Scientific-Expert Journal of Anthropological Aspects of Sports, Physical Education and Recreation

Submitted: 10/05/2024
Accepted: 26/06/2024
Published: 02/08/2024
UDK: to be registered

Original article

Graduate Versus Undergraduate Interrater Reliability of the Landing Error Scoring System (LESS) and Less-Real Time

[1,2]Megan B. Thompson, [3]Pasquale J. Succi, [2]Taylor Dinyer-Mcneely, [4]Robert G. Lockie, [5]Jennifer L. Volberding, [1,2]J. Jay Dawes

[1]Tactical Fitness and Nutrition Laboratory, Oklahoma State University, Stillwater, OK, USA
[2]School of Kinesiology, Applied Health and Recreation, Oklahoma State University, Stillwater, OK, USA;
[3]University of Kentucky, Lexington, KY, USA; [4]California State University, Fullerton, Fullerton, CA, USA. Oklahoma State University Health Sciences Center, Tulsa, OK
*Correspondence: megan.mcguire@okstate.edu

Abstract

The study sought to determine if the LESS and LESS-RT scoring criteria are reliable when scored by graduate and undergraduate kinesiology students with minimal experience. Eleven graduate (7 male: 28.29 ± 3.251 years; 4 female: 28.00 ± 4.082 years) and 19 undergraduate (7 male: 21.57 ± 1.512 years; 12 female: 21.42 ± 4.870 years) students participated as raters. Raters with minimal (< 2 hours) or no experience with the LESS watched 30 videos and evaluated jump-landing mechanics using the LESS and LESS-RT across four sessions (two per criteria). A 4-way repeated measures ANOVA analyzed interactions among trials, groups, videos, and scoring sheets. Intraclass correlation coefficient (ICC), coefficient of variation (CV), and minimal difference (MD) values were calculated, with an alpha level of 0.05. ICC values for the whole group and undergraduates (R = 0.102 – 0.780) demonstrated "poor" to "good" reliability, while graduate students (R = 0.356 – 0.814) demonstrated "poor" to "excellent" reliability. The CV for the whole group, graduate, and undergraduate students (14.24 – 29.90%), were all above the 10% threshold thus, reliable. Prior experience with the LESS may impact the quality of assessment, therefore, providing a single training session could drastically improve the quality of ratings, even for novice raters.

Keywords: Fitness Testing, Jump-landing, Injury Prevention, Biomechanics, Assessment

Citation: Thompson et al. (2024). Graduate Versus Undergraduate Interrater Reliability of the Landing Error Scoring System (LESS) and Less-Real Time. *SportLogia. 20*(2), xx-xx. doi: registering

Introduction

The Landing Error Scoring System (LESS) is a movement competency screen that has been utilized in a variety of athletic and tactical populations to determine the potential for injury (Beutler et al., 2009; Padua et al., 2011; Padua et al., 2009). Specifically, the LESS is used to create a composite score based on different outcome measures that identify errors in jump-landing mechanics (Padua et al., 2011; Padua et al., 2009). In recreationally athletic populations, poor jump landing mechanics have been shown to correlate and predict an increased risk of noncontact anterior cruciate ligament injury (via receiver operator characteristic curve; 86% sensitivity, 64% specificity, $p \leq 0.005$) (Kernozek et al., 2008; Padua et al., 2015). Specific to tactical populations, the LESS has been used to determine deficiencies in landing mechanics among soldiers, as many soldiering tasks (i.e., exiting vehicles, traversing uneven terrain, jumping over walls and other obstacles) require proper jumping and landing technique to mitigate injury risk (Beutler et al., 2009; Mala et al., 2015; Orr et al., 2015; Scott et al., 2015). For these reasons, the LESS may be of value among first responders for reducing injury risks, as proper landing technique could also be important in many tasks (entering and exiting fire engines, maneuvering over curbs and barriers, etc.) required in this population. Despite the potential practical importance of using the LESS in tactical personnel such as first responders, greater analysis of how the LESS is scored by different individuals who may be assisting with occupational fitness testing is warranted.

The LESS composite score is derived from objective viewing of jump-landing mechanics. Therefore, consideration of who is scoring the LESS is important, as from a practical standpoint, the LESS may provide greater insight into musculoskeletal weaknesses and technical flaws that could be addressed via specific strength training interventions. Lower composite scores indicate fewer movement compensations, whereas higher scores reflect more errors and a greater probability of injury (Padua et al., 2015; Everard et al., 2018). Padua et al. (2009) introduced the LESS scoring criteria to include 17 items which each indicate a particular biomechanical compensation within the drop jump and landing patterns (e.g., knee valgus angle and trunk flexion angle at initial contact, hip and trunk flexion at the point of maximal knee flexion, symmetrical foot contact). Following the introduction of the LESS, an abbreviated scoring system consisting of 10 jump-landing characteristics, the LESS-Real Time (LESS-RT), was derived from the full assessment (Hanzlíková & Hébert-Losier, 2020). These assessments were introduced as an easily implementable tool for large-scale screening sessions that would provide the rater with a reliable measure of jump-landing mechanics (Padua et al., 2011; Padua et al., 2009).

Numerous studies have evaluated the reliability of the LESS and LESS-RT, showing good to excellent intrarater and interrater reliability (ICC range 0.71 to 0.96) (Padua et al., 2011; Padua et al., 2009; Everard et al., 2019; Markbreiter et al., 2015). These studies have often utilized individuals with advanced degrees (masters of science, athletic trainer certified, etc.), and with prior experience utilizing the LESS to complete these evaluations (Everard et al., 2019). However, students and interns from various backgrounds often collect data alongside coaches and academic staff, and this data may be utilized for practical application and research. If individuals without advanced degrees or prior experience with the LESS are administering and scoring the LESS and LESS-RT, this could impact the quality of data collected.

There is a paucity of research utilizing individuals without LESS exposure and advanced degrees to evaluate the LESS. Therefore, the purpose of this study was to determine if the scoring criteria of the LESS and LESS-RT are reliable when scored by graduate and undergraduate students in kinesiology degree paths, with minimal or no experience with the LESS. It was hypothesized that graduate students would demonstrate greater intrarater and interrater reliability than undergraduate students. It was also hypothesized that the LESS full sheet (LESS-full) would show greater consistency between groups than the LESS-RT.

Methods

Participants

Eleven graduate (7 male: 28.29 ± 3.2 years; 4 female: 28.00 ± 4.1 years) (3.64 ± 1.6 years of post-baccalaureate schooling) and 19 undergraduate (7 male: 21.57 ± 1.512 years; 12 female: 21.42 ± 4.9 years) students voluntarily agreed to participate as raters in this study. An a priori power analysis was conducted to determine that a sample size of 28 participants was necessary to achieve a power of 0.80 at an alpha level of 0.05 (Faul et al., 2007). All subjects were recruited from exercise science courses at the university where this study was conducted. Prior to the commencement of this study, all protocols were approved by the university's institutional ethics committee (ED-19-139-STW), and all participants completed and signed an informed consent document. This research was carried out fully in accordance with recommended ethical standards within the field of exercise science and the Declaration of Helsinki (Navalta et al., 2019; World Medical Association, 2001).

Measurements and Procedures

Lower extremity injury risk was assessed using the LESS, with procedures adapted from previous research (Beutler et al., 2009; Padua et al., 2011). Individuals from a local law enforcement and fire agency performed a double-leg jump from a 30 cm platform, landing with both feet, at a distance of approximately half their body height, and then immediately completed a maximal effort vertical jump (Beutler et al., 2009). Three individual trials were performed, which were recorded with a mounted camera (Sony CX405 Handycam, Sony Electronics Inc.; San Diego, California, USA; HD/60p frame rate) placed approximately one meter away from the landing position. The camera height was adjusted to the height of the participant's hip, and distance from the participant was increased or decreased to ensure that the participant's face was not captured during the trials (based on the height of the participant), while still maintaining full view of the anatomical points of interest in the LESS sheet. The first two jump trials were recorded from the frontal view, and the final trial was recorded from the sagittal view (Padua et al., 2009). Once trials were recorded, the videos were extracted from the camera via USB, and the unedited videos were placed into a PowerPoint presentation for rater scoring.

Graduate and undergraduate raters with minimal (< 2 hours) or no experience with the LESS watched 30 total video clips (three videos for each view of the individual's jump trial of the LESS) and evaluated jump-landing mechanics using the scoring criteria provided. Specific items and scoring criteria have been detailed previously in the literature (Padua et al., 2009; Padua et al., 2015). No limit was placed on the rater's ability to rewind, pause, slow down, or review videos as they were scoring. Video order and LESS scoring sheet version (LESS-full or LESS-RT) were randomized for all raters using an online number generator (www.random.org). Raters completed two scoring sessions per sheet (a total of four scoring sessions), separated by one week between sessions. The raters watched the 10 participants jump trials during each session, totaling in 20 ratings with the LESS-full, and 20 ratings with the LESS-RT. On average, raters completed each session in approximately 45 minutes.

All raters were instructed to follow the criteria on the LESS scoring sheet. No additional information was provided. The LESS-full was scored using 17-items that provided points for every landing error identified. The items included frontal- and sagittal-plane analysis, and identified errors associated with knee and hip angles, foot contact, displacement at the knee and hip, and overall impressions. The LESS-RT was scored using 10 jump-landing characteristics, derived from the original LESS criteria (Padua et al., 2011; Padua et al., 2009). The LESS-RT removes 7 items related to knee valgus at initial contact, knee and hip flexion angle at initial contact, hip and truck flexion at maximal knee flexion.

Statistical analyses

All statistical analyses were conducted using IBM Statistics Package for Social Sciences (IBM SPSS Inc., version 26; Chicago, Illinois, USA). A 4-way (trial [trial 1 vs trial 2] x scoring sheet [LESS-full vs LESS-RT] x video

[video 1-video 10] x student [undergraduate vs graduate]) repeated measures analysis of variance (ANOVA) was used to determine interactions among the variables, and to identify if main effects were present. Additionally, a pairwise t-test was conducted to determine if undergraduate and graduate raters scored significantly different on any of the individual videos. The test-retest reliability of each of the scoring sheets was calculated using an intraclass correlation coefficient (ICC, relative reliability) (2, k) model (Hopkins et al., 2009; Weir, 2005; Weir & Vincent, 2020). The ICC values were classified as "excellent" (0.80-1.0), "good" (0.60-0.80), or "poor" (<0.60) (Buckthorpe et al., 2012). The coefficient of variation (CV) was calculated using previously described equations to indicate a normalized measure of the standard error of the measurement, and the minimal difference to be considered real (MD) was calculated to examine individual differences for raters from trial 1 to trial 2 of each sheet (Weir, 2005; Weir & Vincent, 2020). A CV of <10% was used as an indication of acceptable absolute reliability (Weir, 2005); however, the overall reliability was characterized by accounting for the ICC value, CV, and MD. An a priori alpha level was set at 0.05 for all analyses.

Results

The trial x sheet x video x student ANOVA lacked significance ($F = 0.453$, $p = 0.905$, $p\eta^2 = 0.016$). However, a significant 3-way ANOVA for trial x video x student ($F = 2.178$, $p = 0.024$, $p\eta^2 = 0.082$) and a significant video x student interaction ($F = 4.004$, $p < 0.001$, $p\eta^2 = 0.125$) was observed. Follow-up pairwise comparisons among the 10 videos indicated that undergraduate students scored one of the 10 videos significantly lower, and one video significantly higher than graduate students ($t = -3.478$, $p = 0.002$; $t = 2.211$, $p = 0.036$, respectively).

The reliability statistics for each video for the LESS and LESS-RT scoring sheets are presented in Table 1 and 2, respectively. Using the LESS-full scoring sheet, the ICC values for the whole group ($R = 0.218 – 0.780$) and the undergraduate students ($R = 0.102 – 0.774$) demonstrated "poor" to "good" reliability, while the graduate students ($R = 0.356 – 0.814$) demonstrated "poor" to "excellent" reliability. However, the CV for the whole group (16.05 – 29.69%), graduate (14.24 – 29.90%), and undergraduate (15.87 – 29.48%) students were all above the 10% threshold to be considered reliable (Atkinson & Nevill, 1998).

Table 1. Intraclass correlation coefficient (ICC), standard error of the measurement (SEM), minimal difference to be considered real (MD), and coefficient of variation (CoV) for the Whole group, Graduate, and Undergraduate Students.

	Whole				Graduates				Undergraduates			
Video	**ICC**	**SEM**	**MD**	**CoV**	**ICC**	**SEM**	**MD**	**CoV**	**ICC**	**SEM**	**MD**	**CoV**
1	0.602	1.44	3.99	23.26	0.678	1.43	3.97	22.30	0.562	1.43	3.97	22.28
2	0.482	1.51	4.20	16.05	0.474	1.57	4.34	18.14	0.302	1.37	3.80	15.87
3	0.654	1.48	4.09	25.23	0.739	1.25	3.47	21.90	0.628	1.59	4.40	27.77
4	0.621	1.39	3.86	22.03	0.502	1.69	4.69	28.58	0.743	1.09	3.01	18.34
5	0.755	1.13	3.14	28.56	0.750	0.67	1.85	14.79	0.735	1.33	3.68	29.48
6	0.470	1.26	3.49	16.15	0.623	1.10	3.04	14.29	0.398	1.35	3.73	17.51
7	0.780	1.23	3.40	16.98	0.814	1.03	2.85	14.24	0.774	1.34	3.70	18.53
8	0.469	1.29	3.56	22.17	0.500	1.20	3.32	20.59	0.472	1.34	3.73	23.11
9	0.218	1.86	5.16	29.69	0.356	1.98	5.50	29.90	0.102	1.74	4.81	26.18
10	0.703	1.53	4.25	18.78	0.482	1.72	4.77	21.87	0.797	1.38	3.81	17.49
Average	0.575	1.41	3.91	21.89	0.592	1.36	3.78	20.66	0.551	1.39	3.86	21.66

Table 2. Intraclass correlation coefficient (ICC), standard error of the measurement (SEM), minimal difference to be considered real (MD), and coefficient of variation (CoV) for the Whole group, Graduate, and Undergraduate Students.

	Whole Group				Graduates				Undergraduates			
Video	**ICC**	**SEM**	**MD**	**CoV**	**ICC**	**SEM**	**MD**	**CoV**	**ICC**	**SEM**	**MD**	**CoV**
1	0.355	1.17	3.24	22.83	0.108	0.85	2.37	18.07	0.392	1.31	3.62	24.45
2	0.471	1.16	3.22	16.50	0.439	0.88	2.43	11.03	0.380	1.26	3.50	19.34
3	0.384	1.07	2.97	22.44	0.349	1.02	2.83	21.01	0.410	1.11	3.08	23.58
4	0.556	1.16	3.22	24.59	0.339	1.14	3.16	21.63	0.653	1.10	3.05	25.03
5	0.281	1.29	3.58	40.17	0.380	0.70	1.94	27.03	0.218	1.51	4.19	42.25
6	0.501	1.08	3.00	16.54	0.123	0.96	2.66	14.76	0.622	1.10	3.05	16.72
7	0.287	1.28	3.55	22.59	0.547	1.05	2.90	18.13	0.177	1.39	3.85	24.81
8	0.542	1.22	3.37	25.26	0.746	0.78	2.15	15.39	0.487	1.40	3.89	29.96
9	0.501	1.13	3.14	21.48	0.400	0.83	2.29	18.76	0.451	1.24	3.43	21.47
10	0.469	1.30	3.59	18.84	0.532	1.24	3.44	16.35	0.387	1.31	3.63	20.20
Average	0.435	1.19	3.29	23.12	0.396	0.94	2.62	18.22	0.418	1.27	3.53	24.78

For the LESS-RT scoring sheet, the whole group (R = 0.287 – 0.556) demonstrated "poor" reliability (Buckthorpe et al., 2012). However, the graduate (R = 0.108 – 0.746) and undergraduate (R = 0.177 – 0.622) students demonstrated "poor" to "good" reliability (Buckthorpe et al., 2012). Similar to the LESS-full score sheet, the CV for the whole group (16.50 – 40.17%), graduate (11.03 – 27.03%), and undergraduate (16.72 – 42.25%) groups were above the 10% threshold to be considered reliable. Lastly, 10.5% of undergraduate subjects exceeded the MD test-retest for the LESS-full scoring sheet (MD = 3.86 ± 0.47; range: 3.01 – 4.81), and 7.9% exceeded the MD for the LESS-RT scoring sheet (MD = 3.53 ± 0.39; range: 3.05 – 4.19) across all ten videos (range for individual videos: 0-21% and 0-16%, respectively). Additionally, 10% of graduate subjects exceeded the MD for test-retest for the LESS-full sheet (MD = 3.78 ± 1.08; range: 1.85 – 5.50), and 18% exceeded the MD for the LESS-RT sheet (MD = 2.62 ± 0.47; range: 1.94 – 3.44) across all ten videos (range for individual videos: 0-36%).

Discussion

These data show that both undergraduate and graduate students were generally unreliable with scoring on the LESS and LESS-RT, with few individual exceptions. This could have been due to several factors, however minimal experience with this scoring criteria most likely contributed. A majority of the raters in this study did not exceed the MD for each video (82-91% of raters), suggesting more reliable scoring, however individual exceptions were observed. Ideally, with the use of a scoring system such as the LESS, it is desired that raters do not exceed the MD, as the MD indicates differences, and consistency is preferred. Overall, undergraduate students were slightly less reliable than graduate students that participated in this study, with both groups exceeding the 10% CV threshold.

The scoring sheets for the LESS and LESS-RT did not show high levels of reliability within this sample of raters. The LESS-Full sheet was shown to be slightly more reliable between the groups of raters (R = 0.575 ± 0.169) than the LESS-RT sheet (R = 0.435 ± 0.101). It should be noted that in previous research evaluating the validity of the LESS against motion capture (considered the gold-standard), certain items within the scoring criteria showed poor and moderate agreement with motion capture (Onate et al., 2010). Based on these findings the investigators concluded that the validity of the LESS was most likely item dependent. Additionally, studies that compare novice versus expert rater reliability with the LESS have often reported small sample sizes, or provide training to the novice raters prior to scoring the task (Padua et al., 2011; Hanzlíková et al., 2020; Markbreiter et al., 2015; Onate et al., 2010). These factors could have contributed to the lower reliability between sheets, as well as the reliability of the rater groups.

This study was not without limitations. Namely, a reference rater that has shown excellent reliability with the scoring sheets was not utilized when analyzing the raters' reliability. Without a reference, it is unclear whether any raters in this study were accurate with their scoring, and any raters that demonstrated acceptable reliability could have been incorrect with their scoring across all videos and sheets consistently.

To the authors' knowledge, this is the first study to assess the inter- and intra-rater reliability of the LESS and LESS-RT among novice raters (minimal/no experience or training with the LESS or LESS-RT) consisting of university students. As exercise science students are often recruited to assist with large-scale testing and assessment for both athletic and tactical populations, it is imperative to provide training and/or experience with the screens administered. Prior experience or practice with the LESS, as these data suggest, may be the limiting factor in the quality of assessment being graded. Regardless of the education level of raters, experience with the screen may have affected the overall reliability observed. Therefore, providing a single training session (as short as one hour; Onate et al., 2010) could drastically improve the quality of ratings, even from novice raters.

Acknowledgements: This research was partially funded through a grant provided by the National Strength and Conditioning Association Foundation. The authors would like to thank AR and KW for their assistance in participant recruitment, data collection, and preparation for the research.

Conflict of interest: All authors declare that they have no conflict of interest relevant to the content of this article.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*, 217-238.

Beutler, A. I., de la Motte, S. J., Marshall, S. W., Padua, D. A., & Boden, B. P. (2009). Muscle strength and qualitative jump-landing differences in male and female military cadets: The jump-ACL study. *Journal of Sports Science & Medicine*, *8*(4), 663.

Buckthorpe, M. W., Hannah, R., Pain, T. G., & Folland, J. P. (2012). Reliability of neuromuscular measurements during explosive isometric contractions, with special reference to electromyography normalization techniques. *Muscle & Nerve, 46*(4), 566-576.

Everard, E., Lyons, M., & Harrison, A. J. (2018). Examining the association of injury with the Functional Movement Screen and Landing Error Scoring System in military recruits undergoing 16 weeks of introductory fitness training. *Journal of Science and Medicine in Sport, 21*(6), 569-573.

Everard, E., Lyons, M., & Harrison, A. J. (2019). Examining the reliability of the Landing Error Scoring System with raters using the standardized instructions and scoring sheet. *Journal of Sport Rehabilitation, 29*(4), 519-525.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Hanzlíková, I., & Hébert-Losier, K. (2020). Is the Landing Error Scoring System reliable and valid? A systematic review. *Sports Health, 12*(2), 181-188.

Hopkins, W., Marshall, S., Batterham, A., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine & Science in Sports & Exercise, 41*(1), 3.

Kernozek, T. W., Torry, M. R., & Iwasaki, M. (2008). Gender differences in lower extremity landing mechanics caused by neuromuscular fatigue. *The American Journal of Sports Medicine, 36*(3), 554-565.

Mala, J., Szivak, T. K., Flanagan, S. D., Comstock, B. A., Laferrier, J. Z., Maresh, C. M., & Kraemer, W. J. (2015). The role of strength and power during performance of high intensity military tasks under heavy load carriage. *US Army Medical Department Journal, 11*(4), 3-11.

Markbreiter, J. G., Sagon, B. K., McLeod, T. C. V., & Welch, C. E. (2015). Reliability of clinician scoring of the Landing Error Scoring System to assess jump-landing movement patterns. *Journal of Sport Rehabilitation, 24*(2), 214-218.

Navalta, J. W., Stone, W. J., & Lyons, T. S. (2019). Ethical issues relating to scientific discovery in exercise science. *International Journal of Exercise Science, 12*(1), 1.

Onate, J., Cortes, N., Welch, C., & Van Lunen, B. (2010). Expert versus novice interrater reliability and criterion validity of the Landing Error Scoring System. *Journal of Sport Rehabilitation, 19*(1), 41-56.

Orr, R. M., Johnston, V., Coyle, J., & Pope, R. (2015). Reported load carriage injuries of the Australian army soldier. *Journal of Occupational Rehabilitation, 25*, 316-322.

Padua, D. A., Boling, M. C., DiStefano, L. J., Onate, J. A., Beutler, A. I., & Marshall, S. W. (2011). Reliability of the Landing Error Scoring System-real time, a clinical assessment tool of jump-landing biomechanics. *Journal of Sport Rehabilitation*, *20*(2), 145-156.

Padua, D. A., DiStefano, L. J., Beutler, A. I., De La Motte, S. J., DiStefano, M. J., & Marshall, S. W. (2015). The Landing Error Scoring System as a screening tool for an anterior cruciate ligament injury–prevention program in elite-youth soccer athletes. *Journal of Athletic Training, 50*(6), 589-595.

Padua, D. A., Marshall, S. W., Boling, M. C., Thigpen, C. A., Garrett Jr, W. E., & Beutler, A. I. (2009). The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: the jump-ACL study. *The American Journal of Sports Medicine, 37*(10), 1996-2002.

Scott, S. A., Simon, J. E., Van Der Pol, B., & Docherty, C. L. (2015). Risk factors for sustaining a lower extremity injury in an Army Reserve Officer Training Corps cadet population. *Military Medicine, 180*(8), 910-916.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research, 19*(1), 231-240.

Weir J. P., & Vincent W. J. (2020). Quantifying Reliability. In J.P Weir & W. J. Vincent (Eds.), *Statistics in Kinesiology* (pp. 213-228). Human Kinetics, Champaign, IL.

World Medical Association. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization, 79*(4), 373.