

Original Research

Reliability of the Tuck Jump Assessment Using Standardized Rater Training

Kevin Racine, PT¹, Meghan Warren, PT, MPH, PhD¹, Craig Smith, PT, DPT², Monica R. Lininger, PhD, LAT, ATC^{1 a}¹ Northern Arizona University, ² Northern Arizona University; Smith Performance Center

Keywords: screening, psychometrics, movement system, injury risk

<https://doi.org/10.26603/001c.18662>

International Journal of Sports Physical Therapy

Vol. 16, Issue 1, 2021

BACKGROUND

The Tuck Jump Assessment (TJA) is a test used to assess technique flaws during a 10-second, high intensity, jumping bout. Although the TJA has broad clinical applicability, there is no standardized training to maximize the TJA measurement properties.

HYPOTHESIS/PURPOSE

To determine the reliability of the TJA using varied healthcare professionals following an online standardized training program. The authors hypothesized that the total score will have moderate to excellent levels of intra- and interrater reliability.

STUDY DESIGN

Cross-sectional reliability.

METHODS

A website was created by a physical therapist (PT) with videos, written descriptors of the 10 TJA technique flaws, and examples of what constituted no flaw, minor flaw, or major flaw (0,1,2) using published standards. The website was then validated (both face and content) by four experts. Three raters of different professions: a PT, an AT, and a Strength and Conditioning Coach Certified (SCCC) were selected due to their expertise with injury and movement. Raters used the online standardized training, scored 41 videos of participants' TJAs, then scored them again two weeks later. Reliability estimates were determined using intraclass correlation coefficients (ICCs) for total scores of 10 technique flaws and Krippendorff α ($K \alpha$) for the individual technique flaws (ordinal).

RESULTS

Eleven of 50 individual technique flaws were above the acceptable level ($K \alpha = 0.80$). The total score had moderate interrater reliability in both sessions (Session 1: $ICC_{2,2} = 0.64$; 95% CI (Confidence Interval) (0.34-0.81); Standard Error Measurement (SEM) = 0.66 technique flaws and Session 2: $ICC_{2,2} = 0.56$; 95% CI (0.04-0.79); SEM = 1.30). Rater 1 had a good reliability ($ICC_{2,2} = 0.76$; 95% CI (0.54-0.87); SEM = 0.26), rater 2 had a moderate reliability ($ICC_{2,2} = 0.62$; 95% CI (0.24-0.80); SEM = 0.41) and rater 3 had excellent reliability ($ICC_{2,2} = 0.98$; 95% CI (0.97-0.99); SEM = 0.01).

a Corresponding author:

Monica R. Lininger, PhD, LAT, ATC
Northern Arizona University
PO Box 15094
Flagstaff, AZ 86011
928.523.7442 (phone)
928.523.4315 (fax)
monica.lininger@nau.edu

CONCLUSION

All raters had at least good reliability estimates for the total score. The same level of consistency was not seen when evaluating each technique flaw. These findings suggest that the total score may not be as accurate when compared to individual technique flaws and should be used with caution.

LEVEL OF EVIDENCE:

3b

INTRODUCTION

Several clinical screening tests have been created to help identify individuals who are at high risk upon observation of jump-landing tasks. One of these tests is the 10-second Tuck Jump Assessment (TJA)^{1,2} which was developed as a “clinician-friendly” screening test to help identify lower extremity landing technique flaws in individuals during a high intensity, plyometric activity.^{1,2} One advantage of the TJA is that it is quick and inexpensive to administer, as it only requires athletic tape and two video cameras. The TJA may better simulate conditions faced in actual sporting activities than other anterior cruciate ligament (ACL) injury screening tools or jump-landing screening tests because it begins and ends from ground level while requiring maximum effort over multiple repetitive jumps.³ The TJA requires participants to jump with maximal effort over 10 seconds; this may also induce fatigue, which may expose jumping or landing technique flaws not seen with other tests that use one or two jumps.³ The TJA performance is scored qualitatively by a clinician based on 10 technique flaws using video recordings. The original TJA assessment has 10 technique flaws, which were scored on a dichotomous scale (0-1) as either flaw present (1) or absent (0).¹ Previous literature reporting the intra- and interrater reliability of the TJA has yielded mixed results, ranging from poor to excellent.³⁻⁵

Recently, because of the inconsistency in scoring interpretations, a modified scoring system of the TJA was developed.⁵ The modified system changed the scoring of the 10 technique flaws to an ordinal scale of 0-2 (0 = no flaw, 1 = small flaw, 2 = large flaw). Initial reliability testing of the modified TJA found both excellent interrater and intrarater reliability.⁵ However, significant limitations in the study by Fort-Vanmeehaeghe, et al.⁵ warranted caution when interpreting the results. Details of the type or the amount of training that raters received on the modified scoring of the TJA were not included.

Understanding the level of training and education of the raters could give important information on possible learning effect associated with scoring and would help with reproducibility.^{3,6} Additionally, the study used two raters who were both certified strength and conditioning coaches with five years of clinical experience; potentially limiting utility of the modified TJA by other professionals in the athletic performance community. Another reliability study of the modified TJA found excellent intrarater and good interrater reliability for total score, but lower reliability for individual technique flaws.⁷ However, this study also did not provide any details regarding the type or amount of training the raters received prior to scoring TJAs.

Currently, there is no standardized TJA training for

raters, nor any standards of how much rater experience with the TJA produces reliable results. The purpose of this work was to determine the reliability of the TJA using varied healthcare professionals following an online standardized training program. The authors hypothesized that the raters would have moderate to excellent levels of intra- and interrater reliability.

METHODS

This cross-sectional reliability study was a secondary analysis of TJA videos obtained as part of a larger study. A website, www.tuckjumpassessment.com, was created by a physical therapist with videos and written descriptors of TJA technique flaws as examples of what constitutes no flaw, minor flaw, or major technique flaws (0,1,2, respectively). The website was created a tool to be included in standardized rater training. The website was then validated by four experts in the field (two athletic trainers (AT) and two physical therapists (PT)), who have scored over 50 TJAs and use the TJA clinically; two of these experts have also been authors of past TJA studies.³ These experts added both face and construct validity to the website by assessing the website using a standardized instrument and providing feedback based on whether or not the videos were accurate representations of scoring the technique flaws. Modifications to the training website were made based on the experts' feedback.

To test intra- and interrater reliability between raters of varying educational and clinical background, our study design utilized three raters of different professions; a PT with a Doctorate of Physical Therapy degree and two years of clinical experience, an AT with five years of clinical experience and a Strength and Conditioning Coach Certified (SCCC) with five years of experience. The PT and AT for this portion of the study were different from the experts involved in the validation process. These three raters were chosen because they represented professions associated with injury screening and athletic performance. Each rater independently scored videos of 41 participants after reviewing the website and reading details of the modified TJA.⁸ Raters were asked to complete two scoring sessions two weeks apart to reduce the likelihood of remembering scores completed during the first session.

Instructions to participants for performance of the TJA were the same as established by Myer, et al.⁸ which consisted of standing in an athletic position with feet shoulder width apart then swinging their arms while jumping straight in the air and pulling the knees up as high as possible while landing as softly as possible and repeating until told to stop at the end of 10 seconds or stopping if they

Table 1. Scoring criteria for technique flaws of the Tuck Jump Assessment

Technique Flaw	Score of 0	Score of 1	Score of 2
Lower extremity valgus at landing	No valgus at landing	Slight valgus	Obvious valgus; both knees touch
Thighs do not reach parallel (peak of jump)	The knees are higher or at the same level as the hips	The middle of the knees are at a lower level than the middle of the hips	The whole knees are under the entire hips
Thighs not equal side to side	Thighs equal side to side	Thighs slightly unequal side to side	Thighs completely unequal side to side (one knee is over the other)
Foot placement not shoulder width apart	Foot placement exactly shoulder width apart	Foot placement mostly shoulder width apart	Both feet fully together and touch at landing
Foot placement not parallel	Foot (the end of the feet) placement parallel	Foot placement mostly parallel	Foot placement obviously unparallel (one foot is over half the distance of the other foot/leg)
Foot contact timing not equal	Foot contact timing equal side to side	Foot contact timing slightly unequal	Foot contact timing completely unequal
Does not land in same footprint	Lands in the same footprint	Does not land in same footprint, but inside the shape	Lands outside the shape
Excessive landing contact noise	Subtle noise at landing (landing on the balls of their feet)	Audible noise at landing (heels almost touch the ground at landing)	Loud and pronounced noise at landing (contact of the entire foot and heel on the group between jumps)
Pause between jumps	Reactive and reflex jumps	Small pause between jumps	Large pause between jumps (or double contact between jumps)
Technique declines prior to 10 seconds	No decline in technique	Technique declines after five seconds	Technique declines before five seconds

could not complete the full 10 seconds. The 10 seconds of jumping were video recorded and each participant was scored on 10 established technique flaws.⁵ If the flaw was seen two or more times during the 10 second period, then it was counted and scored by a magnitude of 1 (small) or 2 (large).⁵ The complete scoring rubric can be seen in [Table 1](#).

The videos of 41 study participants performing the TJA were part of a previous study.

The participants were between 18-30 year-old and recreationally active (participated in physical activity for at least 30 minutes three times a week for the prior five to six months and not participating in formal athletics competition). Those with a concussion within the prior six months and women who were pregnant were excluded. Each participant filled out a Physical Activity Readiness Questionnaire (PAR-Q)⁹ and positive answers were evaluated by a licensed AT or PT to ensure safe participation with data collection.

The Institutional Review Board at Northern Arizona University approved this study, informed consent was obtained prior to the collection of data, and patient confidentiality was protected according to the U.S. Health Insurance Portability and Accountability Act.

STATISTICAL METHODS

The total score for the TJA was treated as continuous data and therefore traditional intraclass correlation coefficients

(ICCs) from repeated measures analysis of variance, specifically a two-way random model with absolute agreement (ICC_{2,2}) was used. ICC values, measures of relative reliability, were classified as excellent (> 0.90), good (0.75-0.89), moderate (0.50-0.74), or poor (<0.50).¹⁰ Standard error of measure (SEM), a measure of absolute reliability, was calculated using $SEM = SD\sqrt{(1 - ICC)}$. The individual technique flaws were ordinal (0,1,2) and therefore reliability was assessed using the Krippendorff α ($K \alpha$), which allows for ordinal data with multiple raters. Values > 0.80 were considered acceptable.^{11,12} Ninety-five percent confidence intervals were constructed using a bootstrapping technique (n=1,000). These procedures were followed for both intrarater (individual rater across the two time points for either the individual technique flaw or the total score) and interrater (across 3 raters for either the individual technique flaw or the total score) reliability.

Level of agreement estimates were calculated using Fleiss's kappa due to multiple raters and use of ordinal data. Fleiss's kappa¹³ were classified as almost perfect agreement (≥ 0.81), substantial agreement (0.61-0.80), moderate agreement (0.41-0.60), fair agreement (0.21-0.40), and slight agreement (0.01-0.20). All analyses were conducted in SPSS version 25 (IBM SPSS, Inc.).

Table 2. Krippendorff alpha coefficients (K α (95% Confidence Interval)) for intrarater reliability estimates on individual technique flaws.

	Rater 1	Rater 2	Rater 3
Technique Flaws			
Lower extremity valgus at landing	0.65 (0.39, 0.85)	0.60 (0.35, 0.83)	0.78 (0.56, 0.95)
Thighs do not reach parallel	0.57 (0.30, 0.82)	0.66 (0.51, 0.82)	0.99 (0.99, 0.99)
Thighs not equal side-to-side	0.31 (0.05, 0.56)	0.48 (0.01, 0.99)	0.99 (0.99, 0.99)
Foot placement not shoulder width apart	0.41 (0.05, 0.74)	0.53 (0.24, 0.83)	0.96 (0.89, 0.99)
Foot placement not parallel	0.33 (0.01, 0.73)	0.18 (0.01, 0.66)	0.99 (0.99, 0.99)
Foot contact timing not equal	0.44 (0.12, 0.72)	0.27 (0.01, 0.85)	0.99 (0.99, 0.99)
Excessive landing contact noise	0.68 (0.47, 0.85)	0.41 (0.01, 0.99)	0.94 (0.85, 0.99)
Pause between jumps	0.80 (0.68, 0.90)	0.86 (0.69, 0.97)	0.98 (0.94, 0.99)
Technique declines prior to 10 seconds	0.14 (0.01, 0.49)	0.02 (0.01, 0.39)	0.88 (0.76, 0.97)
Does not land in same footprint	0.25 (0.01, 0.66)	0.36 (0.12, 0.85)	0.95 (0.84, 0.99)

Table 3. Krippendorff alpha coefficients (K α (95% Confidence Interval)) for interrater reliability estimates on individual technique flaws for each viewing session.

	Session 1	Session 2
Technique Flaws		
Lower extremity valgus at landing	0.64 (0.51, 0.76)	0.50 (0.32, 0.67)
Thighs do not reach parallel	0.54 (0.40, 0.67)	0.42 (0.30, 0.53)
Thighs not equal side-to-side	0.11 (0.01, 0.33)	0.24 (0.06, 0.41)
Foot placement not shoulder width apart	0.32 (0.15, 0.48)	0.31 (0.14, 0.47)
Foot placement not parallel	0.26 (0.06, 0.46)	0.15 (0.01, 0.41)
Foot contact timing not equal	0.12 (0.01, 0.27)	0.10 (0.01, 0.24)
Excessive landing contact noise	0.13 (0.01, 0.31)	0.03 (0.01, 0.22)
Pause between jumps	0.63 (0.50, 0.74)	0.62 (0.51, 0.72)
Technique declines prior to 10 seconds	0.06 (0.01, 0.23)	0.00 (0.00, 0.03)
Does not land in same footprint	0.15 (0.01, 0.32)	0.00 (0.00, 0.13)

RESULTS

Rater 1 had a good reliability ($ICC_{2,2} = 0.76$; 95% CI (0.54 - 0.87); SEM = 0.26), rater 2 had a moderate reliability ($ICC_{2,2} = 0.62$; 95% CI (0.24 - 0.80); SEM = 0.41) and rater 3 had excellent reliability ($ICC_{2,2} = 0.98$; 95% CI (0.97 - 0.99); SEM = 0.01) for the total score. The raters had moderate levels of interrater reliability for the total score in both sessions (Session 1: $ICC_{2,2} = 0.64$; 95% CI (0.34 - 0.81); SEM = 0.66 and Session 2: $ICC_{2,2} = 0.56$; 95% CI (0.04 - 0.79); SEM = 1.30). Of all individual technique flaw reliability estimates (K α) for both intra- and interrater reliability, only 11 (50 total) were above the acceptable level (Tables 2 and 3). For level of agreement (Fleiss's Kappa) (Table 4), within session 1, 3 individual technique flaws (lower extremity valgus at landing, thighs do not reach parallel, and technique declines prior to 10 seconds) had moderate agreement. Thighs not equal side-to-side had a fair level of agreement between raters in session 1. In Session 2, thighs do not reach parallel

and technique declines prior to 10 seconds had moderate and fair agreement, respectively, among raters. All other assessments of agreement were within the slight agreement classification (0.01-0.20).

DISCUSSION

The primary objective of this study was to investigate the intra- and interrater reliability of the modified TJA when using a standardized training tool for raters of different clinical backgrounds who may be likely to use the modified TJA clinically. The main findings were the total scores had good, moderate, and excellent intrarater reliability, respectively, among the three different raters. Additionally, the total TJA scores had moderate interrater reliability between both scoring sessions.

When examining the intra- and interrater reliability for individual technique flaws, only 11/50 of the K α coefficients were above the acceptable level of 0.80. The level of

Table 4. Fleiss's Kappa for agreement of scores (Fleiss's Kappa (95% Confidence Interval)) on individual technique flaws for each viewing session.

	Session 1	Session 2
Technique Flaws		
Lower extremity valgus at landing	0.56 (0.41, 0.70)	0.07 (-0.09, 0.24)
Thighs do not reach parallel	0.56 (0.41, 0.70)	0.43 (0.28, 0.58)
Thighs not equal side-to-side	0.33 (0.20, 0.46)	0.12 (-0.01, 0.25)
Foot placement not shoulder width apart	0.09 (-0.06, 0.24)	0.10 (-0.05, 0.25)
Foot placement not parallel	0.19 (0.04, 0.34)	0.19 (0.03, 0.34)
Foot contact timing not equal	0.15 (0.01, 0.29)	0.10 (-0.04, 0.24)
Excessive landing contact noise	0.01 (-0.16, 0.14)	0.01 (-0.21, 0.82)
Pause between jumps	0.02 (-0.11, 0.15)	0.01 (-0.20, 0.07)
Technique declines prior to 10 seconds	0.43 (0.30, 0.56)	0.35 (0.22, 0.48)
Does not land in same footprint	0.01 (-0.18, 0.11)	0.01 (-0.26, 0.01)

agreement between raters for both scoring sessions using Fleiss' kappa for the majority of individual items showed slight agreement (0.01-0.20). Fort-Vanmeerhaeghe, et al.⁵ found intra- and interrater reliability coefficients for the modified TJA scoring of individual technique flaws as good to excellent with 27/30 of the Fleiss's kappa coefficients above 0.61, which was used as the cutoff for good defined in their statistical analysis. However, the findings of the current study are more aligned with Gokeler and Dingenen⁷, which demonstrated poor level of agreement for the majority of individual technique flaws for both intra- and interrater reliability, using item analysis. One potential explanation for the differences in results for individual flaw reliability could be the lack of clarity of the scoring descriptors. Two of the individual items that showed the lowest level of agreement between raters in both sessions were "excessive landing contact noise" and "pause between jumps." The current scoring protocol for "excessive landing contact noise" is as follows: (0) subtle noise at landing (landing on the balls of their feet), (1) audible noise at landing (heels almost touch the ground at landing), and (2) loud and pronounced noise at landing (contact of the entire foot and heel on the ground between jumps).⁵ Concerns regarding the scoring of this specific technique flaw have been expressed by Smith, et al.⁶ In the work by Smith, et al., the team described the need to find means for quantifying landing noise via standardized volume calibration or by rephrasing the written descriptors regarding types of foot contact for each score.⁶ For example, it is possible that a person could land loudly, but on the balls of their feet, which would provide conflicting scoring options for raters. For the flaw "pause between jumps," a similar issue regarding the written descriptors exists. The current scoring protocol for this flaw is: (0) reactive and reflex jumps, (1) small pause between jumps, and (2) large pause between jumps.⁵ With these descriptors, there is no way to delineate or quantify what constitutes a "small pause" versus a "large pause." Smith, et al.⁶ proposed a change in the scoring protocol for this flaw to standardize a time-based cutoff for a small versus a large pause, for example 0.5 seconds, which could then be determined by watching the video frame by frame.

Another potential factor in the lack of consensus of individual flaw reliability could be due to inconsistency regarding the training of the TJA raters. One reason the original scoring system was revised into the modified version was that it was believed that the original version did not allow the rater to evaluate severity of dysfunction in the outlined criteria due to the dichotomous scoring nature (0,1).⁵ This dichotomous scoring system also made it difficult to determine improvement or reduction in an individual's lower extremity landing techniques. The new modified scoring system was proposed to provide a more objective assessment on an individual's risk of ACL injury.⁵ A reliability study of the original TJA conducted by Dudley, et al.³ included three raters who scored participants on two separate occasions and were analyzed for inter-rater reliability, which demonstrated a learning effect as ICC values increased from 0.52 to 0.69 between the first and second sessions.³ This demonstrated the potential need for scoring practice to be included as a standard in training assessors as a way to potentially improve reproducibility and reliability of the test. The need for training of raters is even more imperative for the current modified version of the TJA due to more scoring options on an ordinal scale. This ordinal scale results in slightly higher degrees of subjectivity, which is why standardized rater training was included in this study. To this research team's knowledge, this is the first study to clearly outline and delineate the training procedures that TJA rates completed prior to scoring the TJA videos. The lack of consensus between the current study and previous studies of excellent reliability points to the need for further research to make a definitive conclusion about several important psychometric property of this test. In addition to a description of the TJA and the participate completing the test, future studies must include information about the experience of the raters with the TJA and other observational movement assessment tests, as well as the training specific to the TJA the raters received.

Initial reliability testing of the modified TJA scoring conducted by Fort-Vanmeerhaeghe, et al.⁵ in a group of 24 athletes with two raters found excellent intrarater (ICC = 0.94; 95% CI = (0.88-0.97); ICC rater 2 = 0.96; 95% CI =

(0.92-0.98)), as well as interrater (ICC = 0.94; 95% CI = (0.88-0.97)) reliability for total score. In a more recent study conducted by Gokeler and Dingenen,⁷ excellent intrarater reliability for the total score (ICC rater 1 = 0.93; 95% CI = (0.78-0.98); ICC rater 2 = 0.96; 95% CI = (0.89-0.99)), and good interrater reliability for total score (ICC rater 1 = 0.85; 95% CI = (0.58-0.95); ICC rater 2 = 0.88; 95% CI = (0.66-0.96)) was reported. A lack of consensus in previous studies and the current study may be explained by differences in statistical methods. There are six different Shrout and Fleiss models for ICCs that are commonly used. So, if two separate researchers use two different models then the findings for reliability estimates might be slightly different. Gokeler and Dingenen⁷ used the same model (ICC_{2,2}) as the current study, but Fort-Vanmeerhaeghe, et al.⁵ did not provide which model was used. One suggestion for future TJA research would be to include which ICC model was used to calculate reliability estimates for replication and direct comparison of results.

Due to the lower individual item reliability and potential variability in scoring combinations, it is possible that the reported higher reliability results could be falsely inflated. For example, a total score of 8/20 could be highly variable and achieved in a variety of ways due to a combination of different scores for individual technique flaws. These findings align with those discovered in a recent critically appraised topic (CAT) of five TJA reliability studies which concluded that total scores are reliable, but individual technique flaws when examined individually are not reliable.¹⁴ Therefore, as previously stated by Gokeler and Dingenen⁷ it is advised to use caution when solely looking at total scores when interpreting injury risk.

This is the first study to provide a standardized protocol for training of raters prior to scoring any of the TJA videos. The website created for training purposes was validated by selected experts of the TJA, providing both face and con-

struct validity to the website as a training platform for future scoring. One potential limitation of this study is that there was a fairly homogenous group of participants (college-aged, recreationally active) and the TJA was originally developed for use with athletes and not a recreationally active population.

CONCLUSION

Using standardized rater training, the modified TJA revealed moderate interrater reliability and moderate to excellent intrarater reliability for total scores, but only slight levels of agreement for the majority of individual technique flaws for both intra- and interrater reliability. These findings demonstrate that caution is warranted when solely interpreting total scores and also indicates the need for certain technique flaws such as “pause between jumps” and “excessive landing contact noise” to be further examined in terms of scoring descriptions and potentially modified to be more reproducible.

Dr. Warren is now working at the Patient-Centered Outcomes Research Institute (PCORI). All statements, findings, and conclusions in this publication are solely those of the authors and do not necessarily represent the views of the PCORI or its Board of Governors.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

Submitted: January 17, 2020 CDT, Accepted: May 23, 2020 CDT



REFERENCES

1. Myer GD, Brent JL, Ford KR, Hewett TE. Real-time assessment and neuromuscular training feedback techniques to prevent ACL injury in female athletes. *Strength Cond J*. 2011;33(3):21-35. doi:10.1519/SSC.0b013e318213afa8
2. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. *Athletic Therapy Today*. 2008;13(5):39-44.
3. Dudley LA, Smith CA, Olson BK, Chimera NJ, Schmitz B, Warren M. Interrater and intrarater reliability of the tuck jump assessment by health professionals of varied educational backgrounds. *Journal of Sports Medicine*. 2013;2013:1-5.
4. Herrington L, Myer GD, Munro A. Intra and inter-tester reliability of the tuck jump assessment. *Physical Therapy in Sport*. 2013;14(3):152-155. doi:10.1016/j.ptsp.2012.05.005
5. Fort-Vanmeerhaeghe A, Montalvo AM, Lloyd RS, Read P, Myer GD. Intra- and inter-rater reliability of the modified Tuck Jump Assessment. *Journal of Sports Science & Medicine*. 2017;16(1):117-124.
6. Smith CA, Chimera NJ, Lininger MR, Warren M. Modified Tuck Jump Assessment: Reliability and training of raters. *Journal Of Sports Science & Medicine*. 2017;16(3):440-442.
7. Gokeler A, Dingenen B. Between-session and inter-rater reliability of the modified tuck jump assessment in healthy adult athletes. *Physical Therapy in Sport*. 2019;37:10-14. doi:10.1016/j.ptsp.2019.02.002
8. Myer GD, Chu DA, Brent JL, Hewett TE. Trunk and hip control neuromuscular training for the prevention of knee joint injury. *Clinics in Sports Medicine*. 2008;27(3):425-448. doi:10.1016/j.csm.2008.02.006
9. PAR-Q. [https://www.nasm.org/docs/default-source/PDF/nasm_par-q-\(pdf-21k\).pdf](https://www.nasm.org/docs/default-source/PDF/nasm_par-q-(pdf-21k).pdf). Accessed December 20, 2019.
10. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Saddle River, NJ: Pearson Prentice Hall; 2009.
11. Krippendorff K. Reliability in content analysis some common misconceptions and recommendations. *Human Communication Research*. 2004;30(3):411-433.
12. Krippendorff K. *Content Analysis: An Introduction to Its Methodology*. Third edition. Los Angeles ; London: SAGE Publications, Inc; 2012.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174. doi:10.2307/2529310
14. Mason ML, Clemons MN, LaBarre KB, Szymczak NR, Chimera NJ. A critically appraised topic on the Tuck Jump Assessment: Does the Tuck Jump Assessment demonstrate interrater and intrarater reliability in healthy individuals? *J Sport Rehabil*. October 2019:1-6. doi:10.1123/jsr.2019-0099