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

Jacob R Pierce MS LAT ATC CSCS1,2, Adam C Raikes MS LAT ATC1,2, Sydney Y Schaefer PhD1,3

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, with when they did not (undistracted [UD]). We also expected that the number of balance errors reported by the participants would be lower under UD compared with D conditions because of the inability to attend to two tasks simultaneously.

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

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” (single-leg, firm).

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.
(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 pub-
lished 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 descrip-
tion 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 partici-
pants 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 cor-
rect 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 continu-
ously 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 _D and _D conditions.

Because clinicians’ ability to visually assess balance in real-world
settings is likely more relevant for diagnostic purposes than their abil-
ity 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 _D and _D 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 _D condi-
tions (see reference 10 for operational definition).

Establishing inter- and intrarater reliability under _D conditions
To establish how reliable the participants were in visually assessing bal-
ance through counting errors, participants first scored 32 video sets in
random order without the concurrent auditory vigilance task (undis-
tracted _D; time 1). After each individual video, participants were
instructed to write down on their scoring sheet the total number of bal-
cance errors they observed, according to the four criteria in Table 1. These
data were used to determine the inter-rater reliability of these error cri-
teria 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 _D conditions. Also
during time 2, the participants scored the other 16 videos under _D
conditions. The order in which the _D and _D 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 reli-
bility of these four error criteria without potential distraction; intra-
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
calculate reliability measures and to compare the participants’ use of
the four error criteria.

To measure the inter- and intrarater reliability in both _D and _D
conditions, two-way, random absolute agreement intraclass correlation
coefficients (ICC2,1) (12,13) were used. Inter-rater reliability of the _D
scoring was computed across participants using the same 16 video sets
and re-scored at times 1 and 2, respectively. Similarly, intrarater reli-
bility of the _D scoring was computed within each participant
using the same 16 video sets between times 1 and 2. More importantly,
intrarater reliability of the _D 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 _D raters’ scoring was
computed within participants using the 16 video sets from time 1 (_D)
that were the same videos from time 2 (_D). 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 _D compared with _D conditions, the mean total scores across
raters was compared using a one-way repeated measures ANOVA with

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Balance ‘error’ criteria used by raters</strong></td>
</tr>
<tr>
<td>Lifts hands off of iliac crests</td>
</tr>
<tr>
<td>Opens eyes</td>
</tr>
<tr>
<td>Steps/stumbles/falls</td>
</tr>
<tr>
<td>Takes &gt;5 s to return to the testing position</td>
</tr>
</tbody>
</table>
**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 as well as in the tandem, firm stance position at time 2. The total score inter-rater ICC value was 0.93 during both the time 1 and time 2 conditions. Time 1 versus time 2 demonstrated reduced intrarater ICC values during single-leg, firm; single-leg, foam; and tandem, foam stance positions when compared with time 1 versus time 2. Conversely, time 1 versus time 2 resulted in a higher inter-rater reliability ICC value during the single-leg, foam stance position when compared with time 1 versus time 2.

---

**DISCUSSION**

The purpose of the present study was to determine whether participants could reliably use four error criteria to visually identify balance errors under potential distraction 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 performance in the ‘good’ reliability range (13). During
pD conditions, we also expected the number of errors the participants scored to be lower when compared with uD conditions. The participants’ reported number of balance errors, however, was not significantly different between the uD and pD 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 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 of errors between both conditions was not significantly different. 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.

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