearman correlations calculated

		change has occurred	change has occurred	
For dichotomous/nominal/ordina	Kappa calculated			Only percentage agreement
calculated? for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	calculated Only percentage agreement calculated actually TEM calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described	With TEM- not sure how to answer this	

C. Hypothesis Testing / Construct Validity n=30

di 11) poulledio 1 es	0/	· · · · · · · · · · · · · · · · · · ·		
Design	Excellent	Good	Fair	Poor
requirements				
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled- all in final analysis	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		

mean differences included in the hypotheses?				
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population-	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Zouita/Annals of Phys & Rehab Medicine/2009

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity n=46

Design	Excellent	Good	Fair	Poor
requirements Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled- all in final analysis	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s) – same test in a healthy population	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator	Adequate measurement properties of the comparator instrument(s) in a	Adequate measurement properties of the comparator instrument(s)	Some information on measurement properties (or a reference to a study on	No information on the measurement properties of the comparator

instrument(s) adequately described?	population similar to the study population-	but not sure if these apply to the study population	measurement properties) of the comparator instrument(s) in any study population	instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Barber/CORR/1990

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity n=93

Design Excellent Good Fair Poor requirements NA NA Was the Percentage of Percentage of percentage of missing items missing items missing items described NOT described given? none Was there a Described how Not described Not clear how description of missing items but it can be missing items were how missing were handled- all deduced how handled items were in final analysis missing items handled? were handled

Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s) same test in a healthy	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	population Adequate measurement properties of the comparator instrument(s) in a population similar to the study population-	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that	Other important methodological flaws in the design or execution of the study –

			measures another construct)	
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Brosky/JOSPT/1999

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
✓	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability n=15 for reliability

Evallant	Cood	Fain	Daam
		rair	Poor
_	O .		
•	0		
described	described		
Described how	Not described	Not clear how	
missing	but it can be	missing	
subjects were	deduced how	subjects were	
handled	missing	handled	
	subjects were		
	handled		
Adequate	Good sample	Moderate	Small sample
	•	sample size	size (<30)
_		_	n=15
At least two			Only one
measurements			measurement
Independent	Assumable	Doubtful	Measurements
measurements	that the	whether the	NOT
	measurement	measurements	independent
	s were	were	•
	independent	independent	
Time interval	-	Time interval	
stated		NOT stated	
Patients were	Assumable	Unclear if	Patients were
stable	that patients	patients were	NOT stable
(evidence	were stable	stable	
provided)			
	Excellent Percentage of missing subjects described Described how missing subjects were handled Adequate sample size (≥100) At least two measurements Independent measurements Time interval stated Patients were stable (evidence	Excellent Percentage of missing subjects described Described how missing subjects were handled Adequate sample size (≥100) At least two measurements Independent measurements Independent measurements Independent swere independent Time interval stated Patients were stable (evidence Percentage of missing subjects NOT described but it can be deduced how missing subjects were handled Adequate size (50-99) Assumable that the measurement swere independent Time interval stated Patients were stable (evidence Percentage of missing subjects NOT described but it can be deduced how missing subjects were handled Adequate size (50-99) Assumable that patients were stable	ExcellentGoodFairPercentage of missing subjects describedPercentage of missing subjects NOT describedNot clear how missing subjects were handledDescribed how missing subjects were handledNot clear how missing subjects were handledAdequate sample size (≥100)Good sample size (50-99)Moderate sample size (30-49)At least two measurementsAssumable that the measurement s were independentDoubtful whether the measurements were independentTime interval statedTime interval NOT statedPatients were stable (evidenceAssumable that patients were stableUnclear if patients were stable

Was the time interval appropriate? Intra Were the test conditions similar for both measurements? e.g. type of administration, environment,	Time interval appropriate Test conditions were similar (evidence	Assumable that test conditions were similar	Doubtful whether time interval was appropriate Unclear if test conditions were similar	Time interval NOT appropriate Test conditions were NOT similar
instructions Were there any important flaws in the design or methods of the study? Intra	provided) No other important methodologica I flaws in the design or execution of the study		Other minor methodologica I flaws in the design or execution of the study	Other important methodologica I flaws in the design or execution of the study
for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated actually TEM calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described	With TEM- not sure how to answer this	

Grindem/Am J Sports Med/2011

51Li 1. Evaluated incasul ement properties in the article					
Check if present	Property	Location			
	Reliability	Box A			
	Agreement/Measurement Error	Box B			
	Hypothesis Testing / Construct validity	Box C			

y	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality D. Criterion/predictive Validity n=91 but 10 were lost in follow-up so n=81

Design	Excellent	Good	Fair	Poor
requirements				
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled-	Not clear how missing items were handled Not clear if there were missing items	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study	Were there any important flaws in the design or methods of the study?
for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated- Single hop is only test correlated with self-reported IKDC function at 1 year			Correlations or AUC NOT calculated
for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated Single hop 71;71; LR+ 2.52 LR- 0.40			Sensitivity and specificity NOT calculated

Logerstedt/Am J Sports Med/2012

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
	Hypothesis Testing / Construct	Box C

	validity	
✓	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality D. Criterion/predictive Validity pre-op to 1 year n=79; 6 mos to 1 year n=85

Design	Excellent	Good	Fair	Poor
requirements	2	dood		1 001
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled-	Not clear how missing items were handled Not clear if there were missing items	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study	Were there any important flaws in the design or methods of the study?
for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated- Single hop is only test correlated with self-reported IKDC function at 1 year			Correlations or AUC NOT calculated
for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated 6m timed hop (87.7% LSI) 53;90; LR+ 5.14 LR- 0.40 Crossover hop (94.9% LSI) 88;47 LR- 0.25			Sensitivity and specificity NOT calculated

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Design	Excellent	Good	Fair	Poor
requirements	2	dood		
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled? none	Described how missing items were handled- all in final analysis	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		

hypotheses?				
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s) same test in a healthy population	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population-	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s) NOTE: First time tested in this study
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Wilk/JOSPT/1994

51Li 1. Evaluated incasurement properties in the article					
Check if present	Property	Location			
	Reliability	Box A			
	Agreement/Measurement Error	Box B			
1	Hypothesis Testing / Construct validity	Box C			
	Criterion/Predictive Validity	Box D			
	Responsiveness	Box E			

	ting / Construct Val	•	1	1
Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled- all in final analysis	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator	Adequate measurement properties of the comparator instrument(s) in a	Adequate measurement properties of the comparator instrument(s)	Some information on measurement properties (or a reference to a study on measurement	No information on the measurement properties of the comparator

instrument(s) adequately described?	population similar to the study population-	but not sure if these apply to the study population	properties) of the comparator instrument(s) in any study population	instrument(s)
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Augustsson/ Knee Surg Sports Traumatol Arthrosc/2004

STEP1: Evaluated measurement properties in the article

51 Li 1: Li ui uuteu meusui ement pi opei ties in the ui tiele					
Check if present	Property	Location			
	Reliability	Box A			
	Agreement/Measurement Error	Box B			
1	Hypothesis Testing / Construct validity	Box C			
	Criterion /Predictive Validity	Box D			
	Responsiveness	Box E			

c. Hypothesis res	ting / constituct var	fulty II = 17		
Design	Excellent	Good	Fair	Poor
requirements				
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled- all in final analysis	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	

adequate description provided of the comparator instrument(s)? for convergent validity: Were the measurement properties of the comparator instrument(s) in a instrument(s) adequately described? Were there any important flaws in the design or methods of the study? No other important flaws in the design or methods of the study? Adequate measurement properties of the comparator instrument(s) in a population No other important methodological flaws in the design or execution of the study adequate measurement properties of the comparator instrument(s) in a population No other important methodological flaws in the design or execution of the study adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population No other important methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another measured by the comparator instrument(s) Adequate measurement properties of the comparator instrument(s) on measurement properties of the comparator instrument on measurement properties (or a reference to a study on measurement properties of the comparator instrument on measurement properties of the compa	Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Was the expected direction of the correlations or mean differences included in the hypotheses?	regarding correlations or mean differences formulated a priori (i.e. before	formulated a	of hypotheses formulate a	or not formulated but possible to deduce what was	
absolute or relative magnitude of the correlations or differences stated correlations or mean differences included in the hypotheses? for convergent validity: Was an adequate description of adequate of the comparator instrument(s)? for convergent validity: Was an adequate description of instrument(s)? for convergent validity: Were the comparator instrument(s) in any study population Were there any important flaws in the design or methodological flaws in the design or methodological flaws in the design or execution of the study? magnitude of the correlations or differences NOT stated Adequate description of the constructs measured by the comparator instrument(s) in any study population Mo other important methodological flaws in the design or execution of the study? magnitude of the correlations or differences NOT stated Adequate description of the constructs measured by the comparator instrument(s) adequate yellow the comparator instrument(s) in any study population No other important methodological flaws in the design or execution of the study? More there any important flaws in the design or execution of the study? More there any important flaws in the design or execution of the study? More there any important flaws in the design or execution of the study? More there any important flaws in the design or execution of the study e.g. only data presented on a comparison with an instrument that measures another	Was the expected direction of correlations or mean differences included in the	direction of the correlations or differences stated-	direction of the correlations or differences NOT		
validity: Was an adequate description of the constructs measured by the comparator instrument(s)? for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	magnitude of the correlations or differences stated	magnitude of the correlations or differences NOT stated		
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described? Were there any important flaws in the design or methods of the study? No other important methodological flaws in the study? Adequate measurement properties of the comparator instrument(s) in a population No other important methodological flaws in the design or execution of the study No other important methodological flaws in the design or execution of the study The study is a similar to the study is a study or execution of the study is a study or execution of the study The study is a similar to the design or execution of the study is a study or execution of the study is a study or execution of the study is a study or execution of the study or execution of the study is a study or execution of the study or execution or execution of the study or execution or execution of the study or execution or execu	validity: Was an adequate description provided of the comparator	description of the constructs measured by the comparator	description of most of the constructs measured by the comparator	the constructs measured by the comparator	of the constructs measured by the comparator
important flaws in the design or methods of the study? important methodological flaws in the design or execution of the study important methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another important methodological flaws in the design or execution of the study –	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately	measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study	on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study	on the measurement properties of the comparator instrument(s)
Were design and Statistical Assumable that Statistical methods Statistical	important flaws in the design or methods of the study?	important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	flaws in the design or execution of the study –

statistical	methods applied	statistical	applied NOT	methods applied
methods	appropriate	methods were	optimal-	NOT appropriate
adequate for the		appropriate, e.g.	correlation	
hypotheses to be		Pearson	coefficients would	
tested?		correlations	be more	
		applied, but	appropriate	
		distribution of		
		scores or mean		
		(SD) not		
		presented		

Jerre/ Scan J Med Sci Sports /2001

STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
	Reliability	Box A
	Agreement/Measurement Error	Box B
1	Hypothesis Testing / Construct validity	Box C
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality C. Hypothesis Testing / Construct Validity n= 275

Design Excellent Good Fair Poor requirements NA Was the Percentage of Percentage of NA percentage of missing items missing items missing items described NOT described given? none Not described Was there a Described how Not clear how description of missing items but it can be missing items were were handled- all how missing deduced how handled items were in final analysis missing items handled? were handled Good sample size Small sample Was the sample Adequate sample Moderate sample size included in (50-99 per size (30-49 per size (≥100 per size (<30 per the analysis analysis) analysis) analysis) analysis) adequate? Were hypotheses hypotheses Minimal number Hypotheses vague Unclear what or not formulated regarding formulated a of hypotheses was expected correlations or priori formulate a but possible to mean differences priori deduce what was formulated a expected priori (i.e. before data collection)? Was the expected Expected Expected direction of direction of the direction of the correlations or correlations or correlations or mean differences differences differences NOT

included in the	stated-	stated -		
hypotheses?				
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population-	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate

Vandermeulen/Physother Can/1995 STEP1: Evaluated measurement properties in the article

Check if present	Property	Location
✓	Reliability	Box A

	Agreement/Measurement Error	Box B
✓	Hypothesis Testing /	Box C
	Construct validity	
	Criterion/Predictive Validity	Box D
	Responsiveness	Box E

Step 2. Determining if a study meets the standards for good methodological quality A. Reliability n=46

Design requirements	Excellent	Good	Fair	Poor
Was the percentage of missing	Percentage of	Percentage of		
subjects given?	missing	missing		
none	subjects	subjects NOT		
	described	described		
Was there a description of	Described how	Not described	Not clear how	
how missing subjects were	missing	but it can be	missing	
handled?	subjects were handled	deduced how	subjects were handled	
none	nanuieu	missing subjects were	Halluleu	
		handled		
Was the sample size included	Adequate	Good sample	Moderate	Small sample
in the analysis adequate?	sample size	size (50-99)	sample size	size (<30)
	(≥100)		(30-49)	
Were at least two	At least two			Only one
measurements available?	measurements			measurement
Intra and inter intra only			D 1.63	
Were the administrations	Independent	Assumable	Doubtful	Measurements
independent? Intra	measurements	that the	whether the	NOT
		measurement s were	measurements were	independent
		independent	independent	
Was the time interval stated?	Time interval	таерепаен	Time interval	
Intra	stated		NOT stated	
Were patients stable in the	Patients were	Assumable	Unclear if	Patients were
interim period on the	stable	that patients	patients were	NOT stable
construct to be measured?	(evidence	were stable	stable	
Intra	provided)			
Was the time interval	Time interval		Doubtful	Time interval
appropriate?	appropriate		whether time	NOT
Intra			interval was	appropriate
More the test and division	Toot	Aggumelala	appropriate	Toot
Were the test conditions similar for both	Test conditions	Assumable that test	Unclear if test conditions	Test conditions
measurements? e.g. type of	were similar	conditions	were similar	were NOT
administration, environment,	(evidence	were similar	were similar	similar
instructions	provided)	were similar		Sillilai
Were there any important	No other		Other minor	Other
flaws in the design or	important		methodologica	important
methods of the study?	methodologica		l flaws in the	methodologica
Intra	l flaws in the		design or	l flaws in the
	design or		execution of	design or
	execution of		the study	execution of
	the study			the study
for continuous scores: Was an	ICC calculated	ICC calculated	Pearson or	No ICC or

intraclass correlation coefficient (ICC) calculated?	and model or formula of the ICC is described	but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has	Pearson or Spearman correlations calculated
For dichotomous/nominal/ordina l scores: Was kappa calculated?	Kappa calculated	333373	3334754	Only percentage agreement calculated
for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated actually TEM calculated
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described	With TEM- not sure how to answer this	

C. Hypothesis Testing / Construct Validity n= 46

Design	Excellent	Good	Fair	Poor
requirements				
Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	NA	NA
Was there a description of how missing items were handled?	Described how missing items were handled- all in final analysis	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
Were hypotheses regarding correlations or mean differences formulated a	hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected

_				
priori (i.e. before data collection)?				
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated-	Expected direction of the correlations or differences NOT stated –		
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population-	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study –
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal- correlation coefficients would be more appropriate	Statistical methods applied NOT appropriate