

Inter-rater reliability, internal consistency and common technique flaws of the Tuck Jump Assessment in elite female football players

MAYHEW, Laurence, JOHNSON, Mark I., FRANCIS, Peter, SNOWDON, Nicky and JONES, Gareth <<http://orcid.org/0000-0002-0313-0092>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/16156/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

MAYHEW, Laurence, JOHNSON, Mark I., FRANCIS, Peter, SNOWDON, Nicky and JONES, Gareth (2017). Inter-rater reliability, internal consistency and common technique flaws of the Tuck Jump Assessment in elite female football players. *Science and Medicine in Football*, 1 (2), 139-144.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Common technique flaws identified by the tuck jump assessment in elite female soccer players

Journal:	<i>Journal of Sports Sciences: Science and Medicine in Football</i>
Manuscript ID	Draft
Manuscript Type:	Original Papers
Keywords:	Tuck Jump Assessment, landing faults, female football, reliability, internal consistency
Abstract:	<p>Although injury rates between elite female and male players are comparable, female players are more likely to sustain an anterior cruciate ligament (ACL) injury. The common mechanism of ACL injury is non-contact trauma sustained when landing from a jump. The reliability of the Tuck Jump Assessment (TJA) has been challenged. The aim of this study was to identify commonly occurring technique flaws during the TJA and to determine inter-rater agreement and internal consistency. Sixty elite female soccer players were recruited (mean (SD): age = 20.27 ± 3.44yrs). Four raters independently assessed each participant post hoc. Six hundred and sixty-five technique flaws were recorded. Criterion 2 'Thighs do not reach parallel' (N=147/665) and criterion 1 'Knee valgus on landing' (N=80/665) were the most common. The most common fault category was 'Knee and thigh motion' (N=234/720, 32.5%). Clinically acceptable levels of agreement were reached for 'Lower extremity valgus at landing' $k = .83$ (95% CI, .72 - .93); 'Thighs do not reach parallel' $k = .84$ (95% CI, .74 - .94); 'Thighs not equal side to side' $k = .86$ (95% CI, .75 - .96). The level of agreement for the composite score of all 10 criteria ranged from $k_w = .62$ (95% CI, .48 - .76) to $k_w = .80$ (95% CI, .70 - .90) suggesting a 'fair-to-very good' level of inter-rater agreement. Internal consistency results suggest that the TJA is not unidimensional. We recommend the 'knee and thigh' motion category of the TJA for screening elite female soccer players.</p>

SCHOLARONE™
Manuscripts

Title

Common technique flaws identified by the tuck jump assessment in elite female football players

Authors:**Author affiliations****Corresponding author**

Article Type: Original Article

Running header: Inter-rater Reliability of Tuck Jump Assessment

Funding: This study did not receive external funding

Ethical Approval: Ethical approval was granted by Sheffield Hallam University

Contribution of Authors:

- Study design: LM, GJ, NS
- Data collection: LM
- Data analysis: LM, GJ, MIJ
- Preparation of the manuscript: GJ, LM, MIJ, PF

Competing Interests: None

Acknowledgments: None

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

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

ABSTRACT

Although injury rates between elite female and male players are comparable, female players are more likely to sustain an anterior cruciate ligament (ACL) injury. The common mechanism of ACL injury is non-contact trauma sustained when landing from a jump. The reliability of the Tuck Jump Assessment (TJA) has been challenged. The aim of this study was to identify commonly occurring technique flaws during the TJA and to determine inter-rater agreement and internal consistency. Sixty elite female football players were recruited (mean (SD): age = 20.27 ± 3.44yrs). Four raters independently assessed each participant post hoc. Six hundred and sixty-five technique flaws were recorded. Criterion 2 'Thighs do not reach parallel' (N=147/665) and criterion 1 'Knee valgus on landing' (N=80/665) were the most common. The most common fault category was 'Knee and thigh motion' (N=234/720, 32.5%). Clinically acceptable levels of agreement were reached for 'Lower extremity valgus at landing' $k = .83$ (95% CI, .72 – .93); 'Thighs do not reach parallel' $k = .84$ (95% CI, .74 - .94); 'Thighs not equal side to side' $k = .86$ (95% CI, .75 - .96). The level of agreement for the composite score of all 10 criteria ranged from $k_w = .62$ (95% CI, .48 – .76) to $k_w = .80$ (95% CI, .70 – .90) suggesting a 'fair-to-very good' level of inter-rater agreement. Internal consistency results suggest that the TJA is not unidimensional. We recommend the 'knee and thigh' motion category of the TJA for screening elite female football players.

INTRODUCTION

The advent of professionalism and the exponential rise in the number of UEFA registered female football players has corresponded with a significant increase in the reported incidence of injury. The estimated incidence of injury for female players is between 12.6 and 24.0 injuries per 1000 hours of match play and between 1.2 and 7.0 injuries per 1000 hours of training (Giza, Mithofer, Farrell, Zarins, & Gill, 2005; Le Gall, Carling, & Reilly, 2008; Nilstad, Andersen, Bahr, Holme, & Steffen, 2014). Although injury rates between elite female and male players are comparable per se (Hagglund, Walden, & Atroshi, 2009), female players are more likely to sustain an anterior cruciate ligament (ACL) injury of the knee than their male counterparts (Walden, Hagglund, Magnusson, & Ekstrand, 2011). ACL injury in female players is more likely to occur at an earlier age (Renstrom et al., 2008) and a previous history of ACL injury is considered to be a significant risk factor for reinjury (Faude, Junge, Kindermann, & Dvorak, 2006). The most common mechanism of ACL injury is an acute non-contact trauma sustained during rapid decelerating movements, for example when landing from a jump and changes of direction when running (Walden et al., 2011).

Observational screening tools are commonly used to identify faulty movement patterns during key athletic tasks (Frohm, Heijne, Kowalski, Svensson, & Myklebust, 2012; Kiesel, Plisky, & Voight, 2007). Reduced neuromuscular control during landing may result in increased knee valgus angles (ligament dominance) and increase the likelihood of an individual sustaining an anterior cruciate ligament (ACL) injury (Hewett, Myer, Ford, et

1
2
3 al., 2005). Screening tools used to assess jump landing tasks include the
4
5 Landing Error Scoring System (LESS) (Padua et al., 2009), the Drop Jump
6
7 test (Barber-Westin, Smith, Campbell, & Noyes, 2010) and the Tuck Jump
8
9 Assessment (TJA) (Myer, Ford, & Hewett, 2008).
10

11
12
13
14 Performance during the TJA is scored using a 10 criterion screening tool to
15
16 identify technique flaws associated with the jump landing action for knee and
17
18 thigh motion, foot position during landing and plyometric technique
19
20 (Herrington, Myer, & Munro, 2013; Myer et al., 2008). The TJA involves
21
22 repetitive jumps over a set period of time and the potential effects of fatigue
23
24 may be observed. Practitioners use a TJA composite score of ≥ 6 to instigate
25
26 interventions to correct technique flaws, although there is a lack of empirical
27
28 evidence to support the choice of this cut-off point (Klugman, Brent, Myer,
29
30 Ford, & Hewett, 2011; Myer et al., 2008; Myer, Ford, Khoury, Succop, &
31
32 Hewett, 2011). Moreover, there has been limited research on inter-rater
33
34 reliability of the TJA in athletic populations (Dudley et al., 2013; Herrington et
35
36 al., 2013; Read, Oliver, de Ste Croix, Myer, & Lloyd, 2016). Herrington et al.
37
38 (2013) found that the level of inter-rater level of agreement was 'very
39
40 good/excellent', although the sample size was small with two Raters scoring
41
42 five female and five male university sports science students. In contrast,
43
44 Dudley et al. (2013) used five Raters to score 40 recreationally active
45
46 university students not involved in college athletics and found that the level of
47
48 inter-rater agreement was 'poor'. They concluded that existing protocols
49
50 were inadequate to ensure consistent TJA scoring. Recently, Read et al.
51
52
53
54
55
56
57
58
59
60

1
2
3 (2016) suggested that the knee valgus criterion of the TJA was a reliable
4
5 measure of landing performance but that the composite TJA score was not.
6
7

8
9
10 To date, there have been no studies that have investigated technique flaws
11 associated with the TJA in elite female football players. The primary aim of
12 this study was to identify the most commonly occurring technique flaws in a
13 large sample of elite female football players. The secondary aim was to
14 determine the inter-rater reliability and degree of internal consistency of the
15 TJA.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHOD

Design

An inter-rater reliability study conducted in the U.K.

Participants

A convenience sample of two physiotherapists (PT) and two strength and conditioning coaches (SC) were approached, and agreed to take part in the study. Study participation involved the rating (scoring) of routinely collected TJA video data. Characteristics of these Raters were:

- PT1 - Itinerant member of staff for the Women's Football Association (FA) with 5 years of experience in elite female football; 10 years post qualifying experience (Health & Care Professions Council registered Physiotherapist)
- PT2 - Head Physiotherapist for a women's super league team with 3 years of experience at an FA girls centre of excellence; 11 years post qualifying experience (Health & Care Professions Council registered Physiotherapist)
- SC1 - Head Strength and Conditioning Coach for the Women's FA with 8 years of experience in elite football; 11 years post qualifying experience (United Kingdom Strength & Conditioning Association accredited)
- SC2 - Strength and Conditioning Coach for a university who had worked with multi-sport elite athletes and had 1 year post qualifying experience with football players of a national standard (United Kingdom Strength & Conditioning Association accredited)

1
2
3 In total the four Raters had 17 years post qualifying experience working with
4
5 female football players at national and international standard.
6
7

8
9
10 There were 60 elite international female football players that were required to
11
12 undertake one TJA as part of their team's mandatory physiological
13
14 screening. All players were briefed about the study and invited to take part.
15
16 This involved a video of their TJA to be recorded and subsequently scored
17
18 by the four Raters. All players agreed to participate in the study, were passed
19
20 medically fit and completed the TJA (mean \pm SD: age = 20.27 \pm 3.44yrs;
21
22 height = 168.02 \pm 5.26cm; mass = 62.54 \pm 6.33kg).
23
24
25
26

27 Ethical approval was granted by Sheffield Hallam University, and each player
28
29 and rater provided signed informed consent to take part in the study.
30
31
32
33

34 **Procedures**

35
36 Each international female football player undertook one TJA facilitated by the
37
38 principal investigator before a regular training session. The TJA was
39
40 conducted on an indoor artificial 4G playing surface. Ambient temperature
41
42 and humidity were not controlled during testing. Standardised verbal
43
44 instructions and a demonstration of the TJA was given to each participant
45
46 immediately prior to the TJA by the Principal Investigator. A video recording
47
48 of individual tuck jumps from the sagittal and coronal was made using two
49
50 Sony PJ410 High Definition cameras on tripods. The TJA was identical to
51
52 previously published protocols ((Dudley et al., 2013; Herrington et al., 2013;
53
54 Myer et al., 2008). Two strips of 2.5cm tape were placed 20cm apart and
55
56
57
58
59
60

1
2
3 aligned parallel to each other. Participants were instructed to stand with one
4
5 foot on each tape strip and to perform repeated tuck jumps, lifting their knees
6
7 to be level with the hips in the horizontal plane, and to return to the start
8
9 position for 10 seconds. Participants were encouraged to use a high level of
10
11 effort. No feedback was given to participants during the assessment.
12
13

14
15
16 Independent scoring of the TJA videos was conducted post-test by four
17
18 raters. Raters were instructed to view each video in real time and to view
19
20 each video no more than 3 times prior to scoring. Each rater was required to
21
22 score each tuck jump across 10 criterion (Dudley et al., 2013; Herrington et
23
24 al., 2013; Myer et al., 2008). A score of 0 was assigned if the participant met
25
26 the individual criterion. A score of 1 was assigned if the participant failed to
27
28 meet an individual criterion on any occasion during the test. A composite
29
30 score was calculated for each participant with a higher composite score
31
32 indicative of reduced performance on the TJA.
33
34
35
36
37
38

39 **Data Analysis**

40 Raw data was screened for anomalies including data inputted incorrectly.
41
42 The minimum number of participants required to detect a kappa coefficient
43
44 as statistically significant when the value of kappa (K) was set at $k = .00$ (with
45
46 80% power) was $n=39$ (Sim & Wright, 2005).
47
48

49 Fleiss Kappa (an extension of the Cohen's kappa coefficient (k) for two
50
51 raters) was utilised to assess multiple inter-rater agreement for each TJA
52
53 criterion with standard error of measurement (SEM) and 95% confidence
54
55
56
57
58
59
60

1
2
3 intervals (CI). The significance level was set at $p < 0.05$. Microsoft Office
4
5 Excel 2010 was used to compute Fleiss Kappa.
6
7

8
9
10 A weighted kappa (K_w) was performed on the composite score to calculate
11 the degree of disagreement. The interpretation of Cohen's kappa coefficient
12 utilised the theoretical values set by Fleiss et al. (2003) as < 0.40 poor, 0.41
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
– 0.75 fair to good and $0.75 - 1.00$ very good, with > 0.75 used as a cut off
for clinically acceptable measure of inter-rater agreement (Sim & Wright,
2005).

Internal consistency associated with the scores derived from TJA composite
score was assessed by Cronbach's alpha reliability coefficient. With no
widely acknowledged lower limit to the coefficient the following rules of
thumb were applied: $> .9$ – Excellent, $> .8$ – Good, $> .7$ – Acceptable, $> .6$ –
Questionable, $> .5$ – Poor, and $< .5$ – Unacceptable (George & Mallory,
2003). Cronbach's alpha analysis was performed using SPSS version 21.

RESULTS

Frequency of technique flaws

A total of 665 technique flaws were identified by the four raters for all criteria contained within the TJA (Table 1). The most frequent technique flaw was Criterion 2 'Thighs do not reach parallel' (N=147/665, 22%), the second most frequent technique flaw was criterion 1 'Knee valgus on landing' (N=80/665, 12%) and the least frequent technique flaw was Criterion 9 'Pause between jumps' (N=23/665, 4%).

[Insert Table 1 here: Frequency counts and relative percentages of TJA technique flaws]

A one-variable χ^2 test was conducted to measure the association between the observed and expected frequencies of flaws recorded for TJA. The χ^2 value of 152.1, DF=9 was found to have an associated probability value of 0.0001. Thus we can accept that there is a significant difference between the observed and expected frequencies (Table 2).

[Insert Table 2 here: one variable χ^2 observed and expected frequencies of TJA technique flaws]

The frequency of technique flaws within the three respective categories of the TJA ('Knee and thigh motion'; 'Foot position during landing'; 'Foot position during landing') were then analysed relative to the maximum

possible number of technique flaws possible using the following calculation:
60 participants x 4 raters=240 multiplied by the number of criteria. 'Knee and thigh motion' (N=234/720 (32.5%) from 3 criteria); 'Foot position during landing' (N=307/1200 (46%) from 5 criteria) and 'Plyometric technique' (N=64/480, from 2 criteria).

Inter-rater agreement

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

[Insert Table 2 here: Fleiss Kappa Inter-rater agreement of TJA criterion]

Weighted kappa (k_w) coefficient values used to determine inter-rater agreement of the composite score ranged from $k_w = .62$ (95% CI, .48 - .76) to $k_w = .80$ (95% CI, .70 - .90) (Table 3) suggesting a 'fair-to-very good' level of inter-rater agreement.

[Insert Table 3 here: Weighted Kappa Inter-rater agreement of TJA criterion]

Internal Consistency

1
2
3 Low alpha values were detected across all four Raters for the entire TJA
4 scale. Internal consistency was reassessed with items 9 and 10 removed
5 (plyometric technique) as the repeated plyometric nature of the TJA over a
6 10 second period differentiates it from previous tests such as the Landing
7 Error Scoring System (Padua et al 2015). As an 8 item scale there were
8 negligible alterations in internal consistency (range $\alpha = .091 - .161$, Table 4).
9 Internal consistency results suggest that the TJA scale and sub items are not
10 unidimensional
11
12
13
14
15
16
17
18
19

20
21
22 [Insert Table 4 here: internal consistency]
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

Statement of principal findings

This is the first study to investigate technique flaws associated with the TJA in elite female football players. The TJA was designed for use with athletic populations to detect technique flaws in jump landing tasks (Myer et al., 2008). In our study four raters identified 665 technique flaws in 60 participants. The most frequent flaws were 'Thighs do not reach parallel' (criterion 2) and 'Knee valgus on landing' (criterion 1), which are 2 of 3 criteria contained within the 'Knee and thigh' motion category of the TJA. The least frequent technique flaws were criterion 9 'Pause between jumps' and criterion 10 'Technique declines prior to 10 seconds'. These criteria form the 'Plyometric technique' category of the TJA. The inter-rater level of agreement for the composite score of the TJA was 'fair-to-very good' ranging from $K_w = 0.62$ (95% CI, 0.48 – 0.76) to $K_w = 0.80$ (95% CI, 0.70 – 0.90). All three 'Knee and thigh motion' criteria reached clinically acceptable levels of agreement; 'Knee valgus on landing' ($\kappa = .83$, 95% CI, .72 – .93); 'Thighs do not reach parallel' ($\kappa = .84$, 95% CI, .74 - .94) and 'Thighs not equal side to side' ($\kappa = .86$, 95% CI, .75 - .96). Low alpha values for internal consistency suggest that the TJA and the individual criteria contained within the assessment are not unidimensional.

Meaning of the study findings

Previous studies investigating the TJA have not clearly identified the frequency of individual technique flaws and this limits our ability to compare between studies. In our study criterion 2 'Thighs do not reach parallel' was

1
2
3 the most frequently identified technique flaw and 'Pause between jumps' was
4 the least frequently identified technique flaw. Dudley et al. (2013) also
5 reported criterion 2 as the most frequently identified technique flaw but did
6 not report the rank of other TJA criteria.
7
8
9
10

11
12
13
14 Herrington et al. (2013) reported the inter-rater level of agreement for the
15 composite score of the TJA using 2 Raters to be very good/excellent
16 ($K=0.88$) in a sample of ten athletes. The inter-rater percentage of exact
17 agreement between raters across all ten criteria was 93% (range 80%-100%,
18 i.e. high). Interestingly, Dudley et al. (2013) reported the inter-rater level of
19 agreement using 5 Raters to be poor in 40 recreationally active university
20 students ($ICC=0.47$, 95% CI 0.33-0.62). Read et al. (2016) used a test-retest
21 design to investigate intersession reliability of the TJA in 50 elite male youth
22 football players. Although reliability was found to be strong ($ICC=0.88$) the
23 authors suggested caution in interpreting the composite score of the TJA due
24 to high within-subject variation in a number of individual criterion.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 The difference in the reported levels of agreement may in part be explained
42 by the statistical test selected by investigators. Sample sizes of at least 50
43 are recommended when using percentage of exact agreement (Birkimer &
44 Brown, 1979). Therefore results from studies containing smaller sample
45 sizes are quite probably the result of chance agreement and should be
46 considered with caution. Each of the TJA criteria is scored in a dichotomous
47 manner i.e. flaw occurred or no flaw occurred and the data is therefore
48 characterised as nominal. Kappa coefficients are recommend for use as the
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 preferred statistical test to determine the inter-rater level of agreement for
4
5 nominal data (Hallgren, 2012). We utilised Fleiss Kappa to determine inter-
6
7 rater agreement for individual TJA criteria and a weighted Kappa to
8
9 determine inter-rater agreement for the composite score. Despite the
10
11 proposed dichotomous nature of the score, differences in the interpretation of
12
13 the occurrence of a technique flaw exist. Dudley et al. (2013) suggests that
14
15 current TJA instructions are unclear as to whether a flaw should be scored
16
17 by the rater if observed only on a single occasion or whether it needs to
18
19 occur repeatedly and consistently throughout the assessment, leading to
20
21 inconsistency of reporting between assessors. Furthermore, there is
22
23 variability in the cut-points for 'clinical acceptance' of reliability scores in
24
25 literature with scores greater than 0.70 regarded as acceptable whereas Sim
26
27 and Wright (2005) suggest that inter-rater agreement should be greater than
28
29 0.75.
30
31
32
33
34
35

36 Read et al. (2016) concluded that only the knee valgus criterion may be
37
38 reliably used to screen elite youth male football players. A prospective study
39
40 by (Hewett, Myer, Ford, et al., 2005) found increased knee abduction angles
41
42 (knee valgus) during a plyometric activity to be a significant predictor of ACL
43
44 injury. 'Knee valgus on landing' was the second most identified technique
45
46 flaw in our study and reached clinically acceptable levels of agreement.
47
48 Although no empirical evidence exists to support the premise that a TJA
49
50 composite score ≥ 6 increases an individual's risk of sustaining an ACL injury,
51
52 inadequate neuromuscular control of the trunk and hip is a contributor and
53
54 predictor of high-risk knee mechanics (Ford et al., 2006). Female athletes
55
56
57
58
59
60

1
2
3 have been found to have significantly higher hip flexion angles on initial
4 ground contact (Boden, Torg, Knowles, & Hewett, 2009) and increased trunk
5 flexion and lateral tilt (Zazulak, Hewett, Reeves, Goldberg, & Cholewicki,
6 2007) when performing jump landing tasks. ACL strain from valgus knee
7 loading has been confirmed through cadaver, in vivo and 3-dimensional
8 motion analysis methods (Fukuda et al., 2003; Kanamori et al., 2000; Markolf
9 et al., 1995). Increased internal hip rotation, coupled with increased external
10 rotation of the tibia (dynamic knee valgus) has been found in female football
11 players during jump landing and these have been used to predict ACL injury
12 (Alentorn-Geli et al., 2009; Barber-Westin et al., 2010). Female athletes have
13 been found to preferentially rely on increased quadriceps recruitment relative
14 to hamstring recruitment during incremental vertical jump test using surface
15 electromyography (Myer, Brent, Ford, & Hewett, 2011). In addition, a
16 quadriceps dominant landing strategy may increase the risk of sustaining an
17 ACL rupture (Alentorn-Geli et al., 2009; Hewett, Myer, & Ford, 2005).

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38 Cronbach's alpha is considered to infer the degree to which the criteria
39 measures a single unidimensional construct. Our internal consistency
40 statistics raise concern about the construct validity of the TJA suggesting
41 redundancy of TJA criteria. Analysis with the 'Plyometric technique' category
42 removed to determine if the psychometric properties of the test would be
43 improved as an 8 item measure found that unidimensionality remained
44 violated. Thus, neither 10 nor 8 item TJA criteria appear to be measuring a
45 specific construct i.e. jump landing tasks.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Myer et al. (2008) recommend interventions to address landing errors and
4 the risk of injury for individuals with a TJA sum score of 6 or more. A sum
5 score that is derived from multiple items is said to be a test of the same
6 latent variable and measurements of a latent variable are assumed to have
7 the same properties as measurements of observed variables (DeVellis,
8 2012). Based on results from the present study, the meaning of the TJA sum
9 score is unclear because the items are not interrelated and the evidence that
10 suggests a sum score of 6 or more increases injury risk is based on expert
11 opinion rather than empirical data. The findings of our study suggest that TJA
12 items are not internally consistent and do not have a coherent empirical
13 structure.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

30 **Implications**

31
32 There is a paucity of studies evaluating the psychometric properties of the
33 TJA. The findings of previous inter and intra-reliability studies were
34 inconsistent (Dudley et al., 2013; Herrington et al., 2013; Read et al., 2016)
35 so our finding that TJA criteria are not measuring the same underlying
36 construct i.e. jump landing task is important. Elite female football players are
37 4-6 times more likely to sustain an ACL injury than their male counterparts
38 (Alentorn-Geli et al., 2009; Myer, Brent, et al., 2011). The TJA was intended
39 for use in elite athletes and we present the first study of its use in a large
40 sample of elite female football players. Our inter-rater reliability data was
41 gathered using raters from different professional backgrounds (physiotherapy
42 & strength and conditioning) improving generalisability.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Limitations of the study

Our finding that there was a descending trend of item frequencies and kappa scores through items 1 to 10 suggests that item order impacts on recall rates. Cronbach's alpha is considered a crude measure of reliability (coefficient of reliability) and can be unduly influenced by the number of scale items and redundant items (DeVellis, 2012). Exploratory factor analysis would have provided a more in-depth assessment of the factor structure and dimensionality of the TJA. However, with such low internal consistency scores pursuing exploratory factor analysis at this stage may not have provided any further meaningful information.

Future research

There is a need to determine whether reliability and validity is improved by changing the item order or improving the clarity of instructions to assessors for the TJA. Factor analysis may help to determine whether certain items should be removed from the scale. Identifying intra-rater reliability would help to identify whether rater bias has any influential effect on the scale and if formal training in TJA administration would improve reliability of the TJA. Factor analysis techniques have been employed for measuring test validation in similar observational screening tools used in the clinical sports setting.

Conclusion

This study assessing tuck jump in elite female football players found that the inter-rater level of agreement for the composite score was 'fair-to-very good',

1
2
3 although caution should be applied in interpreting the composite score due to
4
5 poor internal consistency. Nevertheless, elements of the TJA may provide
6
7 useful information about knee mechanics and potential risk factors for knee
8
9 injury. Previous literature suggests that 'Knee valgus on landing' may be a
10
11 time efficient measure for risk factors in male youth football players and we
12
13 found that 'Knee and thigh motion' reached clinically acceptable levels of
14
15 agreement in a sample of elite female football players. Therefore, we
16
17 recommend the use of the 'Knee and thigh motion' criteria of the TJA for
18
19 screening athletic populations particularly elite football players of either
20
21 gender.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- Alentorn-Geli, E., Myer, G. D., Silvers, H. J., Samitier, G., Romero, D., Lazaro-Haro, C., & Cugat, R. (2009). Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 1: Mechanisms of injury and underlying risk factors. *Knee Surg Sports Traumatol Arthrosc*, *17*(7), 705-729. doi: 10.1007/s00167-009-0813-1
- Barber-Westin, S. D., Smith, S. T., Campbell, T., & Noyes, F. R. (2010). The drop-jump video screening test: retention of improvement in neuromuscular control in female volleyball players. *J Strength Cond Res*, *24*(11), 3055-3062. doi: 10.1519/JSC.0b013e3181d83516
- Birkimer, J. C., & Brown, J. H. (1979). Back to basics: Percentage agreement measures are adequate, but there are easier ways. *J Appl Behav Anal*, *12*(4), 535-543.
- Boden, B. P., Torg, J. S., Knowles, S. B., & Hewett, T. E. (2009). Video analysis of anterior cruciate ligament injury: abnormalities in hip and ankle kinematics. *Am J Sports Med*, *37*(2), 252-259. doi: 10.1177/0363546508328107
- DeVellis, R. F. (2012). *Scale Development: Theory and Applications* (3rd edition). Los Angeles, CA: Sage.
- Dudley, L. A., Smith, C. A., Olson, B. K., Chimera, N. J., Schmitz, B., & Warren, M. (2013). Interrater and Intrarater Reliability of the Tuck Jump Assessment by Health Professionals of Varied Educational Backgrounds. *J Sports Med (Hindawi Publ Corp)*, *2013*, 483503. doi: 10.1155/2013/483503
- Faude, O., Junge, A., Kindermann, W., & Dvorak, J. (2006). Risk factors for injuries in elite female soccer players. *Br J Sports Med*, *40*(9), 785-790. doi: 10.1136/bjism.2006.027540
- Ford, K. R., Manson, N. A., Evans, B. J., Myer, G. D., Gwin, R. C., Heidt, R. S., Jr., & Hewett, T. E. (2006). Comparison of in-shoe foot loading patterns on natural grass and synthetic turf. *J Sci Med Sport*, *9*(6), 433-440. doi: 10.1016/j.jsams.2006.03.019
- Frohm, A., Heijne, A., Kowalski, J., Svensson, P., & Myklebust, G. (2012). A nine-test screening battery for athletes: a reliability study. *Scand J Med Sci Sports*, *22*(3), 306-315. doi: 10.1111/j.1600-0838.2010.01267.x
- Fukuda, Y., Woo, S. L., Loh, J. C., Tsuda, E., Tang, P., McMahon, P. J., & Debski, R. E. (2003). A quantitative analysis of valgus torque on the ACL: a human cadaveric study. *J Orthop Res*, *21*(6), 1107-1112. doi: 10.1016/s0736-0266(03)00084-6
- Giza, E., Mithofer, K., Farrell, L., Zarins, B., & Gill, T. (2005). Injuries in women's professional soccer. *Br J Sports Med*, *39*(4), 212-216; discussion 212-216. doi: 10.1136/bjism.2004.011973
- Hagglund, M., Walden, M., & Atroshi, I. (2009). Preventing knee injuries in adolescent female football players - design of a cluster randomized controlled trial [NCT00894595]. *BMC Musculoskelet Disord*, *10*, 75. doi: 10.1186/1471-2474-10-75

- Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*, 8(1), 23-34.
- Herrington, L., Myer, G. D., & Munro, A. (2013). Intra and inter-tester reliability of the tuck jump assessment. *Phys Ther Sport*, 14(3), 152-155. doi: 10.1016/j.ptsp.2012.05.005
- Hewett, T. E., Myer, G. D., & Ford, K. R. (2005). Reducing knee and anterior cruciate ligament injuries among female athletes: a systematic review of neuromuscular training interventions. *J Knee Surg*, 18(1), 82-88.
- Hewett, T. E., Myer, G. D., Ford, K. R., Heidt, R. S., Jr., Colosimo, A. J., McLean, S. G., . . . Succop, P. (2005). Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med*, 33(4), 492-501. doi: 10.1177/0363546504269591
- Kanamori, A., Woo, S. L., Ma, C. B., Zeminski, J., Rudy, T. W., Li, G., & Livesay, G. A. (2000). The forces in the anterior cruciate ligament and knee kinematics during a simulated pivot shift test: A human cadaveric study using robotic technology. *Arthroscopy*, 16(6), 633-639. doi: 10.1053/jars.2000.7682
- Kiesel, K., Plisky, P. J., & Voight, M. L. (2007). Can Serious Injury in Professional Football be Predicted by a Preseason Functional Movement Screen? *N Am J Sports Phys Ther*, 2(3), 147-158.
- Klugman, M. F., Brent, J. L., Myer, G. D., Ford, K. R., & Hewett, T. E. (2011). Does an in-season only neuromuscular training protocol reduce deficits quantified by the tuck jump assessment? *Clin Sports Med*, 30(4), 825-840. doi: 10.1016/j.csm.2011.07.001
- Le Gall, F., Carling, C., & Reilly, T. (2008). Injuries in young elite female soccer players: an 8-season prospective study. *Am J Sports Med*, 36(2), 276-284. doi: 10.1177/0363546507307866
- Markolf, K. L., Burchfield, D. M., Shapiro, M. M., Shepard, M. F., Finerman, G. A., & Slauterbeck, J. L. (1995). Combined knee loading states that generate high anterior cruciate ligament forces. *J Orthop Res*, 13(6), 930-935. doi: 10.1002/jor.1100130618
- Myer, G. D., Brent, J. L., Ford, K. R., & Hewett, T. E. (2011). Real-time assessment and neuromuscular training feedback techniques to prevent ACL injury in female athletes. *Strength Cond J*, 33(3), 21-35. doi: 10.1519/SSC.0b013e318213afa8
- Myer, G. D., Ford, K. R., & Hewett, T. E. (2008). Tuck Jump Assessment for Reducing Anterior Cruciate Ligament Injury Risk. *Athl Ther Today*, 13(5), 39-44.
- Myer, G. D., Ford, K. R., Khoury, J., Succop, P., & Hewett, T. E. (2011). Biomechanics laboratory-based prediction algorithm to identify female athletes with high knee loads that increase risk of ACL injury. *Br J Sports Med*, 45(4), 245-252. doi: 10.1136/bjism.2009.069351
- Nilstad, A., Andersen, T. E., Bahr, R., Holme, I., & Steffen, K. (2014). Risk factors for lower extremity injuries in elite female soccer players. *Am J Sports Med*, 42(4), 940-948. doi: 10.1177/0363546513518741
- Padua, D. A., Marshall, S. W., Boling, M. C., Thigpen, C. A., Garrett, W. E., Jr., & Beutler, A. I. (2009). The Landing Error Scoring System (LESS) Is a valid and reliable clinical assessment tool of jump-landing

- 1
2
3 biomechanics: The JUMP-ACL study. *Am J Sports Med*, 37(10), 1996-
4 2002. doi: 10.1177/0363546509343200
- 5 Read, P. J., Oliver, J. L., de Ste Croix, M. B., Myer, G. D., & Lloyd, R. S.
6 (2016). Reliability of the Tuck Jump Injury Risk Screening Assessment
7 in Elite Male Youth Soccer Players. *J Strength Cond Res*, 30(6),
8 1510-1516. doi: 10.1519/JSC.0000000000001260
- 9
10 Renstrom, P., Ljungqvist, A., Arendt, E., Beynnon, B., Fukubayashi, T.,
11 Garrett, W., . . . Engebretsen, L. (2008). Non-contact ACL injuries in
12 female athletes: an International Olympic Committee current concepts
13 statement. *Br J Sports Med*, 42(6), 394-412. doi:
14 10.1136/bjsm.2008.048934
- 15 Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use,
16 interpretation, and sample size requirements. *Phys Ther*, 85(3), 257-
17 268.
- 18
19 Walden, M., Hagglund, M., Magnusson, H., & Ekstrand, J. (2011). Anterior
20 cruciate ligament injury in elite football: a prospective three-cohort
21 study. *Knee Surg Sports Traumatol Arthrosc*, 19(1), 11-19. doi:
22 10.1007/s00167-010-1170-9
- 23 Zazulak, B. T., Hewett, T. E., Reeves, N. P., Goldberg, B., & Cholewicki, J.
24 (2007). The effects of core proprioception on knee injury: a
25 prospective biomechanical-epidemiological study. *Am J Sports Med*,
26 35(3), 368-373. doi: 10.1177/0363546506297909
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tables & Figures

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

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

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), $p < .000$

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	$\kappa_w = .65$ (95% CI, .51 – .79)
PT ₁ : SC ₁	$\kappa_w = .80$ (95% CI, .70 – .90)
PT ₁ : SC ₂	$\kappa_w = .67$ (95% CI, .54 – .80)
PT ₂ : SC ₁	$\kappa_w = .70$ (95% CI, .54 – .84)
PT ₂ : SC ₂	$\kappa_w = .79$ (95% CI, .69 – .88)
SC ₁ : SC ₂	$\kappa_w = .62$ (95% CI, .48 – .76)

Abbreviations: TJA: tuck jump assessment,

PT: physiotherapist, SC: strength &

conditioning coach

Table 4: Internal consistency

Cronbach's Alpha (α)	Rater 1 (PT ₁)	Rater 2 (PT ₂)	Rater 3 (SC ₁)	Rater 4 (SC ₁)
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