

1 **Title**

2 **Inter-rater reliability, internal consistency and common technique flaws**
3 **of the tuck jump assessment in elite female football players**

4
5 **Authors:** Lawrence Mayhew (l.mayhew@leedsbeckett.ac.uk) [1], Mark I.
6 Johnson (m.johnson@leedsbeckett.ac.uk) [1, 2], Peter Francis
7 (peter.francis@leedsbeckett.ac.uk) [1], Nicky Snowdon
8 (n.snowdon@shu.ac.uk) [3], Gareth Jones (g.j.jones@leedsbeckett.ac.uk) [1,
9 2]

10

11 **Author affiliations**

12 1. School of Clinical and Applied Sciences, Leeds Beckett University, Leeds,
13 U.K.

14 2. Leeds Pallium Research Group, Leeds, U.K.

15 3. Faculty of Health and Wellbeing, Sheffield Hallam University, Sheffield,
16 U.K.

17

18 **Corresponding author**

19 Gareth Jones, School of Applied and Clinical Sciences, Leeds Beckett
20 University, Leeds, LS1 3HE,

21 U.K., Telephone: +44 113 8123304, Fax: +44 113 2833124, e-mail:

22 G.J.Jones@leedsbeckett.ac.uk

23

24

25 **Article Type:** Original Article

26

27 **Running header:** Inter-rater Reliability of Tuck Jump Assessment

28

29 **Funding:** This study was did not receive external funding

30

31 **Ethical Approval:** Ethical approval was granted by Sheffield Hallam
32 University

33

34 **Contribution of Authors:**

- 35 • Study design: LM, GJ, NS
- 36 • Data collection: LM
- 37 • Data analysis: LM,GJ, MIJ
- 38 • Preparation of the manuscript: GJ, LM, MIJ, PF

39

40 **Competing Interests:** None

41

42 **Acknowledgments:** None

43

44 **Accessing Research Materials:** Underlying research materials related to
45 your paper (for example data, trial master file documents etc.) can be
46 accessed by contacting Mr Lawrence Mayhew

47

48 **Key words:** Tuck Jump Assessment; landing faults; female football;
49 reliability, internal consistency

50

51 **ABSTRACT**

52

53 Injury rates between elite female and male players are comparable, although
54 female players are more likely to sustain an anterior cruciate ligament (ACL)
55 injury. The common mechanism of ACL injury is non-contact trauma
56 sustained when landing from a jump. The Tuck Jump Assessment (TJA) is
57 commonly used in football to assess jump landing technique. The aims of
58 this study were to determine inter-rater agreement and internal consistency
59 of the TJA and to identify commonly occurring technique flaws. Sixty elite
60 female football players (mean (SD): age = 20.27 ± 3.44yrs) were video
61 recorded whilst undertaking the TJA and independently assessed by four
62 raters. Clinically acceptable levels of agreement were reached for 'Lower
63 extremity valgus at landing' $k = .83$ (95% CI, .72 – .93); 'Thighs do not reach
64 parallel' $k = .84$ (95% CI, .74 - .94); 'Thighs not equal side to side' $k = .86$
65 (95% CI, .75 - .96). The level of agreement for the composite score of all 10
66 criteria ranged from $k_w = .62$ (95% CI, .48 – .76) to $k_w = .80$ (95% CI, .70 –
67 .90) suggesting a 'fair-to-very good' level of inter-rater agreement. The most
68 common technique flaws were found for criterion 2 'Thighs do not reach
69 parallel' (N=147/665) and criterion 1 'Knee valgus on landing' (N=80/665).
70 However, internal consistency results suggest that the TJA is not
71 unidimensional. We suggest 'Knee valgus on landing' may have clinical utility
72 although further research is needed.

73

74

75

76 **INTRODUCTION**

77 The advent of professionalism and the exponential rise in the number of
78 UEFA registered female football players has corresponded with a significant
79 increase in the reported incidence of injury. The estimated incidence of injury
80 for female players is between 12.6 and 24.0 injuries per 1000 hours of match
81 play and between 1.2 and 7.0 injuries per 1000 hours of training (Giza,
82 Mithofer, Farrell, Zarins, & Gill, 2005; Le Gall, Carling, & Reilly, 2008; Nilstad,
83 Andersen, Bahr, Holme, & Steffen, 2014). Although injury rates between elite
84 female and male players are comparable (Hagglund, Walden, & Atroshi,
85 2009), female players are more likely to sustain an anterior cruciate ligament
86 (ACL) injury of the knee than their male counterparts (Walden, Hagglund,
87 Magnusson, & Ekstrand, 2011). ACL injury in female players is more likely to
88 occur at an earlier age (Renstrom et al., 2008) and a previous history of ACL
89 injury is considered to be a significant risk factor for reinjury (Faude, Junge,
90 Kindermann, & Dvorak, 2006). The most common mechanism of ACL injury
91 is an acute non-contact trauma sustained during rapid decelerating
92 movements, for example when landing from a jump (Walden et al., 2011).
93 Reduced neuromuscular control during landing may result in increased knee
94 valgus angles termed ligament dominance and increase the likelihood of an
95 individual sustaining an ACL injury (Hewett, Myer, Ford, et al., 2005).
96
97 Observational screening tools are commonly used to assess jump landing
98 tasks to identify faulty movement patterns. These screening tools include the
99 Landing Error Scoring System (LESS) (Padua et al., 2009), the Drop Jump
100 test (Barber-Westin, Smith, Campbell, & Noyes, 2010) and the Tuck Jump

101 Assessment (TJA) (Myer, Ford, & Hewett, 2008). There is a paucity of
102 evidence to support the validity and reliability of these tools (Frohm, Heijne,
103 Kowalski, Svensson, & Myklebust, 2012; Kiesel, Plisky, & Voight, 2007).
104
105 The TJA is widely used to assess jump and landing performance and forms
106 part of the battery of physiological tests used to screen players within the
107 English Women's Super League and the English Women's National Team.
108 Performance during the TJA is scored using a 10 item observational tool that
109 documents technique flaws associated with the jump landing action for knee
110 and thigh motion, foot position during landing, and plyometric technique
111 (Herrington, Myer, & Munro, 2013; Myer et al., 2008). An individual is
112 identified as requiring interventions to correct technique flaws if the TJA
113 composite score is ≥ 6 . There is an ongoing debate about the clinical utility of
114 the TJA and a lack of empirical evidence to support the choice of a cut-off
115 point of ≥ 6 (Klugman, Brent, Myer, Ford, & Hewett, 2011; Myer et al., 2008;
116 Myer, Ford, Khoury, Succop, & Hewett, 2011). Moreover, there is limited
117 research on inter-rater reliability of the TJA (Dudley et al., 2013; Herrington
118 et al., 2013; Read, Oliver, de Ste Croix, Myer, & Lloyd, 2016), although a
119 recent study of 50 elite level male youth football players concluded that the
120 TJA criterion for knee valgus was a reliable measure of landing performance
121 (Read et al., 2016).
122
123 To date, there have been no studies that have investigated technique flaws
124 associated with the TJA in elite female football players. The primary aim of
125 this study was to determine the inter-rater reliability and degree of internal

126 consistency of the TJA. The secondary aim was to identify the most
127 commonly occurring technique flaws in elite female football players.
128

129 **METHOD**

130 **Design**

131 This study was designed to measure inter-rater reliability of two
132 physiotherapists and two sports and conditioning coaches who
133 independently scored elite female football players undertaking a tuck jump
134 test.

135

136 **Participants**

137 Participants were 60 elite international female football players (mean \pm SD:
138 age = 20.27 \pm 3.44yrs; height = 168.02 \pm 5.26cm; mass = 62.54 \pm 6.33kg)
139 who were medically fit to complete mandatory physiological screening. Each
140 participant completed a tuck jump test that was video recorded and
141 subsequently assessed by each of the four raters. Assessment of the tuck
142 jump test is routinely included in a battery of physiological tests used at the
143 team's international training camps and therefore all players were familiar
144 with the tuck jump procedure. Written consent was provided by all
145 participants and raters, and ethical approval was granted by Sheffield Hallam
146 University.

147

148 **Raters**

149 Two physiotherapy and two strength a conditioning coaching staff from the
150 Women's English Football Association independently scored a video
151 recording of each player undertaking the tuck jump test. Raters regularly
152 used the TJA as a screening tool and were experienced with the TJA scoring
153 process. Characteristics of the raters were:

- 154 • Rater 1: Physiotherapist - [REDACTED] Women's Football
155 Association (FA) with 5 years of experience in elite female football; 10
156 years post qualifying experience (Health & Care Professions Council
157 registered Physiotherapist)
- 158 • Rater 2: Physiotherapist - [REDACTED] for a women's super
159 league team with 3 years of experience at an FA girls centre of
160 excellence; 11 years post qualifying experience (Health & Care
161 Professions Council registered Physiotherapist)
- 162 • Rater 3: Strength and Conditioning Coach - [REDACTED] Women's FA
163 with 8 years of experience in elite football; 11 years post qualifying
164 experience (United Kingdom Strength & Conditioning Association
165 accredited)
- 166 • Rater 4: Strength and Conditioning Coach - [REDACTED]
167 who had worked with multi-sport elite athletes and had 1 year post
168 qualifying experience with football players of a national standard
169 (United Kingdom Strength & Conditioning Association accredited)

170 The four raters had a total of 17 years' experience working with female
171 football players at national and international standard.

172

173 **Procedures**

174 A video recording was taken of each player completing a tuck jump test on
175 an indoor artificial 4G playing surface. All players wore 'astro turf' football
176 shoes. Ambient temperature and humidity were not controlled during testing.
177 The tuck jump test was facilitated by the Principal Investigator who provided
178 standardised verbal instructions and a practical demonstration of tuck jumps

179 to each participant immediately before they took the tuck jump test. A video
180 recording of individual tuck jumps from the sagittal and coronal was made
181 using two Sony PJ410 High Definition cameras on tripods. The tuck jump
182 test was standardised in-line with previously published protocols (Dudley et
183 al., 2013; Herrington et al., 2013; Myer et al., 2008). Two strips of 2.5cm tape
184 were placed 20cm apart and aligned parallel to each other. Participants were
185 instructed to stand with one foot on each tape strip and to perform repeated
186 tuck jumps for 10 seconds, lifting their knees to be level with the hips in the
187 horizontal plane, and to return to the start position. Participants were
188 encouraged to use a high level of effort. No feedback was given to
189 participants whilst they performed tuck jumps.

190

191 Each rater independently scored the tuck jump test of each participant by
192 watching the video in real time. In order to standardise the test raters were
193 instructed to view the video of each participant no more than 3 times prior to
194 scoring their tuck jumps across the 10 criterion of the TJA screening tool
195 (Dudley et al., 2013; Herrington et al., 2013; Myer et al., 2008). A score of 1
196 was assigned if the participant failed to meet an individual criterion on any
197 occasion during the test (i.e. had a technique flaw) and a score of 0 was
198 assigned if the participant did not exhibit a technique flaw. The total score
199 was calculated for each participant with higher scores indicative of poorer
200 performance.

201

202 **Data Analysis**

203 Raw data was screened for anomalies including data inputted incorrectly. A
204 one-variable χ^2 test was conducted to measure the association between the
205 observed and expected frequencies of technique flaws. The minimum
206 number of participants required to detect a kappa coefficient as statistically
207 significant when the value of kappa (K) was set at $k = .00$ (with 80% power)
208 was $n=39$ (Sim & Wright, 2005). Fleiss Kappa (an extension of the Cohen's
209 kappa coefficient (k) for two raters) was utilised to assess multiple inter-rater
210 agreement for each TJA criterion with standard error of measurement (SEM)
211 and 95% confidence intervals (CI). The significance level was set at $p < 0.05$.
212 Microsoft Office Excel 2010 was used to compute Fleiss Kappa. A weighted
213 kappa (K_w) was performed on the total score to calculate the degree of
214 disagreement. The interpretation of Cohen's kappa coefficient utilised
215 arbitrary theoretical values set by Fleiss et al. (2003) as < 0.40 poor, $0.41 -$
216 0.75 fair to good and $0.75 - 1.00$ very good, with > 0.75 used as a cut off for
217 clinically acceptable measure of inter-rater agreement (Sim & Wright, 2005).

218

219 Internal consistency of total scores was assessed by Cronbach's alpha
220 reliability coefficient. There is no consensus for the lower limit of the
221 coefficient so the following rules of thumb were applied: $> .9$ – Excellent, $>$
222 $.8$ – Good, $> .7$ – Acceptable, $> .6$ – Questionable, $> .5$ – Poor, and $< .5$ –
223 Unacceptable (George & Mallory, 2003). Cronbach's alpha analysis was
224 performed using SPSS version 21.

225

226 **RESULTS**

227

228 **Frequency of technique flaws**

229 The sum of technique flaws scored by all four raters was 665 (Table 1). The
230 most frequent technique flaw was Criterion 2 'Thighs do not reach parallel'
231 (N=147/665, 22%), the second most frequent technique flaw was criterion 1
232 'Knee valgus on landing' (N=80/665, 12%) and the least frequent technique
233 flaw was Criterion 9 'Pause between jumps' (N=23/665, 4%).

234

235 [Insert Table 1 here]

236

237 The χ^2 value of 152.1, DF=9 had an associated probability value of 0.0001.
238 Thus we can accept that there was a significant difference between the
239 observed and expected frequencies.

240

241 The frequency of technique flaws within each of the categories of the TJA
242 (Knee and thigh motion, comprising 3 criterion; Foot position during landing,
243 comprising 5 criterion; and Foot position during landing, comprising 2
244 criterion) were calculated relative to the maximum number of technique flaws
245 possible was calculated (i.e. (60 participants x 4 raters) x the number of
246 criterion included in each sub-category) There were 234/720 (32.5%)
247 technique flaws for Knee and thigh motion 307/1200 (46%) technique flaws
248 for Foot position during landing and 64/480 (13.3%) technique flaws for
249 Plyometric technique.

250

251 **Inter-rater agreement**

252 The Fleiss kappa coefficient values used to determine inter-rater agreement
253 ranged from 'fair-to-good', $k = .46$ (95% CI, .35 - .56) to 'very good' $k = .86$
254 (95% CI, .74 - .94). Raters reached substantial agreement for 'Lower
255 extremity valgus at landing' $k = .83$ (95% CI, .72 – .93); 'Thighs do not reach
256 parallel (peak of jump)' $k = .84$ (95% CI, .74 - .94); 'Thighs not equal side to
257 side' $k = .86$ (95% CI, .75 - .96). A descending order of inter-rater agreement
258 from criterion 1 to criterion 10 was observed in the results.

259

260 [Insert Table 2 here]

261

262 Weighted kappa (k_w) coefficient values used to determine inter-rater
263 agreement of the composite score ranged from $k_w = .62$ (95% CI, .48 – .76)
264 to $k_w = .80$ (95% CI, .70 – .90) suggesting a 'fair-to-very good' level of inter-
265 rater agreement.

266

267 [Insert Table 3 here]

268

269 **Internal Consistency**

270 Low alpha values were detected across all four raters for the entire TJA
271 scale. Internal consistency was reassessed with items 9 and 10 removed
272 (Plyometric technique) as the repeated plyometric nature of the TJA over a
273 10 second period differentiates it from previous tests such as the Landing
274 Error Scoring System (Padua et al 2015). As an 8 item scale there were
275 negligible alterations in internal consistency (range $\alpha = .091 - .161$, Table 4).

276 Internal consistency results suggest that the TJA scale and sub items are not

277 unidimensional

278

279 [Insert Table 4 here]

280

281 **DISCUSSION**

282 **Statement of principal findings**

283 This is the first study to investigate technique flaws associated with the TJA
284 in elite female football players. The TJA was designed for use with athletic
285 populations to detect technique flaws in jump landing tasks (Myer et al.,
286 2008). In our study four raters identified 665 technique flaws in 60
287 participants. The most frequent flaws were 'Thighs do not reach parallel'
288 (criterion 2) and 'Knee valgus on landing' (criterion 1), which are part of the
289 'Knee and thigh motion' category of the TJA. The least frequent technique
290 flaws were criterion 9 'Pause between jumps' and criterion 10 'Technique
291 declines prior to 10 seconds', which form the 'Plyometric technique' category
292 of the TJA. The inter-rater level of agreement for the total score of the TJA
293 was 'fair-to-very good' with all criteria of the 'Knee and thigh motion' category
294 reaching clinically acceptable levels of agreement. Low alpha values for
295 internal consistency suggest the individual criteria contained within the TJA
296 are not unidimensional therefore they are not measuring the same underlying
297 construct (i.e. jump landing task).

298

299 **Meaning of the study findings**

300 The TJA is currently used by teams within the English Women's Super
301 League and has been used by the English Women's National Teams. The
302 TJA is utilised by coaches and medical staff as a screening tool to assess
303 performance of jump landing tasks, and as an outcome measure in regard of
304 neuromuscular retraining.

305

306 Previous studies investigating the TJA have not clearly identified the
307 frequency of individual technique flaws and this limits our ability to compare
308 between studies. In our study criterion 2 'Thighs do not reach parallel' was
309 the most frequently identified technique flaw and 'Pause between jumps' was
310 the least frequently identified technique flaw. Dudley et al. (2013) also
311 reported criterion 2 as the most frequently identified technique flaw but did
312 not report the rank of other TJA criteria.

313

314 Herrington et al. (2013) reported the inter-rater level of agreement for the
315 composite score of the TJA using 2 raters to be very good/excellent ($K=0.88$)
316 in a sample of ten athletes. The inter-rater percentage of exact agreement
317 between raters across all ten criteria was 93% (range 80%-100%, i.e. high).
318 Interestingly, Dudley et al. (2013) reported the inter-rater level of agreement
319 using 5 Raters to be poor in 40 recreationally active university students
320 ($ICC=0.47$, 95% CI 0.33-0.62). Read et al. (2016) used a test-retest design to
321 investigate intersession reliability of the TJA in 50 elite male youth football
322 players. Although reliability was found to be strong ($ICC=0.88$) the authors
323 suggested caution in interpreting the composite score of the TJA due to high
324 within-subject variation in a number of individual criterion.

325

326 The difference in the reported levels of agreement may in part be explained
327 by the statistical test selected by investigators. Sample sizes of at least 50
328 are recommended when using percentage of exact agreement (Birkimer &
329 Brown, 1979). Therefore results from studies containing smaller sample
330 sizes are quite probably the result of chance agreement and should be

331 considered with caution. Each of the TJA criteria is scored in a dichotomous
332 manner i.e. flaw occurred or no flaw occurred and the data is therefore
333 characterised as nominal. Kappa coefficients are recommend for use as the
334 preferred statistical test to determine the inter-rater level of agreement for
335 nominal data (Hallgren, 2012). We utilised Fleiss Kappa to determine inter-
336 rater agreement for individual TJA criteria and a weighted Kappa to
337 determine inter-rater agreement for the composite score.

338

339 Cronbach's alpha is considered to infer the degree to which the criteria
340 measures a single unidimensional construct. Our internal consistency
341 statistics raise concern about the construct validity of the TJA suggesting
342 redundancy of TJA criteria. Analysis with the 'Plyometric technique' category
343 removed to determine if the psychometric properties of the test would be
344 improved as an 8 item measure found that unidimensionality remained
345 violated. However it is important to note that jump landing is a skill
346 characterised by multiple factors.

347

348 Errors are also introduced into TJA by variability in the interpretation of what
349 constitutes the occurrence of a technique flaw. Dudley et al. (2013) claims
350 that instructions used to assess performance during the tuck jump test do not
351 specify whether a technical flaw should be scored by the rater if observed
352 only on a single occasion or whether it needs to occur repeatedly and
353 consistently throughout the assessment, lead to inconsistency of scoring
354 between assessors. Our raters were instructed to score the presence of a

355 technique flaw when they identified a flaw within the time frame of the test i.e
356 10 seconds.

357

358 Myer et al. (2008) suggest that individuals with a total TJA score of ≥ 6 have
359 an increased risk of sustaining an ACL injury and interventions to address
360 landing errors should be employed. To our knowledge there is no empirical
361 evidence to support the use of a cut point of ≥ 6 . The results from the present
362 study suggest that TJA criterion are not internally consistent and do not have
363 a coherent empirical structure (i.e. are not interrelated). However if multiple
364 items were highly interrelated then a case could be made that some items
365 should be removed as they are measuring the same thing. It is important to
366 note that although the items were not internally consistent it does necessarily
367 mean that the composite score is not meaningful.

368

369 A recent study by Read et al. (2016) concluded that only the knee valgus
370 criterion could be reliably used to screen elite youth male football players as
371 a measure of landing performance. A prospective study by (Hewett, Myer,
372 Ford, et al., 2005) found increased knee abduction angles (knee valgus)
373 during a plyometric activity to be a significant predictor of ACL injury. In our
374 study 'Knee valgus on landing' was the second commonest technique flaw
375 reaching clinically acceptable levels of agreement. ACL strain from valgus
376 knee loading has been confirmed through cadaver, in vivo and 3-dimensional
377 motion analysis methods (Fukuda et al., 2003; Kanamori et al., 2000; Markolf
378 et al., 1995). Increased internal hip rotation, coupled with increased external
379 rotation of the tibia (dynamic knee valgus) has been found in female football

380 players during jump landing and these have been used to predict ACL injury
381 (Alentorn-Geli et al., 2009; Barber-Westin et al., 2010). Female athletes have
382 been found to preferentially rely on increased quadriceps recruitment relative
383 to hamstring recruitment during incremental vertical jump test using surface
384 electromyography (Myer, Brent, Ford, & Hewett, 2011). In addition, a
385 quadriceps dominant landing strategy may increase the risk of sustaining an
386 ACL rupture (Alentorn-Geli et al., 2009; Hewett, Myer, & Ford, 2005).

387

388 **Limitations of the study**

389 A number of limitations need to be considered when interpreting our study
390 findings. It is possible that the sequence of items in the TJA impacts on recall
391 rates because we observed a trend of decreasing item frequencies and
392 kappa scores through items 1 to 10. Furthermore, Cronbach's alpha is
393 considered a crude measure of reliability (coefficient of reliability) and can be
394 influenced by the number of scale items and redundant items (DeVellis,
395 2012). Exploratory factor analysis would have provided a more in-depth
396 assessment of the factor structure and dimensionality of the TJA, although
397 with such low internal consistency scores pursuing exploratory factor
398 analysis at this stage may not have provided any further meaningful
399 information. Thus, in future investigators may wish to consider the 'Knee
400 valgus on landing' criterion during jump landing tasks as a predictor of ACL
401 injury in female football.

402

403 **Conclusion**

404 There is a paucity of studies evaluating the psychometric properties of the
405 TJA and those that exist have inconsistent findings (Dudley et al., 2013;
406 Herrington et al., 2013; Read et al., 2016). Our study found that the criterion
407 used in the TJA are not measuring the same underlying construct (i.e. jump
408 landing task). This raises doubt about the clinical utility of the TJA in its
409 current form. The TJA was intended for use in elite athletes, and assessors
410 that were experienced in its use. Our study was concordant with these
411 directives. Thus, we recommend that assessors should remain cautious
412 when interpreting the composite score of the TJA. The three individual
413 criterion that contribute to the 'Knee and thigh motion' category reached
414 clinically acceptable levels of agreement and may be useful when assessing
415 athletic performance of jump landing tasks. In addition the 'Knee valgus on
416 landing' criterion may have clinical utility and contribute to the screening of
417 elite female football players for potential ACL injury risk. We hope our study
418 catalyses further research in this field.

419
420
421
422
423
424
425
426
427
428
429
430
431
432

433
 434
 435
 436
 437
 438

Tables & Figures

TJA Criterion	Frequency response		
	Total flaws	% of total flaws available	% of relative flaws
Knee & thigh motion			
Lower extremity valgus at landing	80	33.3	12
Thighs do not reach parallel	147	61	21.1
Thighs not equal side to side	67	28	10.1
Foot position during landing			
Foot placement not shoulder width apart	67	28	10.1
Foot placement not parallel	68	28.2	10.2
Foot contact timing not equal	50	20.2	7.5
Excessive landing noise	44	18.3	6.6
Does not land in the same footprint	78	33	11.7
Plyometric technique			
Pause between jumps	23	9.5	3.5
Technique declines prior to 10seconds	41	17	6.2

439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449

Table 1: Frequency response of each TJA criterion listed within respective TJA categories; 'knee & thigh motion'; 'foot position during landing'; 'plyometric technique'

450

451

452

453 Table 2: Fleiss Kappa Inter-rater agreement of TJA criterion

TJA criteria	Fleiss Kappa (κ) Inter-rater agreement
Lower extremity valgus at landing	$\kappa = .83$ (95% CI, .72 - .93), $p < .000$
Thighs do not reach parallel (peak of jump)	$\kappa = .84$ (95% CI, .74 - .94), $p < .000$
Thighs not equal side to side	$\kappa = .86$ (95% CI, .75 - .96), $p < .000$
Foot placement not shoulder width apart	$\kappa = .75$ (95% CI, .65 - .85), $p < .000$
Foot placement not parallel (front and back)	$\kappa = .73$ (95% CI, .62 - .82), $p < .000$
Foot contact timing not equal	$\kappa = .70$ (95% CI, .60 - .81), $p < .000$
Does not land in the same footprint	$\kappa = .60$ (95% CI, .50 - .71), $p < .000$
Excessive landing noise	$\kappa = .63$ (95% CI, .53 - .73), $p < .000$
Pause between jumps	$\kappa = .60$ (95% CI, .49 - .69), $p < .000$
Technique declines prior to 10seconds	$\kappa = .46$ (95% CI, .35 - .56), $p < .000$

454

455

456 Table 3: Weighted Kappa Inter-rater agreement of TJA criterion

Paired raters	Weighted Kappa (K_w) Inter-rater agreement (Sum score)
PT ₁ : PT ₂	$K_w = .65$ (95% CI, .51 - .79)
PT ₁ : SC ₁	$K_w = .80$ (95% CI, .70 - .90)
PT ₁ : SC ₂	$K_w = .67$ (95% CI, .54 - .80)
PT ₂ : SC ₁	$K_w = .70$ (95% CI, .54 - .84)
PT ₂ : SC ₂	$K_w = .79$ (95% CI, .69 - .88)
SC ₁ : SC ₂	$K_w = .62$ (95% CI, .48 - .76)

Abbreviations: TJA: tuck jump assessment,

PT: physiotherapist, SC: strength &
conditioning coach

457

458 Table 4: Internal consistency

459

Cronbach's Alpha (α)	Rater 1 (PT ₁)	Rater 2 (PT ₂)	Rater 3 (SC ₁)	Rater 4 (SC ₂)
Entire scale	.073	-.033	.018	.129
TJA categories	Rater 1 (PT ₁)	Rater 2 (PT ₂)	Rater 3 (SC ₁)	Rater 4 (SC ₂)
Knee & Thigh motion	-.397	-.720	-.653	-.509
Foot position during landing	.288	.163	.220	.191
Plyometric technique	.528	.306	.222	.339
With items 9 & 10 removed	.161	.091	.112	.154

Abbreviations: TJA: tuck jump assessment, PT: physiotherapist, SC: strength &
conditioning coach

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477 **References**

478

- 479 Alentorn-Geli, E., Myer, G. D., Silvers, H. J., Samitier, G., Romero, D.,
480 Lazaro-Haro, C., & Cugat, R. (2009). Prevention of non-contact
481 anterior cruciate ligament injuries in soccer players. Part 1:
482 Mechanisms of injury and underlying risk factors. *Knee Surg Sports*
483 *Traumatol Arthrosc*, *17*(7), 705-729. doi: 10.1007/s00167-009-0813-1
- 484 Barber-Westin, S. D., Smith, S. T., Campbell, T., & Noyes, F. R. (2010). The
485 drop-jump video screening test: retention of improvement in
486 neuromuscular control in female volleyball players. *J Strength Cond*
487 *Res*, *24*(11), 3055-3062. doi: 10.1519/JSC.0b013e3181d83516
- 488 Birkimer, J. C., & Brown, J. H. (1979). Back to basics: Percentage agreement
489 measures are adequate, but there are easier ways. *J Appl Behav*
490 *Anal*, *12*(4), 535-543.
- 491 Boden, B. P., Torg, J. S., Knowles, S. B., & Hewett, T. E. (2009). Video
492 analysis of anterior cruciate ligament injury: abnormalities in hip and
493 ankle kinematics. *Am J Sports Med*, *37*(2), 252-259. doi:
494 10.1177/0363546508328107
- 495 DeVellis, R. F. (2012). *Scale Development: Theory and Applications* (3rd
496 edition). Los Angeles, CA. Sage.
- 497 Dudley, L. A., Smith, C. A., Olson, B. K., Chimera, N. J., Schmitz, B., &
498 Warren, M. (2013). Interrater and Intrarater Reliability of the Tuck
499 Jump Assessment by Health Professionals of Varied Educational
500 Backgrounds. *J Sports Med (Hindawi Publ Corp)*, *2013*, 483503. doi:
501 10.1155/2013/483503
- 502 Faude, O., Junge, A., Kindermann, W., & Dvorak, J. (2006). Risk factors for
503 injuries in elite female soccer players. *Br J Sports Med*, *40*(9), 785-
504 790. doi: 10.1136/bjism.2006.027540
- 505 Ford, K. R., Manson, N. A., Evans, B. J., Myer, G. D., Gwin, R. C., Heidt, R.
506 S., Jr., & Hewett, T. E. (2006). Comparison of in-shoe foot loading
507 patterns on natural grass and synthetic turf. *J Sci Med Sport*, *9*(6),
508 433-440. doi: 10.1016/j.jsams.2006.03.019
- 509 Frohm, A., Heijne, A., Kowalski, J., Svensson, P., & Myklebust, G. (2012). A
510 nine-test screening battery for athletes: a reliability study. *Scand J*
511 *Med Sci Sports*, *22*(3), 306-315. doi: 10.1111/j.1600-
512 0838.2010.01267.x
- 513 Fukuda, Y., Woo, S. L., Loh, J. C., Tsuda, E., Tang, P., McMahon, P. J., &
514 Debski, R. E. (2003). A quantitative analysis of valgus torque on the
515 ACL: a human cadaveric study. *J Orthop Res*, *21*(6), 1107-1112. doi:
516 10.1016/s0736-0266(03)00084-6

- 517 Giza, E., Mithofer, K., Farrell, L., Zarins, B., & Gill, T. (2005). Injuries in
518 women's professional soccer. *Br J Sports Med*, 39(4), 212-216;
519 discussion 212-216. doi: 10.1136/bjism.2004.011973
- 520 Hagglund, M., Walden, M., & Atroshi, I. (2009). Preventing knee injuries in
521 adolescent female football players - design of a cluster randomized
522 controlled trial [NCT00894595]. *BMC Musculoskelet Disord*, 10, 75.
523 doi: 10.1186/1471-2474-10-75
- 524 Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational
525 Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*, 8(1),
526 23-34.
- 527 Herrington, L., Myer, G. D., & Munro, A. (2013). Intra and inter-tester
528 reliability of the tuck jump assessment. *Phys Ther Sport*, 14(3), 152-
529 155. doi: 10.1016/j.ptsp.2012.05.005
- 530 Hewett, T. E., Myer, G. D., & Ford, K. R. (2005). Reducing knee and anterior
531 cruciate ligament injuries among female athletes: a systematic review
532 of neuromuscular training interventions. *J Knee Surg*, 18(1), 82-88.
- 533 Hewett, T. E., Myer, G. D., Ford, K. R., Heidt, R. S., Jr., Colosimo, A. J.,
534 McLean, S. G., . . . Succop, P. (2005). Biomechanical measures of
535 neuromuscular control and valgus loading of the knee predict anterior
536 cruciate ligament injury risk in female athletes: a prospective study.
537 *Am J Sports Med*, 33(4), 492-501. doi: 10.1177/0363546504269591
- 538 Kanamori, A., Woo, S. L., Ma, C. B., Zeminski, J., Rudy, T. W., Li, G., &
539 Livesay, G. A. (2000). The forces in the anterior cruciate ligament and
540 knee kinematics during a simulated pivot shift test: A human cadaveric
541 study using robotic technology. *Arthroscopy*, 16(6), 633-639. doi:
542 10.1053/jars.2000.7682
- 543 Kiesel, K., Plisky, P. J., & Voight, M. L. (2007). Can Serious Injury in
544 Professional Football be Predicted by a Preseason Functional
545 Movement Screen? *N Am J Sports Phys Ther*, 2(3), 147-158.
- 546 Klugman, M. F., Brent, J. L., Myer, G. D., Ford, K. R., & Hewett, T. E. (2011).
547 Does an in-season only neuromuscular training protocol reduce
548 deficits quantified by the tuck jump assessment? *Clin Sports Med*,
549 30(4), 825-840. doi: 10.1016/j.csm.2011.07.001
- 550 Le Gall, F., Carling, C., & Reilly, T. (2008). Injuries in young elite female
551 soccer players: an 8-season prospective study. *Am J Sports Med*,
552 36(2), 276-284. doi: 10.1177/0363546507307866
- 553 Markolf, K. L., Burchfield, D. M., Shapiro, M. M., Shepard, M. F., Finerman,
554 G. A., & Slauterbeck, J. L. (1995). Combined knee loading states that
555 generate high anterior cruciate ligament forces. *J Orthop Res*, 13(6),
556 930-935. doi: 10.1002/jor.1100130618
- 557 Myer, G. D., Brent, J. L., Ford, K. R., & Hewett, T. E. (2011). Real-time
558 assessment and neuromuscular training feedback techniques to
559 prevent ACL injury in female athletes. *Strength Cond J*, 33(3), 21-35.
560 doi: 10.1519/SSC.0b013e318213afa8
- 561 Myer, G. D., Ford, K. R., & Hewett, T. E. (2008). Tuck Jump Assessment for
562 Reducing Anterior Cruciate Ligament Injury Risk. *Athl Ther Today*,
563 13(5), 39-44.
- 564 Myer, G. D., Ford, K. R., Khoury, J., Succop, P., & Hewett, T. E. (2011).
565 Biomechanics laboratory-based prediction algorithm to identify female

566 athletes with high knee loads that increase risk of ACL injury. *Br J*
567 *Sports Med*, 45(4), 245-252. doi: 10.1136/bjism.2009.069351
568 Nilstad, A., Andersen, T. E., Bahr, R., Holme, I., & Steffen, K. (2014). Risk
569 factors for lower extremity injuries in elite female soccer players. *Am J*
570 *Sports Med*, 42(4), 940-948. doi: 10.1177/0363546513518741
571 Padua, D. A., Marshall, S. W., Boling, M. C., Thigpen, C. A., Garrett, W. E.,
572 Jr., & Beutler, A. I. (2009). The Landing Error Scoring System (LESS)
573 Is a valid and reliable clinical assessment tool of jump-landing
574 biomechanics: The JUMP-ACL study. *Am J Sports Med*, 37(10), 1996-
575 2002. doi: 10.1177/0363546509343200
576 Read, P. J., Oliver, J. L., de Ste Croix, M. B., Myer, G. D., & Lloyd, R. S.
577 (2016). Reliability of the Tuck Jump Injury Risk Screening Assessment
578 in Elite Male Youth Soccer Players. *J Strength Cond Res*, 30(6),
579 1510-1516. doi: 10.1519/JSC.0000000000001260
580 Renstrom, P., Ljungqvist, A., Arendt, E., Beynon, B., Fukubayashi, T.,
581 Garrett, W., . . . Engebretsen, L. (2008). Non-contact ACL injuries in
582 female athletes: an International Olympic Committee current concepts
583 statement. *Br J Sports Med*, 42(6), 394-412. doi:
584 10.1136/bjism.2008.048934
585 Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use,
586 interpretation, and sample size requirements. *Phys Ther*, 85(3), 257-
587 268.
588 Walden, M., Hagglund, M., Magnusson, H., & Ekstrand, J. (2011). Anterior
589 cruciate ligament injury in elite football: a prospective three-cohort
590 study. *Knee Surg Sports Traumatol Arthrosc*, 19(1), 11-19. doi:
591 10.1007/s00167-010-1170-9
592 Zazulak, B. T., Hewett, T. E., Reeves, N. P., Goldberg, B., & Cholewicki, J.
593 (2007). The effects of core proprioception on knee injury: a
594 prospective biomechanical-epidemiological study. *Am J Sports Med*,
595 35(3), 368-373. doi: 10.1177/0363546506297909

596