

Modified Tuck Jump Assessment: Reliability and Training of Raters

Dear Editor-in-chief

We are writing with regard to “Intra- and inter-rater reliability of the modified tuck jump assessment,” by Fort-Vanmeerhaeghe et al. (2017) published in the *Journal of Sports Science & Medicine*. The authors reported on the reliability of the modified Tuck Jump Assessment (TJA). The purpose of the article was twofold: to introduce a new scoring methodology and to report on the interrater and intrarater reliability. The authors found the modified TJA to have excellent interrater reliability (ICC = 0.94, 95% CI = 0.88-0.97) and intrarater reliability (rater 1 ICC = 0.94, 95% CI = 0.88-0.9; rater 2 ICC = 0.96, 95% CI = 0.92-0.98) with experienced raters ($n = 2$) in a sample of 24 elite volleyball athletes. Overall, we found the study to be well conducted and valuable to the field of injury screening; however, the study did not adequately explain how the raters were trained in the modified TJA to improve consistency of scoring, or the modifications of the individual flaw “excessive contact noise at landing.” This information is necessary to improve the clinical utility of the TJA and direct future reliability studies.

The TJA has been changed at least three times in the literature: from the initial introduction (Myer et al., 2006) to the most referenced and detailed protocol (Myer et al., 2011) to the publication under discussion (Fort-Vanmeerhaeghe et al., 2017). The initial test protocol was based upon clinical expertise and has evolved over time as new research emerged and problems arose with the original TJA. Initially, the TJA was scored on a visual analog scale (Myer et al., 2006), changed to a dichotomous scale (0 for no flaw or 1 for flaw present) (Myer et al., 2011) and most recently modified using an ordinal scale (Fort-Vanmeerhaeghe et al., 2017). A significant disparity in the reported interrater and intrarater reliability arose with the dichotomously scored TJA, between those involved in the development of the TJA (Herrington et al., 2013) and other researchers who were not involved (Dudley et al., 2013). Dudley, et al. (2013) reported the lack of a clarity in protocol and rater training in the dichotomous TJA description (Myer et al., 2011), and these limitations may have contributed to the poor to moderate reliability found in their study of varied raters with differing educational backgrounds. Possibly in reference to the issues brought up in Dudley, et al. (2013), Fort-Vanmeerhaeghe et al. (2017) suggested that a lack of background information and the specific training in the TJA led to reliability issues in the dichotomous TJA scoring, which they believed necessitated changing the TJA protocol. However, the authors did not provide a detailed explanation for the training of the raters, nor their involvement with the creation of the modified TJA, which would have provided important information as a significant learning effect with scoring was seen with the dichotomous TJA (Dudley et al., 2013) which may inflate the reliability in this study

(Fort-Vanmeerhaeghe et al., 2017). Further and perhaps more importantly, the clinical applicability of the new ordinal scoring methods is limited because it is not clear what is required to train raters for reliable scoring, especially with a new, more complicated scoring system. Beyond a simple explanation that the raters “watched as many times as necessary and at whatever speeds they needed to score each test,” no other methodology on video scoring was reported (Fort-Vanmeerhaeghe et al., 2017). Several questions are not answered in the study but will significantly impact replication of the findings and the use in a clinical setting. Were the raters instructed on calibrating volume? Were the raters instructed in the criteria for scoring? Did the raters work together to calibrate their scoring prior to the study? If so, for how long and by what methods?

To illustrate, for “pause between jumps,” the following criteria are reported: (0) reactive and reflex jumps, (1) small pause between jumps, and (2) large pause between jumps. The authors do not explain the difference between small and large. If the frame rate is not controlled while watching the video frame by frame, a rater may incorrectly score a severe pause between jumps when there is no flaw present. To limit this error, a possible solution is for the rater to watch the video at normal speed and only mark a flaw present if a pause is noticeable. The difference between a large and small pause could then be determined by determining time during the pause by going frame by frame. Pauses longer than half a second could constitute a large flaw (2), while those below are a small flaw (1). The method of scoring for each flaw needs to be clear and outline common errors in methodology, especially with a new scoring criteria.

The flaw “excessive contact noise at landing” seems to have two separate criteria in modified TJA compared with the dichotomously scored TJA. Fort-Vanmeerhaeghe et al. (2017) provided the following criteria: (0) subtle noise at landing (landing on the balls of their feet), (1) audible noise at landing (heels almost touch the ground at landing), (2) loud and pronounced noise at landing (contact of the entire foot and heel on the ground between jumps). The text in parentheses was not included in other research on the TJA (Myer et al., 2011). No explanation for this addition is present in the study, and the ambiguity of these criteria will limit reproducibility. If an athlete lands softly and the entire foot and heel touch the ground between jumps, this may be related to the pause between jumps flaw. Would this still be scored as excessive contact noise and scored as a severe flaw even when the noise is not excessive? From the study, it is unclear what constitutes excessive contact noise, if noise was considered in the scoring, if the raters calibrated volume to a certain level during video analysis, and if foot landing strategy should impact scoring—this clarity is needed for reliability, clinical utility, and validity.

In closing, our team has found the TJA to be clinically valuable in practice. We suggest more detail on training methodology for adequate reliability in raters with the modified TJA (Dudley et al., 2013), and an improved method for quantifying excessive contact noise.

Craig A. Smith^{1,2}✉, **Nicole J. Chimera**³, **Monica R. Lininger**² and **Meghan Warren**²

¹Smith Performance Center, Tucson AZ USA; ²Department of Physical Therapy and Athletic Training, Northern Arizona University, USA; ³Department of Athletic Training, Daemen College, USA

References

Dudley, L.A., Smith, C.A., Olson, B.K., Chimera, N.J., Schmitz, B. and Warren, M. (2013) Interrater and intrarater reliability of the tuck jump assessment by health professionals of varied educational backgrounds. *Journal of Sports Medicine* **2013**, 483503.

Fort-Vanmeerhaeghe, A., Montalvo, A.M., Lloyd, R.S., Read, P. and Myer, G.D. (2017) Intra- and inter-rater reliability of the modified tuck jump assessment. *Journal of Sports Science and Medicine* **16**, 117-124.

Herrington, L., Myer, G.D. and Munro, A. (2013) Intra and inter-tester reliability of the tuck jump assessment. *Physical Therapy in Sport* **14**, 152-155.

Myer, G.D., Paterno, M.V., Ford, K.R., Quatman, C.E. and Hewett, T.E. (2006) Rehabilitation after anterior cruciate ligament reconstruction: criteria-based progression through the return-to-sport phase. *Journal of Orthopaedic and Sports Physical Therapy* **36**, 385-402.

Myer, G.D., Brent, J.L., Ford, K.R. and Hewett, T.E. (2011) Real-time assessment and neuromuscular training feedback techniques to prevent ACL injury in female athletes. *Strength and Conditioning Journal* **33**, 21-35.

✉ **Craig Smith, PT, DPT**

Physical Therapist and Director, Smith Performance Center, Tucson AZ, Adjunct Faculty, Department of Physical Therapy, Northern Arizona University

E-mail: craig@smithperformancecenter.com

Authors' response

The authors would like to thank Dr. Smith and colleagues for their thoughtful comments regarding the most recent attempt to improve the clinical utility of the Tuck Jump Assessment Tool (TJA). Based on prior evidence and practitioner feedback, we aimed to improve the clarity of the assessment tool and add a further layer of objectivity in the scoring of each criteria to enhance its clinical utility. Here we have included our responses to Dr. Smith's questions and concerns to provide further clarification for the readers.

Dr. Smith first asked us to clarify how the raters were trained in the modified TJA to improve consistency. As mentioned in our original manuscript, raters were certified strength and conditioning coaches with over five years of clinical experience. In order to ensure that they could achieve maximum reliability, the two raters underwent training once per week for three months under the guidance of a third person who is an expert in scoring the TJA. The expert (AF) had previously trained with the creator of the original test (GDM). During these weekly trainings, the three raters individually scored the same 15 athletes on each criteria of the TJA. After the raters scored the athletes, they debriefed to discuss differences among scores. Each week, a new set of 15 athletes were scored. By the time the study commenced, each rater had scored and debriefed over 100 different athletes on the TJA. To maintain consistency with regard to the number of times videos were watched and the speed of movement during the scoring of each test, the raters followed a specific procedure. Most times, the raters watched the videos once in slow motion (speed reduced by 50%) and once at normal speed in both the frontal and sagittal planes for a total of four observations of the tuck jump. If this procedure did not provide the rater with sufficient information to score the item(s) with the required clarity, the raters were allowed to review the videos again to clarify criteria in question. This approach is likely replicable of how the

test would be scored in most clinical applications.

With reference to Dr. Smith's point concerning the measurement of the length of pauses between jumps, raters were informed that fast stretch shortening cycle actions indicative of 'true' plyometric tasks typically require ground contact times less than 250 milliseconds (Chu and Myer, 2013). Athletes who displayed minimal ground contact times observed at normal speed during video playback were considered to have met this criteria and were assigned a score of "0". Athletes who were perceived to demonstrate small pauses (likely in the 250–500 milliseconds) were assigned a score of "1". Finally, athletes who displayed noticeably longer pauses (e.g. heel contact at landing, a visual pause in movement or a double ankle bounce) were assigned a score of "2". While the raters scored this criterion subjectively as per the other items in the revised TJA, observations were based on extensive experience in screening this test and coaching plyometric activities as has been previously stated. Practitioners who wish to more objectively quantify the length of the pause between jumps should consider the use of jump mats that are capable of measuring ground contact times; however, this may reduce the practical nature of the test.

The final point raised by Dr. Smith pertains to potential flaws in the measurement of excessive contact noise during landing. We acknowledge that there are limitations in subjectively quantifying this criterion, but feel it is a critical component of the TJA as loud landings are likely indicative of excessive ground reaction forces and poor force dissipation strategies when landing. We demonstrated, is a measure in which acceptable intra-rater reliability can be achieved. In order to reduce the variability and subjectivity in this item, it was helpful during the training and testing periods for raters to consider both the actual noise during landing and the distance between the athlete's heels and the floor at the point of ground contact, which was an addition based on experience (Myer et al.,

2013; Stroube et al., 2013). While it is true that these could be considered two separate criteria, this additional information helped less experienced raters to become more consistent and reliable. Specifically, the raters followed a designated procedure during which they were instructed to firstly determine the noise during landing. If the raters were unable to accurately determine the appropriate score based on the noise alone, they were then instructed to evaluate the distance between the heels and the ground to assign a score. It should also be acknowledged that while we agree with Dr. Smith that it is possible for trained athletes to execute landings with minimal noise even when the heels touch the ground, in our experience, rebound jumps that display effective plyometric technique (i.e. no heel contact) are generally quieter.

In closing, we would like to thank Dr. Smith for his letter and hope that our response provides further clarification on how to correctly score the revised TJA, enhancing its practical application and clinical utility. It is hoped that the ordinal scale proposed in the revised TJA provides a further layer of objectivity in assessing clients and athletes, whereby, the degree in which technical flaws can be more effectively rated as opposed to the original dichotomous scale which may not accurately depict the range in the observed deficits shown. This has important connotations for risk stratifying athletes and allows for the development of more targeted training programs to reduce possible injury risk. Finally, we encourage and look forward to future evidence-based approaches to enhance the utility of the TJA from Dr. Smith and other colleagues.

Azahara Fort-Vanmeerhaeghe^{1,2}✉, **Alicia M. Montalvo**³, **Rhodri S. Lloyd**⁴, **Paul Read**⁵ and **Gregory D. Myer**^{6,7,8,9}

¹ School of Health and Sport Sciences (EUSES) Universitat de Girona, Salt, Spain; ² Blanquerna Faculty of Psychology, Education Sciences and Sport (FPCEE), Universitat Ramon Llull, Barcelona, Spain; ³ Florida International University, Nicole Wertheim College of Nursing and Health Sciences, Department of Athletic Training, Miami, FL; and Pennsylvania State University, Department of Kinesiology, Athletic Training/Sports Medicine Program, University Park, PA, USA; ⁴ Youth Physical Development Unit, Cardiff Metropolitan University, Cardiff, Wales; UK; ⁵ School of Sport, Health and Applied Science, St Mary's University, London, UK; ⁶ Division of Sports Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; ⁷ Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH; ⁸ Sports Health and Performance Institute, Ohio State University, Sports Medicine, Ohio State University Medical Center, Columbus, OH; ⁹ Micheli Center for Sports Injury Prevention, Waltham, MA, USA

References

- Chu, D.A. and Myer, G. (2013) *Plyometrics*. Human Kinetics.
- Myer, G.D., Stroube, B.W., DiCesare, C.A., Brent, J.L., Ford, K.R., Heidt, R.S.Jr. and Hewett, T.E. (2013) Augmented feedback supports skill transfer and reduces high-risk injury landing mechanics: a double-blind, randomized controlled laboratory study. *American Journal of Sports Medicine* **41**(3), 669-677.
- Stroube, B.W., Myer, G.D., Brent, J.L., Ford, K.R., Heidt, R.S., Jr. and Hewett, T.E. (2013) Effects of task-specific augmented feedback on deficit modification during performance of the tuck-jump exercise. *Journal of Sport Rehabilitation* **22**(1), 7-18.

✉ **Azahara Fort-Vanmeerhaeghe**

School of Health and Sport Sciences (EUSES) Universitat de Girona, Salt, Spain