Original research

Intra and inter-tester reliability of the tuck jump assessment

Lee Herrington,*, Gregory D. Myer, Allan Munro

Objective: To assess the inter-tester and intra-tester reliability of the tuck jump test.

Design: Repeated measures.

Setting: University Human Performance laboratory.

Participants: Five male and 5 female athletes undertook the Tuck jump test which was then assessed by two independent assessors.

Main outcome measures: Score from the video assessment of the tuck jump test by two independent assessors on two separate occasions.

Results: Average percentage of exact agreement (PEA) between the two testers across all scoring criteria for all subjects was 93% (range 80–100%). Both testers were in absolute 100% agreement in 5 out of 10 subjects for all of the scoring criteria. The kappa measure of agreement was \( k = 0.88 \) which is very good/excellent. The intra-tester PEA ranged 87.2%–100%, with kappa values of \( k = 0.86–1.0 \).

Conclusion: The study showed very good/excellent inter-tester and inter-tester reliability for both examiners when comparing their individual scores of the tuck jump test across two analysis sessions. These findings indicate that the proposed tuck jump assessment is reliable to identify abnormal landing mechanics.

*Corresponding author. Tel.: +441612952326.
E-mail address: l.c.herrington@salford.ac.uk (L. Herrington).

1. Introduction

Analysis of athletes’ motion during specific tasks may enable the identification of athletes at risk of injury because of poor movement patterns, such as knee valgus. Chmielewski et al. (2007) studied the reliability of using observational rating to assess frontal plane trunk and lower limb motion during a unilateral squat and a lateral step-down task; while Krosshaug, Nakamae, Boden (2007) investigated the ability of raters to estimate subjects’ hip, knee, and ankle joint angles during running and cutting trials at predetermined points on a video. Both of these studies found poor intra and inter-rater reliability, and previous authors proposed that their rating guidelines were limited by the nondiscrete and subjective criteria (Ekegren, Miller, Celebrini, Eng, & Macintyre, 2009). Subsequent studies using more defined rating criteria have shown superior intra-tester and inter-tester reliability (Ekegren et al., 2009; Padua et al., 2009).

The studies of both Ekegren et al. (2009) and Padua et al. (2009) subjectively rated subjects knee motion on landing during the drop jump landing, a test which had been previously shown to be sensitive in detecting ACL injury risk when undertaken using 3-D motion capture (Ford, Myer, & Hewett, 2003; Hewett et al., 2005). This test assesses the landing phase of a single bilateral loading task and could be regarded as a low level test, when compared to...
undertaking multiple and unilateral limb landings. Myer, Ford, and Hewett (2008) proposed a tuck jump test as a higher effort-load test of limb control and movement. The tuck jump exercise may be useful as it enables the identification of lower extremity landing technique flaws during a repetitive plyometric activity. The tuck jump test has received little attention in the literature; with only Myer, Paterno, Ford, Quatman, and Hewett (2006) reporting on the tests intra-rater reliability \(r = 0.84\). This original scale employed continuous measures of each criterion which may have limited the potential for optimized inter-rater reliability. More recently, the tuck jump assessment has been developed with dichotomized criteria to potentially improve both intra- and inter-rater reliability as well as improve the clinical utility of the assessment (Myer, Brent, Ford, & Hewett, 2011; Myer, Ford et al., 2008). The objective of this paper is to assess the inter-tester and intra-tester reliability of the new assessment tool, to determine if it shows similar reliability to other qualitative assessment methods of lower extremity landing techniques. The aim of this study is to test the hypothesis that the Tuck jump test will show excellent intra and inter-tester reliability as evidenced by minimal differences in scores being present between examiners and session. If the test proves to have strong reliability then it gives the clinician another testing option when looking to assess lower limb alignment control, especially if they wish to test ability during a repetitive bilateral plyometric activity.

2. Method

2.1. Subjects

Two testers (LH & GM) independently viewed and scored 5 female & 5 male subjects’ recorded video performance of the tuck jump test. All the participants videoed were free from lower limb, pelvic or spinal injury and gave informed consent to participate in the study which was approved by the university research ethics committee. The participant group comprised of five male and five female subjects all recruited from the sports science course of the host university (mean age 19.3 ± 0.8 years range 18–21 years, mean height 175(±10) cm, mean weight 70.8(±10.2) kg), who were all physically active participating in a minimum of 3 h aerobic exercise per week.

2.2. Procedure

2.2.1. Tuck jump test

Each participant carried out continuous tuck jumps on the spot for 10 s after being given basic instructions about how to carry out the test. This included information on lifting the knees to hip height and attempting to land on the same spot and a demonstration of the exercise. They carried out this activity once that is 10 s of jumping only. Their performance of the test was videoed using two cameras (Sony Handycam DCR-HC37) on tripods with height set to that of the participants waist, one aligned 2 m away in the sagittal plane (side view of subject) and one aligned 2 m away in the frontal plane (in front of subject).

2.2.2. Scoring criteria

Each participants video performance of the tuck jump test was scored across 10 criteria grouped into three areas (knee & thigh motion; foot position during landing and plyometric technique). The individual criteria along with the scoring sheet are shown in Fig. 1. If the participant failed to meet the criteria on any given repetition described in Fig. 1 then they scored 1, if they met the criteria they scored 0 for the respective category. The lowest scores indicate a superior performance in this case.

2.3. Analysis

Two raters (LH & GM) independently viewed and scored 5 female & 5 male subjects’ recorded video performance of the tuck jump test. When scoring performance, each film (sagittal and frontal plane views) was viewed three times at normal speed and the score was then marked. Another investigator (AM) masked to the identity of the examiners, compared the scores and then analysed the scores from both testers for each participant. The two raters then re-examined the same videos 1 month later, blind to original scores and the third investigator then analysed these scores. The scores were analysed for percentage of exact agreement (PEA) \[\text{PEA} = \left(\frac{\text{agreed} + \text{agreed} \times 100}{\text{agreed} + \text{agreed} + \text{disagreed}}\right)\] and Kappa coefficient for both intra and inter-tester reliability. The equation for \(\kappa\) is:

\[
\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}
\]

where \(\Pr(a)\) is the relative observed agreement among raters, and \(\Pr(e)\) is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then \(\kappa = 1.0\). If there is no agreement among the raters other than what would be expected by chance (as defined by \(\Pr(e)\)), \(\kappa = 0\) (Haley & Osberg, 1989).

The kappa coefficient was interpreted based on the scale of Landis and Koch (1977) with 0.01–0.2 being slight, 0.21–0.4; fair, 0.41–0.6; moderate, 0.61–0.8; good and 0.81–1.0 almost perfect (excellent).

3. Results

3.1. Overall

3.1.1. Inter-tester

Average percentage exact agreement (PEA) between the two testers across all scoring criteria for all subjects was 93% (range 80–100%). Both testers were in absolute 100% agreement in 5 out of 10 subjects for all of the scoring criteria. The kappa measure of agreement was \(k = 0.88\) which is very good/excellent. In three cases the testers disagreed by a single point (scoring criteria) once on Q2 and twice on Q9, in two cases the testers disagreed by two points (scoring criteria) once on Q2 & Q3 and once on Q6 & Q7. Therefore out of the seven cases of disagreement (out of 100) the testers failed to agree twice each on criteria Q2 and Q9 and once on Q3, Q6 and Q7.

3.2. Male subjects

3.2.1. Inter-tester

Average percentage exact agreement (PEA) between the two testers across all scoring criteria for male subjects was 92.7% (range 80–100%). Both testers were in absolute 100% agreement in 3 out of 5 cases for all of the scoring criteria. In two cases the testers disagreed by a single point (scoring criteria) once on Q2 and once on Q9. The kappa measure of agreement was \(k = 0.86\) which is very good/excellent.

3.2.2. Intra-tester

The average PEA for tester 1 (GM) across the two viewing sessions was 100% for male subjects, Kappa measure of agreement was \(k = 1.0\) which is excellent agreement. For tester 2 (LH) there was a 96% average PEA, disagreeing on two points one for subject 2
Marking criteria
If the participant fails to meet the criteria below then they score 1, if they meet the
criteria they score 0 for the respective category.

Knee & Thigh Motion
1. Knee valgus on landing
   o Hip, knee and foot aligned, no collapse of the
     knee inwards
2. Thighs not reaching parallel (peak of jump)
3. Thighs not equal side to side (during flight)

Foot position during landing
4. Foot placement not shoulder width apart
   o Inside of tape marks
5. Foot placement not parallel (front to back)
6. Foot contact timing not equal
   o Asymmetrical landing
7. Does not land in same foot print
   o Consistent point of landing
8. Excessive landing contact noise

Plyometric technique
9. Pause between jumps
10. Technique declines prior to 10seconds

<table>
<thead>
<tr>
<th>Name:</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knee &amp; thigh motion</td>
<td></td>
</tr>
<tr>
<td>Knee valgus on landing</td>
<td></td>
</tr>
<tr>
<td>Thighs not reaching parallel (peak of jump)</td>
<td></td>
</tr>
<tr>
<td>Thighs not equal side to side (during flight)</td>
<td></td>
</tr>
<tr>
<td>Foot position during landing</td>
<td></td>
</tr>
<tr>
<td>Foot placement not shoulder width apart</td>
<td></td>
</tr>
<tr>
<td>Foot placement not parallel (front to back)</td>
<td></td>
</tr>
<tr>
<td>Foot contact timing not equal</td>
<td></td>
</tr>
<tr>
<td>Does not land in same foot print</td>
<td></td>
</tr>
<tr>
<td>Excessive landing contact noise</td>
<td></td>
</tr>
<tr>
<td>Plyometric technique</td>
<td></td>
</tr>
<tr>
<td>Pause between jumps</td>
<td></td>
</tr>
<tr>
<td>Technique declines prior to 10 seconds</td>
<td></td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Tuck jump test scoring criteria.
(Q2) and one for subject 5 (Q9). The kappa measure of agreement was $k = 0.86$ which is very good/excellent.

3.3. Female subjects

3.3.1. Inter-tester

Average percentage exact agreement (PEA) between the two testers across all scoring criteria for female subjects was 87.2% (range 75–100%). Both testers were in absolute 100% agreement in 2 out of 5 cases for all of the scoring criteria. In two cases the testers disagreed by two points (scoring criteria) on Q2 & Q3 and once on Q6 & Q7, in 1 case the two testers disagreed by single point on Q9. The kappa measure of agreement was $k = 0.86$ which is very good/excellent across all tests.

3.3.2. Intra-tester

The average PEA for tester 1 across the two viewing sessions was 100% for female subjects, for tester 2 it was 96% disagreeing on two points one for subject 3 (Q6) and one for subject 4 (Q7). The kappa measure of agreement was $k = 0.81$ which is very good/excellent across all tests.

4. Discussion

Laboratory based screening tools demonstrate that altered neuromuscular strategies or decreased neuromuscular control during the execution of sports movements, as evidenced by abnormal lower limb joint mechanics (motions and loads), may during the execution of sports movements, as evidenced by neuromuscular strategies or decreased neuromuscular control across all tests.

A measure of agreement was $k = 100\%$ for female subjects, for tester 2 it was 96% disagreeing on two out of 5 cases for all of the scoring criteria. In two cases the testers disagreed by single point on Q6 & Q7. The kappa measure of agreement was $k = 0.81$ which is very good/excellent across all tests.

5. Conclusion

The research undertaken demonstrated that the tuck jump test shows very good—excellent intra and inter-tester reliability when the test is analysed from video. Future work is required to assess this test in real time and also its sensitivity in highlighting those individuals at risk of knee injury and ability to detect performance improvements.

Conflict of interest statement
None declared.

Ethical statement
The project was approved by the University of Salford Research ethics committee.

All subjects signed consent documents to participate.

Funding
None declared.

References


