ORIGINAL ARTICLE

Is a visual-based balance assessment reliable under potentially distracting conditions?

Jacob R Pierce MS LAT ATC CSCS[1,2], Adam C Raikes MS LAT ATC[1,2], Sydney Y Schaefer PhD[1,3]

JR Pierce, AC Raikes, SY Schaefer. Is a visual-based balance assessment reliable under potentially distracting conditions? Curr Res Concussion 2015;2(2):43-48.

BACKGROUND: Visual-based balance assessments have reliability levels that may decrease in the presence of auditory distraction.

OBJECTIVES: To determine whether clinicians could reliably visually assess balance while potentially distracted. The authors hypothesized a decrease in participant reliability and fewer reported balance 'errors' under potentially distracted conditions compared with undistracted conditions.

METHODS: On two separate occasions (undistracted and potentially distracted) separated by one week, six certified athletic trainers used four error criteria to 'score' 32 prerecorded video sets of individuals completing six standardized stance positions of a balance test. Two-way random absolute agreement intraclass correlation coefficients (ICCs) measured inter- and intrarater reliabilities of participants during undistracted and potentially distracted conditions for each stance position individually, and for the summed number of reported 'errors' across all stance positions. Repeated measures ANOVAs tested for effects of condition (undistracted versus potentially distracted) and stance position on the reported number of 'errors'.

RESULTS: For the undistracted condition, total error inter- and intrarater reliability ICCs were 0.93 and 0.92, respectively. For the potentially distracted condition, total error inter- and intrarater reliability ICCs were 0.89 and 0.92, respectively. There were no significant effects of condition or stance position on the number participant reported 'errors'.

DISCUSSION: Contrary to the hypothesis, participants were reliable not only under undistracted conditions, but also under potential distraction. These data suggest that using four error criteria rather than six may account for increased visual-based balance assessment reliability.

CONCLUSION: Participants were reliable at visually identifying balance errors, even under potentially distracting conditions.

Key Words: *Auditory distraction; Balance; Posture*

Balance is defined as controlling the body's centre of mass within the limits of stability with external disturbances present (1), and is critical for many activities of daily life. Although there are many instrumented technologies available for assessing balance, such as force platforms, accelerometers or camera-based motion capture, they may not be readily available to all clinicians. Thus, clinicians may rely on visual-based balance assessments to determine whether an individual has any impairment relative to some prespecified criteria. For example, a clinician may ask a patient to maintain a heel-to-toe (ie, 'tandem') stance position for a set duration of time and count how many times the patient steps out of that position (2).

One advantage of such assessments is their real-world utility outside of a laboratory. However, potentially distracting external stimuli outside of a controlled laboratory setting are not uncommon. The effects of such distractions on balance performance itself have been well documented (3-5), such that various measures of balance tend to be worse when performing two tasks at once or attending to external auditory stimuli such as crowd noise (6). Given the potential for distraction to affect balance performance, clinicians visually assessing balance may be distracted in the presence of uncontrolled or irrelevant stimuli. These stimuli may reduce assessment reliability in real-world settings. If such an assessment is to be used in a clinical setting, its reliability must also be known when the clinician is potentially distracted. Although visual-based balance assessments, such as the Balance Error Scoring System, have moderate reliability (7), how may this reliability change when clinicians are challenged to perform multiple tasks at the same time? Thus, the purpose of the present study was to determine whether clinicians could reliably visually identify balance errors under potentially distracted ($_{p}$D) conditions while using four balance 'error' criteria. We hypothesized that both the inter- and intrarater reliability of a valid visual-based balance assessment would be lower when the participants performed a concurrent auditory vigilance task ($_{p}$D) compared with when they did not (undistracted [$_{u}$D]). We also expected that the number of balance errors reported by the participants would be lower under $_{p}$D compared with $_{u}$D conditions because of the inability to attend to two tasks simultaneously.

METHODS

Participants

Six board-certified athletic trainers (mean [± SD] age 29.04±7.87 years) participated in the present study. Participants sat in front of a computer screen and viewed prerecorded videos of 32 unidentifiable individuals performing a balance task (8). Participants were instructed to visually rate the balancers' performance using four criteria from a valid assessment (8). The amount of clinical experience for participants in the present study ranged from 0.74 to 17.07 years.

All participants provided informed consent before participation in compliance with the university's institutional review board (#5218), as did each individual who was video-recorded in advance performing the balance task.

Experimental design

Balancer videos: Before recruiting participants for the present study, 32 individuals without a history of lower extremity injury in the previous year and no history whatsoever of any physician-diagnosed concussion were recorded. These individuals (referred to as 'balancers' [mean age 22.94±2.54 years]) were recorded using a handheld video camera (HC-V100, Panasonic, Japan) from a frontal view as they completed six stance positions. These stance positions were adapted directly from a valid visual-based assessment of balance (8). The balancers were instructed to perform all of the following: "Stand with both feet together" (double-leg, firm); "Stand on your right foot with your hip/knee of your 'up' leg both flexed to approximately 90°" (single-leg, firm; 3) "Stand heel-to-toe with your left foot in front of your right"

[1]Motor Rehabilitation and Learning Laboratory, Utah State University; [2]Utah State University Sports Medicine, Athletics Department, Logan; [3]Department of Physical Therapy, University of Utah, Salt Lake City, Utah

Correspondence: Mr Jacob R Pierce, Utah State University, Motor Rehabilitation and Learning Laboratory, 7000 Old Main Hill, Logan, Utah 84322, USA. Telephone 435-797-1894, e-mail jrpierce@me.com

TABLE 1
Balance 'error' criteria used by raters

Lifts hands off of iliac crests
Opens eyes
Steps/stumbles/falls
Takes >5 s to return to the testing position

(tandem, firm); "Stand with both feet together on the foam pad" (double-leg, foam); "Stand on your right foot on the foam pad with your hip/knee of your 'up' leg both flexed to approximately 90°" (single-leg, foam); and "Stand heel-to-toe with your left foot in front of your right on the foam pad" (tandem, foam). During the foam conditions, balancers stood on a 50.03 cm × 40.89 cm × 6.10 cm closed-cell foam balance pad (Airex AG Specialty Foams, Switzerland). Balancers were instructed to maintain each stance position for 20 s with eyes closed and hands on hips (eg, correct stance position) to the best of his or her ability. Additional details regarding these conditions have been published previously (2,9).

All videos were archived according to balancer and stance position such that each balancer yielded six separate 20 s video files, which equaled one set of videos. These sets were then inserted in random order offline by the experimenter into PowerPoint (Microsoft Corporation, USA) presentations for the participants to view.

Visual-based balance assessment

Participants in the present study were athletic trainer clinicians who visually 'scored' the balance performance recorded in the videos using four error criteria. All participants sat alone in front of a computer screen in a quiet room and were instructed to score the sets of videos using four criteria of balance performance. Table 1 provides a description of each balance 'error', which was explained to each participant at the beginning of the study. These criteria have been tested previously for validity (8) when used to assess balance performance in the six stance positions described above in balancer videos. Thus, the participants were asked to 'score' the number of balance errors they observed during each video and record this on a score sheet. If the participants observed multiple errors at the same time, such as opening of the eyes and falling out of position, they were instructed to record only one error to maintain criteria validity (8). Participants were also instructed not to count another error until the balancer had returned to the correct stance position. To ensure that the participants used the correct error criteria in their scoring, they were asked to list, from memory, the criteria by which they scored balance errors at the end of the study. All (100%) participants were able to correctly write the four criteria used.

Concurrent auditory vigilance task: Although an auditory task can affect balance itself (3-5), its effect on balance assessment is not well understood. Therefore, a concurrent auditory vigilance task was used in the present study to potentially distract the participants as they viewed the balance videos. During this task, the participants listened to a 20 s sequence of letters. In each sequence, the letters A, G, M and O were repeated in a pseudorandom order at a frequency of 1.75 Hz. The sequence started at the same time as the video and ran continuously for the duration of the video. Before each sequence, participants were instructed to "count the number of times you hear the letter <A, G, M, or O>." The instructed letter (A, G, M or O) was randomized for each trial. Immediately after each sequence, participants were asked to record the number of times a target letter was heard on their scoring sheet. Performance on the concurrent auditory vigilance task was quantified as the number of listening errors, which was calculated as the absolute difference between the actual and reported number of times the target letter was present in the sequence. For example, if the letter A was repeated nine times in the sequence and the participant reported only seven, then the number of listening errors would be two.

All sequences were prerecorded as audio (.mp3) files using an external microphone (Gigaware Omnidirectional model 33-119, Ignition LP, USA), embedded into the PowerPoint video files and played at a comfortable volume through headphones (MDR-V700, Sony Corporation, Japan). Participants completed two familiarization trials of the task before data collection, followed by two more trials that established his or her baseline performance on the concurrent auditory vigilance task alone. During these trials, participants remained seated with their eyes closed. Listening data from the participants was then collected during $_uD$ and $_pD$ conditions.

Because clinicians' ability to visually assess balance in real-world settings is likely more relevant for diagnostic purposes than their ability to attend to and process irrelevant auditory stimuli, participants' performance on the auditory vigilance task in either condition in the present study was not part of the hypothesis. Instead, listening error data in both the $_uD$ and $_pD$ conditions were collected and compared for the a priori power calculation. This indicated that at least five participants were needed to provide 80% power with an alpha level of 0.05 to detect attentional differences between baseline and $_pD$ conditions (see reference 10 for operational definition).

Establishing inter- and intrarater reliability under $_pD$ conditions

To establish how reliable the participants were in visually assessing balance through counting errors, participants first scored 32 video sets in random order without the concurrent auditory vigilance task (undistracted [$_uD$]; time 1). After each individual video, participants were instructed to write down on their scoring sheet the total number of balance errors they observed, according to the four criteria in Table 1. These data were used to determine the inter-rater reliability of these error criteria before any potential distractions were introduced in the study.

One week later (time 2), participants rescored 16 of the 32 sets of videos that they had observed during time 1 under $_uD$ conditions. Also during time 2, the participants scored the other 16 videos under $_pD$ conditions. The order in which the $_uD$ and $_pD$ sets were viewed was randomized among participants. The one-week separation between time 1 and time 2 was to avoid short-term learning, as recommended previously (11). These data were used to determine the intrarater reliability of these four error criteria without potential distraction; inter-rater reliability with a potential distraction; and overall scoring performance with a potential distraction.

Data and statistical analyses

Each set of videos comprised one 20 s video for each stance position. The participants reported the number of balance errors that they observed for each stance position (double-leg, firm; single-leg, firm; tandem, firm; double-leg, foam; single-leg, foam; tandem, foam). The reported the number of errors from the individual stance positions were also summed to yield a total score for each set of videos. These individual stance positions, along with the total score, were used to compute reliability measures and to compare the participants' use of the four error criteria.

To measure the inter- and intrarater reliability in both $_uD$ and $_pD$ conditions, two-way, random absolute agreement intraclass correlation coefficients ($ICC_{2,1}$) (12,13) were used. Inter-rater reliability of the $_uD$ scoring was computed across participants using the same 16 video sets scored and re-scored at times 1 and 2, respectively. Similarly, intrarater reliability of the $_uD$ scoring was computed within each participant using the same 16 video sets between times 1 and 2. More importantly, inter-rater reliability of the $_pD$ participants' scoring was computed across participants using the 16 video sets that were viewed while simultaneously performing the concurrent auditory vigilance task at time 2. Furthermore, intrarater reliability of the $_pD$ raters' scoring was computed within participants using the 16 video sets from time 1 ($_uD$) that were the same videos from time 2 ($_pD$). All ICC values were reported as 0.00 to 1.00, with larger values indicating higher reliability. Reliabilities were computed for all six separate stance positions as well as the total scores.

To test whether the participants reported fewer balance errors under $_uD$ compared with $_pD$ conditions, the mean total scores across raters was compared using a one-way repeated measures ANOVA with

TABLE 2
Inter-rater reliability intraclass correlation coefficients (ICCs) and and CIs

Stance position	ICC			CI		
	$_uD_1$	$_uD_2$	$_pD_2$	$_uD_1$	$_uD_2$	$_pD_2$
Double-leg, firm	0.00*	0.00*	0.00*	0.000*	0.000*	0.000*
Single-leg, firm	0.96	0.95	0.86	0.926–0.985	0.911–0.981	0.741–0.938
Tandem, firm	0.89	0.85	0.90	0.805–0.955	0.736–0.936	0.816–0.958
Double-leg, foam	0.00*	0.00*	0.00*	0.000*	0.000*	0.000*
Single-leg, foam	0.68	0.82	0.56	0.489–0.849	0.690–0.923	0.344–0.780
Tandem, foam	0.93	0.89	0.91	0.869–0.972	0.797–0.953	0.834–0.965
Total score	0.93	0.93	0.89	0.845–0.964	0.847–0.970	0.769–0.958

**Both double-leg stance ICCs and CIs could not be calculated due to a lack of errors. $_uD_1$ Time 1, undistracted; $_uD_2$ Time 2, undistracted; $_pD_2$ Time 2, potentially distracted*

TABLE 3
Intrarater reliability intraclass correlation coefficients (ICCs) and and CIs

Stance position	ICC		CI	
	$_uD_1$ versus $_uD_2$	$_uD_1$ versus $_pD_2$	$_uD_1$ versus $_uD_2$	$_uD_1$ versus $_pD_2$
Double-leg, firm	0.00*	0.00*	0.000*	0.000*
Single-leg, firm	0.96	0.95	0.926–0.985	0.911–0.981
Tandem, firm	0.89	0.85	0.805–0.955	0.736–0.936
Double-leg, foam	0.00*	0.00*	0.000*	0.000*
Single-leg, foam	0.68	0.82	0.489–0.849	0.690–0.923
Tandem, foam	0.93	0.89	0.869–0.972	0.797–0.953
Total score	0.93	0.93	0.845–0.964	0.847–0.970

**Both double-leg stance ICCs and CIs could not be calculated due to a lack of errors. $_uD_1$ Time 1, undistracted; $_uD_2$ Time 2, undistracted; $_pD_2$ Time 2, potentially distracted*

condition $_uD$ versus $_pD$ as a within-subject factor. To test whether the concurrent auditory vigilance task did, in fact, distract the participants, a 2×4 repeated measures ANOVA with condition $_uD$ versus $_pD$ and stance position (single-leg, firm versus tandem, firm versus single-leg, foam versus tandem, foam) was used as within-subject factors. While there were six separate stance positions in the current protocol (2), the two double-leg stance positions (firm and foam) yielded zero reported errors (see Results). Thus, only the four error-inducing stance positions were included in the ANOVA. Significant main effects of condition and stance position were expected on the reported number of errors, yet no interaction between condition and stance position. Significant effects were tested post hoc using Tukey-Kramer Honestly Significant Different tests (14,15). To test the effect of $_pD$ conditions on listening errors, a one-way ANOVA comparing the effect of condition (baseline versus while rating all six stance positions) was used. All ANOVAs and subsequent post hoc analyses were performed with alpha=0.05. Effect sizes were calculated for statistically significant differences using Cohen's d. All statistical analyses were performed using JMP version 10 (SAS Institute Inc, USA).

RESULTS

Table 2 summarizes the inter-rater reliability ICC values and CIs for each stance position and for the total scores. These scores are further subdivided by the 16 video sets at both times 1 and 2 $_uD$ and the 16 video sets at time 2 $_pD$. Compared with $_uD$ conditions at times 1 and 2, total score inter-rater ICC value was only reduced to 0.89 under $_pD$ conditions. $_pD$ conditions also decreased both the single-leg; firm and foam stance position ICC values when compared with times 1 and 2 $_uD$. Conversely, values improved under $_pD$ conditions during scoring of the tandem, firm stance position at times 1 and 2 $_uD$ as well as in the tandem, foam stance position at time 2 $_uD$.

To further determine reliability of this assessment, the intrarater reliability ICC values and CIs for each stance position and the total scores were calculated (Table 3). These scores are further subdivided by the 16 $_uD$ video sets at time 1 versus the same 16 $_uD$ video sets at time 2 and the 16 $_uD$ video sets at time 1 versus the same 16 video sets under $_pD$ conditions at time 2. The total score inter-rater ICC value was 0.93 during both the time 1 $_uD$ versus time 2 $_uD$ and time 1 $_uD$ versus time 2 $_pD$ conditions. Time 1 $_uD$ versus time 2 $_pD$ demonstrated reduced intrarater ICC values during single-leg, firm; single-leg, foam; and tandem, foam stance positions when compared with time 1 $_uD$ versus time 2 $_pD$. Conversely, time 1 $_uD$ versus time 2 $_pD$ resulted in a higher inter-rater reliability ICC value during the single-leg, foam stance position when compared with time 1 $_uD$ versus time 2 $_uD$.

Reported errors under $_uD$ and $_pD$ conditions

Figure 1A shows the reported number of errors for the total error score in $_uD$ and $_pD$ conditions. The one-way repeated measures ANOVA reported no significant effect of condition $_uD$ versus $_pD$ on the reported number of errors in the total error score ($F[1,5] = 0.045$; P=0.83). The 2×4 repeated measures ANOVA reported no significant main effect of condition ($F[1,5] = 0.101$; P=0.75) nor an interaction between condition and stance position ($F[3,5] = 0.057$; P=0.98) on the reported number of errors. There was, however, a significant main effect of stance position on reported number of errors ($F[3,5]=262.831$; P<0.0001; Figure 1B). Post hoc analyses revealed that the reported number of errors in all error-inducing stance positions (single-leg, firm; tandem, firm; single-leg, foam; and tandem, foam) were significantly different from one another (P<0.05). Effect sizes ranged from d=0.53 (single-leg, firm versus tandem, foam) to d=3.32 (tandem, firm versus single-leg, foam).

DISCUSSION

The purpose of the present study was to determine whether participants could reliably use four error criteria to visually identify balance errors under $_pD$ conditions as measured using ICC values. General guidelines for interpreting ICC values indicate that values >0.75 have 'good' reliability, and those ≤0.75 have'poor to moderate' reliability (13). We hypothesized that both the inter- and intrarater reliability of participants visually scoring balance using the four error criteria would be lower when performing a concurrent auditory vigilance task, compared with $_uD$ conditions. Contrary to this hypothesis, differences between $_pD$ and $_uD$ scoring were not significantly different statistically, and were both in the 'good' reliability range (13). During

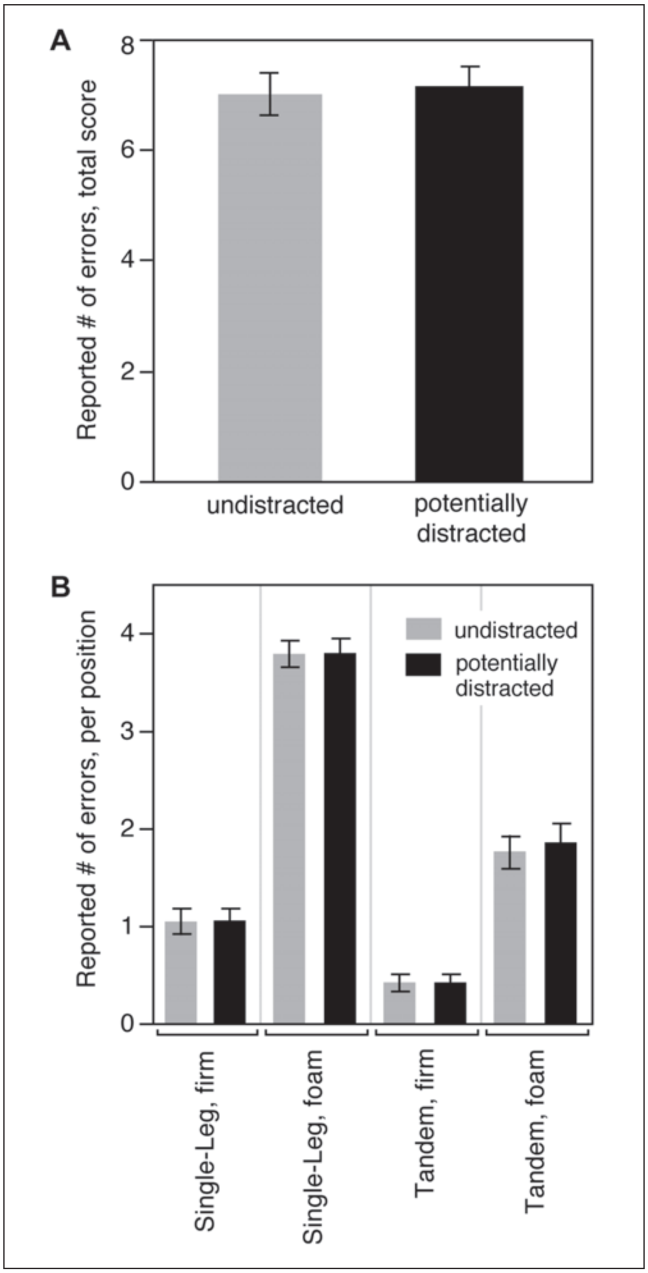

Figure 1) *Mean reported number of errors in undistracted and potentially distracted conditions for total score (***A***) and each stance position (***B***). Error bars indicate standard error*

$_p$D conditions, we also expected the number of errors the participants scored to be lower when compared with $_u$D conditions. The participants' reported number of balance errors, however, was not significantly different between the $_u$D and $_p$D conditions. Interestingly, in both conditions, the four error criteria (derived from the Balance Error Scoring System) yielded higher inter- and intrarater reliability compared with that of all six (7), which suggests that the remaining two error criteria not included in our error rating may potentially reduce inter- and intrarater reliability. Future studies are necessary, however, to determine the comparative reliabilities between standard and modified error criteria within this visual assessment.

Not only have numerous studies reported changes in postural sway (10,16), fall risk (17) and postural responses to perturbations (18) under $_p$D conditions, but balance performance on a similar assessment has also been shown to decline in situations when the patient is susceptible to distraction (6). If a distraction is sufficient to disrupt a patient, the clinician may also be disrupted. We chose, therefore, to analyze the effects of potentially distracting elements on individuals scoring a visual-based balance assessment using four error criteria. Our findings indicated that participants recorded similar numbers of errors between $_u$D and $_p$D conditions, suggesting that the four error criteria can be used reliably in assessing balance errors regardless of whether a potential auditory distraction is present.

Although our results suggested reliable rating regardless of condition, participants in the present study scored prerecorded videos of healthy adult balancers free of musculoskeletal injury and concussion history. Future studies are needed to determine whether the reliability of this visual balance assessment is preserved when the patient is injured or concussed. We also acknowledge that clinician experience may play a role in reliability in visually assessing balance; the present study was underpowered to determine that at this time. To address this question in the future, studies could test the extent to which reliability of a particular assessment is dependent on how long clinicians have been certified, or how long/much they have used that assessment specifically.

CONCLUSION

Participants in the present study viewed prerecorded videos of balancers performing six different stance positions to determine reliability in visually assessing balance using four error criteria. The results were contrary to our hypotheses in that not only were the reliability values the same with and without the potentially distracting auditory task, but the number of errors between both conditions were not significantly different. Clinicians who visually assess balance may not be susceptible to auditory distraction; it is, however, important to note that the balance performance may be influenced by that same distraction.

ACKNOWLEDGEMENTS: The authors thank Dr Eadric Bressel and Dr Richard Gordin for their assistance with study design and statistical analyses. They also thank Dale Mildenberger MS LAT ATC and Dr Jonathan Finnoff for developing the error criteria used in this study.

FUNDING: This study was funded, in part, by the Department of Health, Physical Education, and Recreation at Utah State University.

REFERENCES

1. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010;46:239-48.
2. Riemann BL, Guskiewicz KM, Shields EW. Relationship between clinical and forceplate measures of postural stability. J Sport Rehabil 1999;8:71-82.
3. Ross LM, Register-Mihalik JK, Mihalik JP, et al. Effects of a single-task versus a dual-task paradigm on cognition and balance in healthy subjects. J Sport Rehabil 2011;20:296-310.
4. Teel EF, Register-Mihalik JK, Troy Blackburn J, Guskiewicz KM. Balance and cognitive performance during a dual-task: Preliminary implications for use in concussion assessment. J Sci Med Sport 2013;16:190-4.
5. Broglio SP, Tomporowski PD, Ferrara MS. Balance performance with a cognitive task: A dual-task testing paradigm. Med Sci Sports Exerc 2005;37:689-95.
6. Onate JA, Beck BC, Van Lunen BL. On-field testing environment and balance error scoring system performance during preseason screening of healthy collegiate baseball players. J Athl Train 2007;42:446-51.
7. Stuart EA, Rodenberg RE, Dotson ML, Young JA, MacDonald JP. Reliability measures of the balance error scoring system as administered by certified athletic trainers. Curr Res Concussion 2015;2:22-4.
8. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. Sports Health 2011;3:287-95.
9. Guskiewicz KM, Register-Mihalik J, McCrory P, et al. Evidence-based approach to revising the SCAT2: Introducing the SCAT3. Br J Sports Med 2013;47:289-93.

10. Woollacott M, Shumway-Cook A. Attention and the control of posture and gait: A review of an emerging area of research. Gait Posture 2002;16:1-14.
11. Valovich TC, Perrin DH, Gansneder BM. Repeat administration elicits a practice effect with the balance error scoring system but not with the standardized assessment of concussion in high school athletes. J Athl Train 2003;38:51-6.
12. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. Tutor Quant Methods Psychol 2012;8:23-34.
13. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice, 3rd edn. Upper Saddle River: Pearson Education, Inc, 2009.
14. Kramer CY. Extension of multiple range tests to group means with unequal numbers of replications. Biom 1956;12:307-12.
15. Stoline MR. The status of multiple comparisons – simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. Am Stat 1981;35:134-41.
16. Shumway-Cook A, Woollacott M, Kerns KA, Baldwin M. The effects of two types of cognitive tasks on postural stability in older adults with and without a history of falls. J Gerontol Ser A Biol Sci Med Sci 1997;52:M232-40.
17. Toulotte C, Thevenon A, Watelain E, Fabre C. Identification of healthy elderly fallers and non-fallers by gait analysis under dual-task conditions. Clin Rehabil 2006;20:269-76.
18. Brown LA, Shumway-Cook A, Woollacott MH. Attentional demands and postural recovery: The effects of aging. J Gerontol Ser A Biol Sci Med Sci 1999;54:M165-71.