A Critically Appraised Topic on the Tuck Jump Assessment: Does the Tuck Jump Assessment Demonstrate Interrater and Intrarater Reliability in Healthy Individuals?

Marissa L. Mason, Marissa N. Clemons, Kaylyn B. LaBarre, Nicole R. Szymczak, and Nicole J. Chimera

Clinical Scenario: Lower-extremity injuries in the United States costs millions of dollars each year. Athletes should be screened for neuromuscular deficits and trained to correct them. The tuck jump assessment (TJA) is a plyometric tool that can be used with athletes. Clinical Question: Does the TJA demonstrate both interrater and intrarater reliability in healthy individuals? Summary of Key Findings: Four of the 5 articles included in this critically appraised topic showed good to excellent reliability; however, caution should be taken in interpreting these results. Although composite scores of the TJA were found to be reliable, individual flaws do not demonstrate reliability on their own, with the exception of knee valgus at landing. Aspects of the TJA itself, including rater training, scoring system, playback speed, volume, and number of views allotted, need to be standardized before the reliability of this clinical assessment can be further researched. Clinical Bottom Line: The TJA has shown varying levels of reliability, from poor to excellent, for both interrater and intrarater reliability, given current research. Strength of Recommendation: According to the Centre for Evidence Based Medicine levels of evidence, there is level 2b evidence for research into the reliability of the TJA. This evidence has been demonstrated in elite, adolescent, and college-level athletics in the United Kingdom, Spain, and the United States. The recommendation of level 2b was chosen because these studies utilized cohort design for interrater and intrarater reliability across populations. An overall grade of B was recommended because there were consistent level 2 studies.

Keywords: rehabilitation, sport management, sports physiotherapist, tuck jump assessment, testing and measurement

Clinical Scenario

Lower-extremity (LE) injuries are common in the athletic population, especially injuries to the anterior cruciate ligament (ACL). These injuries can be career ending for athletes and result in costs in the United States of approximately 1 billion dollars annually. The large number of LE injuries and ACL injuries in particular may be able to be reduced with proper screening, which could reduce costs to the US health care system. Currently, plyometric screening tests include the landing error scoring system, the drop jump video screening test, and the tuck jump assessment (TJA). The TJA is a clinician-friendly tool that requires minimal equipment, is time efficient (it only takes 10 s to administer), and evaluates jumping performance for 10 flaws. Two-dimensional video (frontal and sagittal) is taken of an individual performing repeated tuck jumps for 10 seconds, and the rater watches the videos to retrospectively score the TJA for the presence or absence of each flaw. If a flaw is present one or more times, a score of 1 is given, and if a flaw is absent during the 10 seconds, a score of 0 is given. An athlete may earn a score between 0 and 10; a score of 10 represents the poorest neuromuscular performance, with all flaws present, and a score of 0 represents the best performance, with no flaws present.

Focused Clinical Question

Does the TJA demonstrate both interrater and intrarater reliability in healthy individuals?

Summary of Search, Best Evidence Appraised, and Key Findings:

- Five cohort studies were found in the literature review that met the inclusion criteria for this critically appraised topic.
- Three studies involved participants who were healthy, college-aged athletes, and 2 studies involved healthy youth athletes.
- Two of the studies provided a full description of the raters. Dudley et al included 5 raters: a physical therapist with a Doctor of Physical Therapy and 4 years of clinical experience, a certified strength and conditioning coach with 7 years of clinical experience, a certified athletic trainer with 17 years of clinical experience, a third-year Doctor of Physical Therapy student, and a first-year Doctor of Physical Therapy student. Mayhew et al included 4 raters: 2 physical therapists with 10 and 11 years of experience, respectively, and 2 strength and conditioning coaches with 1 and 11 years of experience, respectively. Another study stated that the raters were certified strength and conditioning coaches with more than 5 years of experience each. In addition, 2 of the studies did not mention any education of the raters or related clinical experience.
- Interrater reliability was determined in 4 of the studies and intrarater reliability was also determined in 4 studies.

Clinical Bottom Line

There is insufficient evidence to determine the reliability of the TJA in healthy individuals. According to Centre for Evidence Based Medicine levels of evidence, there is level 2b evidence for research.
into the reliability of the TJA. This evidence has been demonstrated in elite, adolescent, and college-level athletics in the United Kingdom, Spain, and the United States.

Search Strategy

Terms Used to Guide Search Strategy
- Patient/Client group: healthy subjects OR athletes OR collegiate athletes
- Intervention: tuck jump assessment OR TJA
- Comparison: no comparison
- Outcomes: reliability OR interrater reliability OR intrarater reliability OR like term

Sources of Evidence Searched
- PubMed
- MEDLINE
- CINAHL
- SPORTDiscus

Inclusion and Exclusion Criteria

Inclusion Criteria
- Studies with healthy participants
- English language studies published from May 2012 to October 2018
- Oxford Centre for Evidence-Based Medicine rating of 2b or higher
- Articles that have been peer reviewed

Exclusion Criteria
- Studies that did not evaluate interrater and/or intrarater reliability

Results of Search

Five relevant studies were found and are presented in Table 1. Interrater reliability was determined in 4 of the studies,8-12 and intrarater reliability was also determined in 4 of the studies.8,9,11,12

Best Evidence

The 5 studies chosen for this critically appraised topic were identified as the best match when considering the inclusion and exclusion criteria. These 5 studies investigated interrater and intrarater reliability (Table 2).

Table 1 Summary of Study Designs of Articles

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Study design</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2b</td>
<td>Cohort study</td>
<td>Herrington et al8</td>
</tr>
<tr>
<td>2b</td>
<td>Cohort study</td>
<td>Dudley et al9</td>
</tr>
<tr>
<td>2b</td>
<td>Cohort study</td>
<td>Mayhew et al10</td>
</tr>
<tr>
<td>2b</td>
<td>Cohort study</td>
<td>Read et al11</td>
</tr>
<tr>
<td>2b</td>
<td>Cohort study</td>
<td>Fort-Vanneerhaeghe et al12</td>
</tr>
</tbody>
</table>

Implications for Practice, Education, and Future Research

Musculoskeletal conditions were ranked as the third-highest expenditure in health care spending in 2013, accounting for 190 billion health care dollars spent that year.13 A large portion of musculoskeletal care involves treatment of lower extremity (LE) injuries. Of all sports-related injuries, 66% are LE injury injuries, with knee injuries being most common.14 A study conducted by Herzog et al15 found that between 2005 and 2013 the immediate cost for 229,446 ACL surgeries was over 2 billion dollars. This statistic does not account for the rehabilitation process following the surgery. ACL injuries can be career-ending injuries for many athletes. Therefore, athletes’ risk for LE injury should be recognized, treated, and prevented via clinical assessment tools and strength and endurance training programs. This will not only prevent devastating consequences to athletic careers but also assist in reducing health care costs in our nation.

To reduce the risk of career-ending ACL injuries, several plyometric screening tools have been developed, including the landing error scoring system,3 the drop jump video screening test,4 and the TJA.5-7 Tools like the landing error scoring system and the drop jump video screening are more time consuming than the TJA, which only takes 10 seconds to conduct. The TJA has been deemed clinician friendly by researchers of the tool because it requires minimal equipment.9,15

The TJA starts with camera set up in both the frontal and sagittal planes.7 Two pieces of tape on the ground, 35 cm apart, indicate the starting and finishing point for the athlete’s feet. Myers et al7 described that the athlete should jump straight up and at the highest point of the jump bring the knees as high as possible. The jumper should land with the same foot placement after each jump and land softly each time. The effort of the athlete should not decline prior to the completion of the 10 seconds. The instructions to raters differ between studies. However, all 5 raters in the study by Dudley et al9 received the same instructions, videos, and original Myer et al7 article that describes the scoring. The raters in the study by Dudley et al9 were allowed to watch the videos as many times as necessary. The raters scored each video independently, with no discussion allowed.

Scoring of the TJA is quick but not necessarily simple, with 10 flaws to score. The 10 flaws include LE valgus at landing, thighs not reaching parallel, thighs not equal side to side, foot placement not shoulder width apart, foot placement not parallel (front to back), foot contact timing not equal, excessive landing contact noise, pause between jumps, technique declines prior to 10 seconds, and does not land in the same foot placement.5-7 The different flaws must be viewed from either the frontal or sagittal plane and with the volume on. It can be difficult to discern whether a flaw is present or not, especially when rating videos at full speed. Because of the difficulty of scoring a flaw, scorers are likely to slow down the video replay speed in order to view the flaws unless given specific instructions not to. To maintain consistency between scoring attempts, there is a need for standardization of video playback speed because some flaws may be better viewed—and, therefore, better scored—at different speeds. There is also a need for standardization for the video volume because one of the flaws is excessive contact noise. Currently, there is no recommendation on the video playback speed or volume used when viewing the videos.

Two different scoring systems exist for the TJA, the original version and the modified version.5-7,12 The original scoring system
Participants: Participants included a total of 50 elite male soccer players who were split into 2 groups: PHV, also known as the time of adolescent growth spurt, because it has been concluded that athletes are more susceptible to injuries during and post-PHV than they are pre-PHV.

Participants: 60 elite female football players participated in the study. Participants were familiarized with testing procedures 1 wk prior to testing. Participants were instructed to place their feet in the middle of a rectangle on the floor.

Table 2 Characteristics of Included Studies

<table>
<thead>
<tr>
<th>Design</th>
<th>Article</th>
<th>Herrington et al²⁰</th>
<th>Dudley et al²⁰</th>
<th>Mayhew et al¹⁰</th>
<th>Read et al¹¹</th>
<th>Fort-Vanmeerhaeghe et al¹²</th>
</tr>
</thead>
</table>
| Intervention  | Participants: 10 physically active participants including 5 males and 5 females, aged 18–21 y old. Participants were given instructions and a demonstration of the TJA. The instructions to the participants included lifting knees to parallel and landing in the same spot as takeoff. Warm-up: not provided. Number of jumps performed: 1 round of 10 seconds. Camera setup: 2 video cameras were placed 2 m away from the participant in both the frontal and sagittal planes to view all flaws. Camera speed was not provided. Raters: To determine interrater reliability, the raters watched each video 3 times at normal speed, watching for the 10 flaws to determine interrater reliability. They watched the videos again 1 mo later to assess intrarater reliability. | Cohort study | Participants: 10 participants, 5 males and 5 females, were given instructions and a demonstration of the TJA. Participants were instructed on the purpose and protocol of the TJA. Each participant was verbally instructed on the TJA and were given a demonstration. The instructions were to stand with 1 foot on the tape and perform the tuck jump with high effort for 10 s. They were told to lift their knees to be level with their hips and return to the start position after every jump. Warm-up: not provided. Number of jumps performed: 1 round of 10 seconds. Camera setup: 2 high definition cameras, one in the sagittal plane and other in the frontal plane. Distance from jumper and speed of camera not provided. Raters: The raters were instructed to view each video no more than 3 times prior to scoring, and scores were to be determined independently from other raters. | Cohort study | Participants: 60 elite female football players participated in the study. These females were healthy and had a mean age of 20.27 (3.44) y. Raters: The raters who scored the TJA were a convenience sample of SC coaches and PT. The first PT had 5 y of experience with the women’s FA and 10 y of clinical experience. PT 2 had a total of 11 y of clinical experience, including 3 y with the FA. The first SC had a total of 11 y of experience and 8 y of experience with the men’s FA. Finally, the second SC had 1 y of experience with football players. | Cohort study | Participants: Participants included a total of 50 elite male soccer players who were split into 2 groups: 25 pre-peak height velocity (PHV) males with an average age of 11.93 y and 25 post-PHV males with an average age of 17.26 y. The groupings were based on the age of the adolescent growth spurt, because it has been concluded that athletes are more susceptible to injuries during and post-PHV than they are pre-PHV. PHV was determined using individuals’ maturity offset value, which allowed researchers to predict when an individual would reach PHV based on his gender, date of birth, standing and sitting height, and weight. All participants were healthy and active in soccer training and competitions. Raters: Authors mentioned that there was a single rater who scored the videos, but there was no description provided by the authors regarding the training or experience of the rater. | Cohort study | Participants: 24 elite youth volleyball athletes, 12 males and 12 females. Athletes were excluded if they had overuse or acute injuries at the time of testing. Raters: Both of the raters were certified SC coaches with more than 5 y of experience each. | (continued)
Interrater reliability was determined to $\kappa = 0.46$; 95% CI, the 2 raters was good to excellent with $\kappa = 0.35$ – $\kappa = 0.52$; 95% CI, .35 – .68, and the second scoring session improved to ICC = 0.69; 95% CI, 0.55 – 0.81. Interrater reliability ranged from poor (0.44; 95% CI, 0.22 – 0.68) to moderate (0.72; 95% CI, 0.55 – 0.84). The most frequently identified and agreed-upon flaw by the raters was thighs not reaching parallel, and the average number of flaws identified was 6.30 (1.76). The interrater reliability between the 5 raters was poor (ICC = 0.47; 95% CI, 0.33 – 0.62), but interrater reliability between the 3 raters improved for the second scoring session. The first session interrater reliability results were ICC = 0.52; 95% CI, 0.35 – 0.68, and the second scoring session improved to ICC = 0.69; 95% CI, 0.55 – 0.81. Interrater reliability ranged from poor (0.44; 95% CI, 0.22 – 0.68) to moderate (0.72; 95% CI, 0.55 – 0.84).

### Table 2 (continued)

<table>
<thead>
<tr>
<th>Article</th>
<th>Primary outcome: interrater reliability and intrarater reliability</th>
<th>Primary outcome: interrater reliability and intrarater reliability</th>
<th>Primary outcome: interrater reliability</th>
<th>Primary outcome: intrarater reliability</th>
<th>Primary outcome: interrater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of evidence</td>
<td>Level 2</td>
<td>Level 2</td>
<td>Level 2</td>
<td>Level 2</td>
<td>Level 2</td>
</tr>
<tr>
<td>Strobe</td>
<td>15/22</td>
<td>16/22</td>
<td>17/22</td>
<td>16/22</td>
<td>15/22</td>
</tr>
<tr>
<td>Conclusion</td>
<td>The TJA has very good to excellent interrater and intrarater reliability. No information was given with regard to the training of the raters in this study. One of the raters in this study is one of the founders of the TJA, which could explain the lack of information regarding the raters. This study also had the smallest number of participants completing the tuck jump. Speed and number of views allowed by the raters were consistent with 3 views per video at normal speed, minimizing the risk of viewers watching the videos at different speeds or different numbers of times. Authors demonstrated poor interrater reliability. This could be attributed to the fact that the raters in the study all had different levels of clinical experience. The authors did try to create some consistency in the training of the raters by requiring that they read an article by an original creator of the TJA and score a practice video; however, there has to be a more standardized training for the raters. In addition, the authors did not standardize the playback speed or how many times the raters could watch the videos. The TJA demonstrated strong intrarater reliability. The researchers also suggested that variation in test scores were likely due to variability within the performance of participants across trials rather than the examiner. Issues arise with this study due to the fact that no background information was given for the rater. Therefore, readers cannot determine if this principal rater was proficient in scoring the TJA. The researchers also pointed out that knee valgus is the only flaw that was reproducible among both participant groups; they suggested that not all 10 of the TJA flaws were reliable for determining an athlete’s risk of ACL injury. Therefore, the researchers suggested that only the flaw of knee valgus should be used to determine athletes’ risk for LE injury. The modified TJA demonstrated scores of good to excellent for almost all of the individual criteria assessed, and the percentage of exact agreement of interrater and intrarater reliability was excellent across all criteria. However, the authors failed to elaborate on the training provided to the raters regarding the grading of the modified TJA, which is important for replicating results of the same quality in the future.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The authors have no conflicts of interest to disclose.
