Brazilian Journal of Physical Therapy 2019;23(6):541–548

Brazilian Journal of Physical Therapy

https://www.journals.elsevier.com/brazilian-journal-of-physical-therapy

ORIGINAL RESEARCH

# Reliability of two functional clinical tests to evaluate trunk and lumbopelvic neuromuscular control and proprioception in a healthy population

**Cedric De Blaiser***, **Roel De Ridder**, **Tine Willems**, **Lieven Danneels**, **Philip Roosen**

*Department of Rehabilitation Sciences, Ghent University, Campus UZ Gent, Ghent, Belgium*

Received 19 July 2018; received in revised form 25 October 2018; accepted 26 October 2018
Available online 10 November 2018

**KEYWORDS**
Reliability;
Proprioception;
Injury prevention;
Screening;
Lumbopelvic hip complex;
Core stability

**Abstract**
*Objectives:* The need to accurately assess trunk and lumbopelvic proprioception and neuromuscular control is widely accepted. However, based on current literature, there is a lack of reliable clinical tests to evaluate these aspects in clinical practice. The objective of this study is to investigate intra- and inter-tester reliability of the lateral step down test and the lumbopelvic position–reposition test in a healthy population.
*Methods:* Protocol and scoring methods were developed for the lateral step down test and lumbopelvic position–reposition test, used to assess trunk and lumbopelvic neuromuscular control and proprioception respectively. Each test was performed once by thirty participants and video analysis for test scoring was performed. Three items on the lateral step down test were scored to evaluate neuromuscular control and, four items on the lumbopelvic position–reposition test were scored to evaluate proprioception. Aggregate scores for each test were calculated based on the separate item scores. Intraclass correlation coefficients and linear weighted kappa coefficients were determined for intra- and inter-tester reliability.
*Results:* Based on the aggregate score, excellent intra- and inter-tester reliability (ICC (2,1) = 0.73–0.88) was found for both tests. Moderate/almost perfect intra-and inter-tester agreement ($K$ = 0.62–0.91) was found for the separate items of the lateral step down test and fair/substantial agreement ($K$ = 0.25–0.76) for the items of the lumbopelvic position–reposition test.
*Conclusion:* Current testing protocol and scoring method for the lateral step down test is reliable. Adjustments for the scoring method of the lumbopelvic position–reposition test are warranted to improve reliability.


* Corresponding author at: Department of Rehabilitation Sciences, Faculty of Medicine and Health Sciences, Ghent University, Campus UZ Gent, Corneel Heymanslaan 10, 2B3, Entrance 46, BE-9000 Ghent, Belgium.
*E-mail:* Cedric.DeBlaiser@Ugent.be (C. De Blaiser).

https://doi.org/10.1016/j.bjpt.2018.10.014
1413-3555/

## Introduction

Assessment of core stability has gained widespread attention over the last decade since it is associated with low back pain, musculoskeletal injury risk, and athletic function.[1–5] Core stability is defined as dynamic trunk and lumbopelvic control that allows for production and regulation of force, which is transferred throughout the kinetic chain during movement.[4] Core stability requires the integration of core muscle strength, endurance, neuromuscular control and proprioception.[2,6,7] Neuromuscular control is the ability to produce efficient movement during task performance as a result of precisely coordinated muscular activity at the right time, for the correct duration and with the right combination of forces.[8] Proprioception is the ability to sense joint position and movement based on afferent sensory input from joints, tendons, and associated deep tissue proprioceptors.[9]

Inadequate core neuromuscular control and proprioception have been proven to be risk factors for lower extremity injuries since they compromise dynamic joint stability and lead to altered movement patterns.[7,10] In general, neuromuscular control is assessed by evaluating motion quality and control during a specific movement whereas proprioception of a segment can be assessed with motion detection threshold testing or through evaluating the ability to position and reposition the segment in a specific posture.[9,11] These aspects are often evaluated in lab situations with expensive or not commercially available equipment and have a limited use in clinical practice compared to clinical screening tests.

The lateral step down test[12–14] and the lumbar position–reposition test[15] are suggested as feasible clinical tests to evaluate trunk and lumbopelvic neuromuscular control and proprioception respectively and, they require a minimal amount of equipment to perform. Adequate reliability of such clinical screening tests is relevant since it is a prerequisite for test validity.[16–18] Although evaluation of reliability of these tests has been performed in current literature, there is contradictory evidence for reliability of the lateral step down test and, reliability of the lumbar position–reposition test was based on a scoring method using laboratory equipment.[12–15]

In conclusion, reliable tests with feasible scoring methods to measure trunk and lumbopelvic neuromuscular control and proprioception are lacking. The available protocols for the lateral step down test and the lumbar position–reposition test were adapted into novel test protocols and scoring methods and, scoring criteria and video based test assessment were developed. The aim of this study is to investigate the intra- and inter-tester reliability of the novel lateral step down test and lumbopelvic position–reposition test, which evaluate trunk and lumbopelvic neuromuscular control and proprioception respectively, by comparing video based test scores in a healthy population.

## Methods

### Participants

Sixty healthy participants over the age of 18, enrolled in teacher training studies at the Artevelde University College and the University College Ghent, voluntarily participated and were randomly assigned to perform one of both tests. Thirty participants (17♂, 13♀; age: 19.38 ± 1.04 years; BMI (body mass index): 22.34 ± 2.25 kg/m$^2$) were selected for the lateral step down test and thirty participants (15♂, 15♀; age: 19.5 ± 1.5 years; BMI: 21.44 ± 1.94 kg/m$^2$) were selected for the lumbopelvic position–reposition test. The following exclusion criteria were used: A history of traumatic pelvic and/or trunk injury; a history of acute, sub-acute or chronic low back pain or neurological diseases or disorders.

### Study design

Tests were performed in a controlled clinical setting and, video based test scoring was independently performed after test completion by two examiners. The same examiner provided test instructions and supervised all test performances. Intra-tester reliability was investigated by comparing two test scores of the same examiner (given two weeks apart to prevent recall bias). Inter-tester reliability was investigated by comparing test scores given independently by 2 different examiners. The examiners are experienced physical therapists (6 and 13 years) and were trained to standardize test scoring. One training session was held to familiarize the examiners with the scoring protocols. Afterwards, both examiners scored 15 test subjects for each test and a consensus meeting was held to compare scores and fine-tune scoring methods. The local Commissie voor Medische Ethiek (UZ Gent)/Commission for Medical Ethics (UZ Ghent) ethics committee approved the study protocol number: (2014/0780). Participants gave their written informed consent prior to participation. Public trials registry number: NCT03343379 (https://clinicaltrials.gov/ct2/show/NCT03343379).

### Test protocol and scoring method development

Development of the test protocols and scoring methods involved a combination of judgment and expert consensus opinion methodology.

An extensive literature search was performed on the topic of neuromuscular control, proprioception and the possible assessment methods. Investigator agreement was reached on a dynamic movement control test and an active position–reposition test as appropriate clinical assessment tools to measure trunk and lumbopelvic neuromuscular control and proprioception respectively. Existing tests were used as basis for protocol development and were adapted with regards to standardization.

Items for scoring, item scoring descriptions and scoring methods were determined for each test and a consensus opinion method was used in which a panel of expert clinicians ($n$ = 3; between 13 and 25 years of experience in motor control, low back pain and musculoskeletal rehabilitation) was involved in testing and commenting on the appropriateness of the items, their descriptions and the scoring methods.

## Test protocols

### Lateral step down test protocol

The current protocol was based on existing tests.[12–14] Participants stood on the edge of a 30 cm high box (25 cm for people shorter than 170 cm) in unipodal stance (foot pointed forward) and were asked to perform a series of lateral step downs, which were executed at a self-chosen, comfortable speed with the contralateral knee extended and contralateral hip slightly flexed. The correct test performance was demonstrated by the same examiner while the following instructions were given for standardization: ''Cross arms at the chest and do not use them to keep balance. Keep looking forward. Bend the knee and lower the free leg toward the ground in a controlled manner. Keep the trunk upright. Keep the pelvis horizontal. Keep the knee of the stance leg on a line which connects the hip with the second toe of the stance leg. When the opposite heel touches the ground, do not transfer weight and return to the initial position. Repeat this sequence 5 times.''

Two practice trials were allowed to provide feedback in case of a faulty test performance, followed by the actual execution of five consecutive lateral step downs. Trials were repeated if the subject failed to touch the ground with the contralateral heel or stepped off the box. For this study, only one leg was tested in a randomized order.

### Lumbopelvic position–reposition test protocol

The protocol was based on the study of Stevens et al.[15] Proprioception was evaluated by determining the repositioning accuracy of the trunk and lumbopelvic regions into a reference position, after having actively moved from full spinal flexion to extension during sitting. The participant was seated with 85 degrees of knee flexion, feet placed at hip width and arms hanging freely alongside the trunk. The participant was placed in a neutral spine reference position. The neutral position was halfway between full extension and a flat position of the spine.[19] The participant was then asked to perform three movements between spinal flexion and extension at a self-chosen, comfortable speed. After performing these movements, the participant was asked to accurately reassume the reference position.

## Test scoring

Tests were filmed using a video camera (Canon Legria hfG10, Canon Inc., Tokyo, Japan; 25 frames/second; effective resolution: 1.56 pixels). For both tests, the camera was placed at a distance of 3 m and a height of 50 cm. The camera was placed in line with the stance leg for the lateral step down test and filmed the test in the frontal plane. It was placed in line with the lumbar spine for the lumbopelvic position–reposition test and filmed the test in the sagittal plane.

### Lateral step down test

Movement patterns of the trunk, pelvis and lower extremity during the step down were evaluated in three items: dynamic balance (1), knee valgus/hip internal rotation (HIR) (2) and pelvic control (3). A score of 0, 1, 2 or 3 could be attributed to each item. Item scores were combined into an aggregate score with a minimum of 0 and a maximum of 9 points with 9 the best possible score. An open source video editing program for analyzing kinematics (Kinovea, Version 0.8.15) was used for test scoring. This program enabled the examiners to indicate reference points, draw lines between body segments and calculate angles in order to quantify movement pattern deviations. In general, examiners were instructed to score the items with as few as possible video playback repetitions, however, repeated viewing was allowed when in doubt. An overview of the scoring method is presented in Table 1.

For 'dynamic balance', movements of the trunk and free leg were evaluated during step downs. 'Dynamic balance' was evaluated in general for the 5 consecutive trials with normal video playback speed. Perfect balance (no movement of the trunk or free leg for 5 repetitions) received a score of 3. Small movement of trunk or free leg in 2 or more repetitions received score 2. Score 1 corresponded with moderate movement of trunk or free leg in 2 or more repetitions. Score 0 was given in case of large movement of the trunk or free leg in 2 or more repetitions or when the hands were not held on the shoulders during the execution of the test. In case of doubt between two scores the lower score was given.

Scoring the items 'knee valgus/HIR' and 'pelvic control' was performed on one and the same of the 5 step downs. It is the objective of the scoring for these items to assess the best representative of the 5 repetitions. The examiner viewed 5 consecutive repetitions and was instructed to choose the best one for scoring if 5 similar repetitions were observed. However, if variability between the different repetitions was seen, the best and the worst quality repetitions were left out and the best repetition of the remaining repetition was chosen as the best representative. A good performance on the lateral step down test was in accordance with the standardized instructions given to the subject as described above.

For the item 'knee valgus/HIR', the video was slowed down to half speed with Kinovea and paused at the moment of heel-contact of the contralateral leg for the chosen repetition. The amount of knee valgus was measured to evaluate medial knee collapse and consequently the amount of functional hip internal rotation[20]. Therefore, a line was drawn from the hip toward the second metatarsal (Fig. 1). If this line passed through the center of the patella, the participant scored 3 or 2. Differentiation between a score 2 and 3 was based on normal speed video playback: score 2 if the knee made small oscillating movements around the neutral line during the step down; score 3 if a perfect steady step down was performed with no oscillating movements. If the center patella was not in line with the second metatarsal and the hip joint at the moment of heel-contact, the participant scored 1 or 0: score 1 if the middle of the patella was above or lateral of the hallux (evaluated with a vertical line drawn upwards from the hallux); score 0 if the center patella was medial of the hallux.

'Pelvic control' was assessed by evaluating the height difference of the left and right ASIS in Kinovea. The video was slowed down to half speed with Kinovea and paused at the moment of heel-contact. The left and right ASIS were connected with a line and the angle with the horizontal plane was calculated (Fig. 1). Score 3 was given when the line

Table 1 Scoring method used for the lateral step down test and the lumbopelvic position–reposition test.

| Point value | Lateral step down test | | | |
|---|---|---|---|---|
| | Dynamic balance | Knee valgus/HIR | Pelvic control | Total score |
| 0 | Large movement of trunk or free leg | Patella medial to hallux | ASIS angle $\geq 20°$ | /9 |
| 1 | Moderate movement of trunk or free leg | Patella straight above hallux | ASIS angle between $10°$ and $\leq 20°$ | |
| 2 | Small movement of trunk or free leg | Patella on hip-MT2 line (oscillations) | ASIS angle between $0°$ and $\leq 10°$ | |
| 3 | No movement of trunk or free leg | Patella on hip-MT2 line (no oscillations) | ASIS on the same level | |

| Point value | Lumbopelvic position–reposition test | | | | |
|---|---|---|---|---|---|
| | Position of the pelvis | Lumbar spine curvature | Thoracic spine curvature | Inclination of the thorax | Total score |
| 0 | Impossible to reposition | Impossible to reposition | Large deviation | Large deviation | /10 |
| 1 | Large deviation | Large deviation | Small deviation | Small deviation | |
| 2 | Small deviation | Small deviation | Perfect repositioning | Perfect repositioning | |
| 3 | Perfect repositioning | Perfect repositioning | NA | NA | |

ASIS, anterior superior iliac spine; HIR, hip internal rotation; MT2, metatarsal 2; NA, not applicable.

connecting both ASIS was horizontal. An angle between $0°$ and $\leq 10°$ received score 2. An angle between $10°$ and $\leq 20°$ resulted in score 1. Score 0 was given with an angle $>20°$ or in case of explicit weight transfer on the contralateral leg.

### Lumbopelvic position–reposition test

A scoring method was developed based on repositioning of the trunk and lumbopelvic regions. The recording of this test was viewed once and a screenshot of the neutral reference position and a screenshot of the repositioning into the reference position after three movement cycles were compared. Four items were evaluated: position of the pelvis (1), lumbar spine curvature (2), thoracic spine curvature (3) and inclination of the thorax (4). Item scores were combined into an aggregate score ranging from 0 to 10, with 10 being the best score. Possible scoring options for the first two items were: score 3: repositioning = original position (perfect repositioning, no differences between the two screenshots for this region); score 2: small deviation (near perfect repositioning, small difference discernible between the two screenshots for this region); score 1: large deviation (undeniable difference between the two screenshot for this region); score 0: impossibility to reposition (repositioning is impossible due to obvious coordination dysfunction). Possible scoring options for item 3 and 4 were: score 2: repositioning = original position (perfect repositioning, no differences between the two screenshots for this region); score 1: small deviation (near perfect repositioning, small difference discernible between the two screenshots for this region); score 0: large deviation (undeniable difference between the two screenshot for this region). An overview of the scoring method is presented in Table 1.

## Statistical analyses

Intra- and inter-tester reliability, based on the aggregate scores of the lateral step down test and the lumbopelvic position–reposition test were calculated. Intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) were calculated with a two-way random effects model with single measure reliability (ICC (2,1)). Interpretation of the ICC values was in accordance with Cicchetti and Sparrow.[21] Standard error of measurement (SEM) was calculated ($SEM = SD\sqrt{1 - ICC}$). Minimal detectable changes (MDC), based on the 95% confidence intervals, were calculated with $1.96 \times \sqrt{2} \times SEM$.

Intra- and inter-tester reliability of the ordinal scores on the separate test items of both tests was investigated with linear weighted kappa coefficients and the corresponding percentage agreement. Interpretation of the kappa coefficients was in accordance with Landis and Koch.[22] For each separate item, the percentage agreement between two given scores was calculated. Sample size requirements to achieve a significance level of 0.05 with a power of 0.8 were


**Figure 1** Screenshot with reference lines and angle calculation for the items 'knee valgus/HIR' and 'pelvic control' of the lateral step down test.

calculated based on the study of Walter et al.[23] Statistical analyses were conducted with the statistical software package SPSS (v25.0).

## Results

Descriptive statistics for both tests are presented in Table 2. ICC values with 95% confidence intervals and accompanying SEM and MDC values are presented in Table 3. Good to excellent intra- and inter-tester reliability values were found for the aggregate scores for both tests.

Kappa coefficients with 95% confidence intervals and accompanying percentage agreement are presented in Table 4. Intra-tester agreement for the items 'dynamic balance' and 'knee valgus/HIR' of the lateral step down test revealed almost perfect agreement and, substantial intra-tester agreement was found for the item 'pelvic control'. Inter-tester agreement for all three items of the lateral step down test was substantial.

Intra-tester agreement for the lumbopelvic position–reposition test was substantial for the items 'position of the pelvis', 'lumbar spine curvature' and 'inclination of the thorax' and moderate for the item 'thoracic spine curvature'. Inter-tester agreement for the item 'position of the pelvis' and 'lumbar spine curvature' was substantial. Inter-tester agreement for the items 'thoracic spine curvature' and 'inclination of the thorax' was fair.

## Discussion

The aim of this study was to investigate reliability of the current scoring methods for the lateral step down test and the lumbopelvic position–reposition test to measure trunk and lumbopelvic neuromuscular control and proprioception in a healthy population.

Based on the ICC value (with 95% CI) of the aggregate scores on the lateral step down test, good to excellent intra- and inter-tester reliability can be expected. The separate items showed substantial to almost perfect intra- and inter-tester agreement. The lowest kappa coefficient was shown for the intra-tester agreement of the item 'pelvic control', but was nonetheless above 0.61, which is used as clinically acceptable agreement.[24–28] Based on the expected good to excellent reliability of the aggregate scores, and the substantial to perfect agreement on the separate items, the authors of this study consider the current scoring method of the lateral step down test reliable for use in clinical practice with healthy participants. However, caution is needed for the item 'pelvic control'. The 95% CI for the kappa coefficient was 0.36–0.84, which possibly results in low intra-tester reliability when scoring this item separately. This might be the result of subjective anatomical reference determination in Kinovea. It is advised to mark both ASIS reference points with tape to aid scoring of this item.

The results are in agreement with Piva et al.[13] They evaluated a lateral step down on 5 items and depending on the movement deviation severity, 1 or 2 points were added per item and, based on the total scores, participants were classified as having poor (score 4–10), medium (score 2–3) or good quality (score 0–1) of movement. However, reliability was calculated on the aggregate score without taking into account the separate items. This could result into reliably classifying two participants as having the same

**Table 2** Descriptive statistics of the lateral step down test and lumbopelvic position–reposition test.

| | *N* | Mean | SD | Min–Max |
|---|---|---|---|---|
| Lateral step down test (Examiner 1) | 30 | 4.5/9 | ±1.2 | 1/9 – 7/9 |
| Lateral step down test (Examiner 2) | 30 | 4.4/9 | ±1.4 | 0/9 – 8/9 |
| Lumbopelvic position–reposition test (Examiner 1) | 30 | 6.9/10 | ±1.8 | 2/10 – 10/10 |
| Lumbopelvic position–reposition test (Examiner 2) | 30 | 7.3/10 | ±1.4 | 5/10 – 10/10 |

*N*, number of subjects; SD, standard deviation.

**Table 3** Intra- and inter-tester reliability of the lateral step down test and lumbopelvic position–reposition test (ICC values, based on the aggregate test scores, with corresponding SEM values and MDC values).

| | *N* | ICC (2,1) | 95% CI for ICC | SEM[a] | MDC[a] |
|---|---|---|---|---|---|
| Intra-tester reliability Lateral step down test | 30 | 0.88 | 0.76–0.94 | 0.44 | 1.23 |
| Inter-tester reliability Lateral step down test | 30 | 0.81 | 0.64–0.9 | 0.51 | 1.42 |
| Intra-tester reliability Lumbopelvic position–reposition test | 30 | 0.87 | 0.75–0.94 | 0.64 | 1.76 |
| Inter-tester reliability Lumbopelvic position–reposition test | 30 | 0.73 | 0.51–0.86 | 0.85 | 2.34 |

*N*, number of subjects; CI, confidence interval; SEM, standard error of measurement; MDC, minimal detectable change.
[a] SEM values and MDC values are represented on a scale from 0 to 9 for the lateral step down test; on a scale from 0 to 10 for the lumbopelvic position–reposition test.

**Table 4** Intra- and inter-tester reliability of the different items of the lateral step down test and the lumbopelvic position–reposition test (linear weighted kappa coefficients, with percentage agreement between examiners for each item scored).

| | *N* | Kappa | 95% CI for kappa | % agreement |
|---|---|---|---|---|
| *Lateral step down test* | | | | |
| Intra-tester reliability Dynamic balance | 30 | 0.81 | 0.66–0.97 | 83.3 |
| Intra-tester reliability Knee valgus/HIR | 30 | 0.91 | 0.74–1 | 96.7 |
| Intra-tester reliability Pelvic control | 30 | 0.62 | 0.36–0.84 | 76.7 |
| Inter-tester reliability Dynamic balance | 30 | 0.65 | 0.45–0.85 | 70 |
| Inter-tester reliability Knee valgus/HIR | 30 | 0.75 | 0.51–0.98 | 90 |
| Inter-tester reliability Pelvic control | 30 | 0.78 | 0.57–0.99 | 86.7 |
| *Lumbopelvic position–reposition test* | | | | |
| Intra-tester reliability Position of the pelvis | 30 | 0.62 | 0.39–0.86 | 70 |
| Intra-tester reliability Lumbar spine curvature | 30 | 0.76 | 0.56–0.96 | 83.3 |
| Intra-tester reliability Thoracic spine curvature | 30 | 0.5 | 0.19–0.81 | 70 |
| Intra-tester reliability Inclination of the thorax | 30 | 0.7 | 0.47–0.94 | 83.3 |
| Inter-tester reliability Position of the pelvis | 30 | 0.61 | 0.32–0.87 | 80 |
| Inter-tester reliability Lumbar spine curvature | 30 | 0.72 | 0.49–0.95 | 83.3 |
| Inter-tester reliability Thoracic spine curvature | 30 | 0.29 | 0.01–0.57 | 60 |
| Inter-tester reliability Inclination of the thorax | 30 | 0.25 | 0.01–0.55 | 56.7 |

CI, confidence interval; HIR, hip internal rotation; *N*, number of subjects.

quality of movement, whilst having very dissimilar test performances. Furthermore, their test was designed to detect altered movement patterns in patients with patellofemoral pain syndrome. It is possible that the step down height (20 cm) in their study was not physically demanding enough to identify inadequate neuromuscular control in healthy subjects. Therefore, the step down height for the current study was adjusted to 30 cm as proposed by Norcross et al.[29]

Chmielewski et al.[12] did not find high agreement between raters for scoring the lateral step down test. They scored excessive, moderate, small or no movement deviation from the neutral position for the trunk, pelvis and hip during a similar test protocol. These less than favorable results were attributed to the absence of explicit scoring guidelines for the examiners and a timed scoring period of 30 seconds after viewing the test performance.

The protocol for the lumbopelvic position−reposition test was based on the study of Stevens et al.[15] However, they evaluated repositioning capacities based on ultrasound 3D movement analysis, which was reliable but its use in clinical practice is limited. Therefore, in our study, four different items were developed to quantify repositioning capacities of the trunk and lumbopelvic region.

Based on the ICC value (and 95% CI) of the aggregate scores on the lumbopelvic position−reposition test, moderate to excellent reliability can be expected for the evaluation of the test. Substantial agreement was found for the intra- and inter-tester reliability for the separate items 'position of the pelvis' and 'lumbar spine curvature', with all kappa coefficients for these items obtaining clinically acceptable agreement. On the other hand, only fair to substantial agreement was found for the items 'thoracic spine curvature' and 'inclination of the thorax' of which only intra-tester reliability for 'inclination of the thorax' obtained clinically acceptable agreement. Although acceptable reliability can be expected when using the aggregate score, it is not recommended to use the lumbopelvic position−reposition test and its scoring method in the current form. Agreement for the items 'thoracic spine curvature' and 'inclination of the thorax' is not clinically acceptable, especially when multiple examiners are used for comparing test scores. If the same examiner performs scoring, acceptable agreement can be expected. However, taken into account the 95% CI, it is warranted for future research to adjust the scoring method in order to improve agreement on these separate items. More specific criteria, combined with the use of Kinovea could improve the scoring and the agreement for these two items.

With each test taking less than 10 min to perform and score, no extra costs attached, simple training needed and detailed operating instructions for scoring readily available, both tests are considered as clinically feasible.[30]

A few methodological considerations need to be taken into account. First, it should be noted that only two raters were used to perform inter-tester reliability since using three or more raters should provide a better representation of the reliability. However, in a similar study, Chmielewski et al.[12] found comparable results when calculating reliability coefficients for two raters and three raters. Second, anatomical reference points were used for measuring angles and quantifying movement deviations during the lateral step down test with Kinovea. However, no markers were used during test performance to indicate these reference points. Nonetheless, reliable results were yielded when using the current method of determining anatomical reference points to evaluate movement deviations except for the item 'pelvic control'. Using tape to mark anatomical reference points might improve reliability for this item.

## Conclusion

Scoring the lateral step down test with video analysis and the software program Kinovea is reliable. This test can be performed and scored in clinical practice to evaluate trunk and lumbopelvic neuromuscular control in a healthy population. On the contrary, scoring the lumbopelvic position−reposition test using the novel scoring method is not recommended for use in clinical practice due to low inter-tester agreement on two items. Future research should focus on defining more specific criteria to score the items 'thoracic spine curvature' and 'inclination of the thorax' for this test. Furthermore, investigating concurrent validity for both tests is warranted and could be executed with 3D video analysis. As such, the performance quality of the separate items can be compared against criterion standards.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Conflicts of interest

The authors declare no conflicts of interest.

## References

1. Oliveira IO, Pilz B, Santos RLG, et al. Reference values and reliability for lumbopelvic strength and endurance in asymptomatic subjects. *Braz J Phys Ther*. 2018;22(1):33−41.
2. Leetun DT, Ireland ML, Willson JD, et al. Core stability measures as risk factors for lower extremity injury in athletes. *Med Sci Sports Exerc*. 2004;36:926−934.
3. Wilkerson GB, Giles JL, Seibel DK. Prediction of core and lower extremity strains and sprains in collegiate football players: a preliminary study. *J Athl Train*. 2012;47:264−272.
4. Kibler WB, Press J, Sciascia A. The role of core stability in athletic function. *Sports Med*. 2006;36:189−198.
5. Lust KT, Sandrey MA, Bulger SM, et al. The effect of 6-week training programs on throwing accuracy, proprioception, and core endurance in baseball. *J Sport Rehabil*. 2009;18:407−426.
6. Sahrmann S, Azevedo DC, Dillen LV. Diagnosis and treatment of movement system impairment syndromes. *Braz J Phys Ther*. 2017;21(6):391−399.
7. Zazulak BT, Hewett TE, Reeves NP, et al. The effects of core proprioception on knee injury. *Am J Sports Med*. 2007;35:337−368.
8. Comerford MJ, Mottram SL. Movement and stability − contemporary developments. *Man Ther*. 2001;6:15−26.
9. Hewett TE, Paterno MV, Myer GD. Strategies for enhancing proprioception and neuromuscular control of the knee. *Clin Orthop Relat Res*. 2002;402:76−94.
10. Verrelst R, De Clerq D, Vanrenterghem J, et al. The role of proximal dynamic joint stability in the development of exertional medial tibial pain: a prospective study. *Br J Sports Med*. 2013;48:388−393.
11. Zazulak BT, Cholewicki J, Reeves NP. Neuromuscular control of trunk stability: clinical implications for sports injury prevention. *J Am Acad Orthop Surg*. 2008;16:497−505.
12. Chmielewski TL, Hodges MJ, Horodyski M, et al. Investigation of clinician agreement in evaluating movement quality during

unilateral lower extremity functional tasks: a comparison of 2 rating methods. *J Orthop Sports Phys Ther*. 2007;37:122–129.
13. Piva SR, Fitzgerald K, Irrgang JJ, et al. Reliability of measures of impairments associated with patellofemoral pain syndrome. *BMC Musculoskelet Disord*. 2006;7:33.
14. Rabin A, Kozol Z. Measures of range of motion and strength among healthy women with differing quality of lower extremity movement during the lateral step-down test. *J Orthop Sports Phys Ther*. 2010;40:792–800.
15. Stevens VK, Bouche KG, Mahieu NN, et al. Reliability of a functional clinical test battery evaluating postural control, proprioception and trunk muscle activity. *Am J Phys Med Rehabil*. 2006;85:727–736.
16. Evans K, Refshauge KM, Adams R. Trunk muscle endurance tests: reliability, and gender differences in athletes. *J Sci Med Sport*. 2007;10:447–455.
17. Hayen A, Dennis RJ, Finch CF. Determining the intra- and inter-observer reliability of screening tools used in sports injury research. *J Sci Med Sport*. 2007;10:201–210.
18. Thomas JR, Nelson JK, Silverman S. *Research Methods in Physical Activity*. 7th ed. Champaign, Illinois: Human Kinetics Publishers; 2015.
19. Danneels L, Coorevits P, Cools A, et al. Differences in electromyographic activity in multifidus muscle and the iliocostalis lumborum between healthy subjects and patients with subacute and chronic low back pain. *Eur Spine J*. 2002;11:13–19.
20. Hollman JH, Ginos BE, Kozuchowski J, et al. Relationship between knee valgus, hip-muscle strength, and hip-muscle recruitment during a single-limb step-down. *J Sport Rehabil*. 2009;18:104–117.
21. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic*. 1981;86:127–137.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
23. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101–110.
24. Brouwer S, Reneman MF, Dijkstra PU, et al. Test-retest reliability of the Isenhargen work systems functional capacity evaluation in patients with chronic low back pain. *J Occup Rehab*. 2003;13:207–218.
25. Hunt DG, Zuberbier OA, Kozlowski AL, et al. Reliability of the lumbar flexion, lumbar extension, and passive straight leg raise test in normal populations embedded within a complete physical examination. *Spine (Phila Pa 1976)*. 2001;26: 2714–2718.
26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
27. Nomden JG, Slagers AJ, Bergman GJD, et al. Interobserever reliability of physical examination of the shoulder girdle. *Man Ther*. 2009;14:152–159.
28. Smart KM, Curley A, Blake C, et al. The reliability of clinical judgements and criteria associated with mechanisms-based classifications of musculoskeletal pain in patients with low back pain disorders: a preliminary reliability study. *J Man Manip Ther*. 2010;18:102–110.
29. Norcross M, Halverson SD, Hawkey TJ, et al. Evaluation of the lateral step-down test as a clinical assessment of hip musculature strength. *Athl Train Sports Health Care*. 2009;1: 272–278.
30. Tyson SF, Brown P. How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far. *Clin Rehabil*. 2014;28:804–816.