ARTICLE

Berg Balance Scale: inter-rater and intra-rater reliability of the Spanish version with incomplete spinal cord injured subjects

Orestes Freixes¹ · Diego Alejandro Passuni¹ · Eliana Buffetti¹ · Milagros Elizalde¹ · Francisco Lastiri²

Received: 8 November 2019 / Revised: 8 April 2020 / Accepted: 9 April 2020
© The Author(s), under exclusive licence to International Spinal Cord Society 2020

Abstract

Study design Cross sectional.

Objectives To determine the inter-rater and intra-rater reliability of the Spanish version of the Berg Balance Scale with incomplete spinal cord injured subjects.

Setting CINER Rehabilitation Center.

Methods We administered and video recorded the Spanish version of the Berg Balance Scale to 20 incomplete SCI patients. Two raters scored the videos on two different occasions at least three weeks apart. We used intraclass correlation coefficient (ICC) and a confidence interval (CI) of 95% to evaluate the inter-rater and intra-rater (test–retest) reliability of the BBS total scores.

Results ICC values for inter-rater reliability at first and second observation were 0.99 (95% CI 0.97–1.00) and 0.99 (95% CI 0.99–1.00), respectively. Intra-rater ICC for rater 1 was 1.00 (95% CI 1.00–1.00) and for rater 2 was 1.00 (95% CI 0.99–1.00). All of them were excellent.

Conclusions The results indicate that the Spanish version of the Berg Balance Scale is a reliable tool to evaluate spinal cord injured patients' balance.

Introduction

It is considered that in Europe and the United States, 10–83 persons out of every million sustain a spinal cord injury (SCI) every year and that approximately half of these injuries are incomplete (sensory and/or motor function is partially spared below the level of the lesion) [1]. The majority of people classified within the first week of injury as motor incomplete will eventually become ambulatory by the time of rehabilitation discharge [2]. In fact, walking is a highly desired aim for persons who have a SCI [3]. This

walking capacity can be influenced by many variables, including the environment [4] and the correct functioning and integration of many subsystems (somatosensory information and lower limb strength) [5].

As a consequence of the aforementioned sensorimotor impairment provoked by the lesion, balance capacities can be altered, leading to a higher risk of falls when walking. Falls are common for individuals with SCI and fall rates have been estimated to be higher than in healthy adults [6]. Persons with SCI perceived reduced strength in lower limbs, balance loss, and environmental barriers as contributing factors to falls [7].

It is essential to use an evaluation tool that can be helpful in assessing and predicting risk of falls in this population in order to develop a customized rehabilitation program to address the problem of balance loss.

The Berg Balance Scale (BBS) is a widely used clinical assessment tool [8]. It was initially developed to measure balance in the elderly [9]. It is a 14-item test that requires an individual to perform static and dynamic everyday tasks of diverse difficulty. Each of the 14 items is scored from 0 to 4 and they are added to create a total score between 0 and 56; a higher score indicates better balance [10].

Supplementary information The online version of this article (<https://doi.org/10.1038/s41394-020-0278-5>) contains supplementary material, which is available to authorized users.

✉ Orestes Freixes
orestesfreixes@gmail.com

¹ Physical Therapy Unit, CINER Rehabilitation Institute, Buenos Aires, Argentina

² Computer Analyst Academic Affiliation, CINER Rehabilitation Institute, Buenos Aires, Argentina

Since its original development, it has been used to measure balance in a wide variety of patients. It has been validated for patients with stroke [11], brain injury [12], Parkinson's disease [13], SCI [14], and multiple sclerosis [15], and its English version has shown high inter-rater and intra-rater reliability [16]. However, as far as we are aware, no studies have measured the inter-rater and intra-rater reliability of the Spanish version of the BBS in SCI patients.

The aim of our study is to establish the inter-rater and intra-rater reliability of the Spanish version of the BBS in incomplete SCI subjects.

Methods

Sample size

Twenty subjects were studied. We calculated the required sample size to identify a desired intraclass correlation coefficient (ICC) of $\rho_1 = 0.90$, with a lower CI of $\rho_0 = 0.70$, given $\alpha = 0.05$ and a power $1 - \beta = 0.80$. This gave a sample size of at least 19 participants, according to Walter et al.'s [17] method of optimal design for reliability studies.

Recruitment

We included 20 incomplete SCI patients receiving outpatient rehabilitation therapy at CINER rehabilitation center at the time of the study. Inclusion criteria were having an incomplete SCI lesion ASIA C or D in accordance with the American Spinal Injury Association classification [18], time since injury ≥ 12 months, age ≥ 18 years, being at least community ambulator and speaking Spanish as primary language. We did not include patients who were unable to give informed consent owing to cognitive impairment. The study was approved by the ethics committee of UAI (Universidad Abierta Interamericana) University. All the participants gave written consent to take part in the study after being given written and verbal information about its purpose.

Translation of the Berg Balance Scale into Spanish

The English Berg Balance Scale was translated into Spanish by a native Spanish-speaking physical therapist with advanced English language skills (Certificate of Proficiency in English. Council of Europe level C2. Cambridge English Language Assessment). Back translation into English was carried out by an official translator. The translation into Spanish only suffered minor modifications and the final Spanish version was created by consensus between the two bilingual professionals.

Procedures

We administered the Spanish version of the Berg Balance Scale (supplementary Material) to 20 incomplete SCI patients. All the tests were administered by the same test administrator. The test administrator was a physical therapist specialized in the treatment of spinal cord injuries and experienced in the use of the BBS. He was in charge of reading the instructions to each subject before each individual task and ensuring the safety of patients. When scoring of an item depended in part on time, the test administrator was also in charge of pressing down the button of a stopwatch to start and finish the task. The action of depressing the stopwatch button was done in such a way that it could be clearly seen by the raters on the video (the administrator partially raised the arm and said “start” or “finish” aloud to the patient simultaneously). All the 20 complete assessments were videotaped with a camera that remained in a static position throughout the process and that stood 3.3 meters away from the patients, so that both the test administrator's and the patient's entire bodies could be filmed.

Participants completed all the assessments at CINER Rehabilitation Center in one measurement session.

The 14 tasks were evaluated in the following order:

1. Sitting unsupported
2. Sitting to standing
3. Standing unsupported for 2 minutes without holding on
4. Standing with eyes closed for 10 seconds
5. Standing unsupported with feet together
6. Reaching forward with outstretched arm while standing
7. Retrieving object from floor from a standing position
8. Turning to look behind over left and right shoulders while standing
9. Turning 360 degrees
10. Placing alternate foot on stool while standing unsupported
11. Standing with one foot extended
12. Standing on one foot
13. Standing to sitting
14. Transfers

Each participant was video recorded only once. Patients were allowed to perform each task only once and they were not permitted to use any walking aid during the tasks. Equipment used included two 18-inch high chairs (only one chair was used in tasks “sitting unsupported”, “sitting to standing” and “standing to sitting” and two chairs were used in task “transfers”), a measuring tape (task “reaching

forward with outstretched arm while standing”), a pencil (task “retrieving object from floor from a standing position”), a 9-inch high step stool (task “placing alternate foot on stool”) and a stopwatch (tasks “sitting unsupported”, “standing unsupported for 2 minutes without holding on”, “standing with eyes closed for 10 seconds”, “standing with feet together for 10 seconds”, “turning 360 degrees”, “placing alternate foot on a stool”, “standing with on foot”, and “standing with one foot extended). In case patients needed to rest between the tasks, they were allowed to sit on one of the chairs available for the study.

Scoring

After completing all the BBS assessments, videotapes were numbered, titled, and saved in a personal computer. Two raters (R1 and R2), who were not physically present during the video recording process, scored the videos on two different occasions with at least a 3-week interval, time 1 (T1) and time 2 (T2). Raters were not allowed to compare the results between T1 and T2 scores. Like the test administrator, the raters were physiotherapists experienced in the use of the BBS. They were allowed to re-play the videos as many times as necessary to improve their judgment accuracy but not to discuss their scores or to have knowledge of each other's rating. Depending on the patient's performance, each of the 14 items was scored from 0 to 4 (0 = lowest performance, 4 = highest performance). Progressively more points were subtracted if distance or time requirements were not met, if the subject needed supervision or if he/she had to receive external assistance to prevent a fall. Scores were then added to create a total score between 0 and 56.

If a participant stated being incapable of performing a task this was video-recorded and 0 points were awarded by the rater after watching the video.

Statistical analysis

Data analysis was performed using R software (version 3.5.1). Numerical variables were summarized by means and standard deviations (SD), whereas categorical variables were summarized by frequencies and percentages.

We used ICC using ICC (2,1) model according to Shrout and Fleiss classification [19] and a confidence interval (CI) of 95% to evaluate inter-rater and intra-rater (test–retest) reliability of the BBS total scores.

The following classification was used for the ICC values: poor reliability ICC < 0.50, moderate reliability ICC between 0.50 and 0.75, good reliability between > 0.75 and 0.90, and excellent reliability ICC > 0.90.

Results

Demographic characteristics

Twenty incomplete (1 ASIA C and 19 ASIA D) SCI subjects participated in this study. Mean age was 48.2 (16.6). Fourteen were male and six were female. Characteristics are listed in Table 1.

Distribution of scores

The mean total score $\pm$ SD of the BBS was similar between the raters (Table 2).

Reliability

ICC values for inter-rater reliability at first and second observation were 0.99 (95% CI 0.97–1.00) and 0.99 (95% CI 0.99–1.00), respectively. Both were excellent (Table 3).

Table 1 Demographic characteristics.

	Total (<i>n</i> = 20)
<i>Sex</i>	
Women	6 (30%)
Men	14 (70%)
<i>Age</i>	
Mean (SD)	48.2 (16.6)
Median [min, max]	52.5 [22,75]
<i>SCI level</i>	
C2	2 (10%)
C4	2 (10%)
C5	1 (5%)
C6	2 (10%)
C7	1 (5%)
L2	2 (10%)
L3	3 (15%)
L4	1 (5%)
T1	1 (5%)
T10	1 (5%)
T12	1 (5%)
T6	2 (10%)
T7	1 (5%)
<i>ASIS</i>	
C	1 (5%)
D	19 (95%)

ASIS ASIA impairment scale, *C* motor and sensitive incomplete, *D* motor and sensitive incomplete, *C* cervical, *T* thoracics, *L* lumbar, *SCI level* level of spinal cord injury.

Table 2 Descriptive analysis.

	(<i>n</i> = 20)
<i>BBS-rater #1 (T1)</i>	
Mean (SD)	43.7 (14.8)
Median [Min, Max]	50.5 [9.00, 56.0]
<i>BBS-rater #1 (T2)</i>	
Mean (SD)	43.8 (14.8)
Median [Min, Max]	50.5 [9.00, 56.0]
<i>BBS-rater #2 (T1)</i>	
Mean (SD)	43.1 (15.0)
Median [Min, Max]	50.0 [12.0, 56.0]
<i>BBS-rater #2 (T2)</i>	
Mean (SD)	43.6 (14.5)
Median [Min, Max]	50.0 [12.0, 56.0]

Table 3 ICC for inter-rater and intra-rater reliability.

Inter-rater reliability			
Time	ICC	Reliability classification	IC 95%
T1	0.99	Excellent	0.97–1.00
T2	0.99	Excellent	0.99–1.00
Intra-rater reliability			
Rater	ICC	Reliability classification	IC 95%
R1	1.00	Excellent	1.00–1.00
R2	1.00	Excellent	0.99–1.00

ICC Intraclass correlation coefficient, IC 95% confidence interval 95% of ICC. Intra-rater ICC for rater 1 was 1.00 (95% CI 1.00–1.00) and for rater 2 was 1.00 (95% CI 0.99–1.00). Both were excellent (Table 3).

Discussion

In this study, inter-rater and intra-rater reliabilities were both excellent. Our results coincide with those reported by Witz [20] for the Berg Balance Scale in the original English version, who found its inter-rater reliability to be excellent as well. Our analysis of individual tasks scoring (Table 4) showed moderate to excellent levels of agreement between raters (inter-rater reliability) with the exception of task “transfers” (poor). This poor reliability in this last-mentioned task might be explained by the fact that in this item there are two scoring options that include the possibility of using the hands to assist in the transfer (able to transfer with minor use of hands vs able to transfer safely

Table 4 ICC for individual tasks scoring reliability.

Task	Inter-rater		Intra-rater	
	T1	T2	R1	R2
1	–	–	–	–
2	0.99	0.97	1.00	0.99
3	0.64	0.64	1.00	1.00
4	0.92	0.91	1.00	0.97
5	0.83	0.83	1.00	1.00
6	0.86	0.86	1.00	1.00
7	0.94	0.94	1.00	1.00
8	0.89	0.88	0.98	1.00
9	0.97	0.97	0.99	0.99
10	0.96	0.96	1.00	1.00
11	0.85	0.85	1.00	1.00
12	0.89	0.86	1.00	0.99
13	0.74	0.76	1.00	0.98
14	0.29	0.37	1.00	0.82

(All scores = 4)

ICC intraclass correlation coefficient, (–) not calculated because of lack of variability.

with definite use of hands). This may have led to disagreement between raters on how to score the task when patients helped themselves with their hands to perform the transfer.

There were tasks raters found it simple to score. This is the case of task “sitting unsupported” in which both the inter-rater and intra-rater scores were exactly the same because all the patients were considered to be able to sit safely (all scores = 4). The rationale for this might be that all our 20 study subjects were at least community ambulators. Trunk control is directly related to the ability to walk [21]. This means that if they were able to walk it is because they already had enough trunk control to sit without assistance. In other words, when including in our sample only patients who could walk (with or without assistive device), we were indirectly including patients who could sit safely.

With respect to the intra-rater reliability, rater 1 demonstrated an excellent (ICC 1.00) level of agreement between both observational times (T1 and T2) in all tasks, whereas rater 2 demonstrated an excellent (ICC 1.00) concordance percentage in all tasks except for item “transfers” (good reliability). We chose to separate the two observational moments (T1 and T2) with a time lapse of at least 3 weeks to minimize the likelihood of raters remembering the first scoring (T1). Moreover, bearing in mind that each rater watched a total of 280 videotapes, we consider that the results cannot be explained by a remembering process.

Tasks were timed by using a stopwatch. Pressing a stopwatch is not a purely objective action (for example, the test administrator can mistakenly finish a task one or two

seconds before or after the required time has elapsed). However, we consider that the margin of error was very small and that this question did not affect any of the results obtained.

Study limitations

The present study had certain limitations. The reliability values obtained may be higher than those physical therapists would probably obtain in real clinical practice. The reason likely is twofold: the sample selection and the rating process. Sample selection only included community ambulatory patients. The homogeneity of the studied population (only high-performing patients) may have influenced the creation of such high-reliability estimates. More realistic results may have been expected if the sample had included patients with a broader range of abilities. Second, the rating process was artificial. The use of video tape methods offers the advantage of a high degree of reproducibility when measuring observations [22]. In our study, raters could re-watch the tape as many times as they wanted. In real clinical practice, the test administrator is also the rater and has to judge the performance of a patient immediately without the possibility of repeating the task to make sure of the score, and is prone to missing specific details of a task (for example, a patient momentarily opening the eyes when required to keep them closed). This unusual scoring process certainly resulted in higher values than physiotherapists would probably obtain in daily practice. Further research is needed to determine the inter-rater and intra-rater reliability of this version with a more varied sample and with a different scoring method.

Conclusion

The results indicate that the Spanish version of the Berg Balance Scale is a reliable tool to evaluate spinal cord injured patients' balance.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Wyndaele M, Wyndaele J-J. Incidence, prevalence and epidemiology of spinal cord injury: what learns a worldwide literature survey? *Spinal Cord* 2006;44:523–529.
- Burns SP, Golding DG, Rolle WA Jr, Graziani V, Ditunno JF Jr. Recovery of ambulation in motor-incomplete tetraplegia. *Arch Phys Med Rehabil* 1997;78:1169–1172.
- Ditunno PL, Patrick M, Stineman M, Ditunno JF. Who wants to walk? Preferences for recovery after SCI: a longitudinal and cross-sectional study. *Spinal Cord* 2008;46:500–506.
- Olmos LE, Freixes O, Gatti MA, Cozzo DA, Fernandez SA, Vila CJ, et al. Comparison of gait performance on different environmental settings for patients with chronic spinal cord injury. *Spinal Cord* 2017;46:331–334.
- Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? *Age Ageing* 2006;35:ii7–ii11.
- Jørgensen V, Butler Forslund E, Franzén E, Opheim A, Seiger Å, Ståhle A, et al. Factors associated with recurrent falls in individuals with traumatic spinal cord injury: a multicenter study. *Arch Phys Med Rehabil* 2016;97:1908–1916.
- Brotherton SS, Krause JS, Nietert PJ. A pilot study of factors associated with falls in individuals with incomplete spinal cord injury. *J Spinal Cord Med* 2007;30:243–250.
- La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale an internally valid and reliable measure of balance across different etiologies in neurorehabilitation? A revisited Rasch analysis study. *Arch Phys Med Rehabil* 2012;93:1209–1216.
- Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health* 1992;83:S7–11.
- Park SH, Lee YS. The diagnostic accuracy of the Berg Balance scale in predicting falls. *West J Nurs Res* 2017;39:1502–1525.
- Wee JY, Wong H, Palepu A. Validation of the Berg Balance Scale as a predictor of length of stay and discharge destination in stroke rehabilitation. *Arch Phys Med Rehabil* 2003;84:731–735.
- Newstead AH, Hinman MR, Tomberlin JA. Reliability of the Berg Balance Scale and balance master limits of stability tests for individuals with brain injury. *J Neurol Phys Ther* 2005;29:18–23.
- Qutubuddin AA, Pegg PO, Cifu DX, Brown R, McNamee S, Carne W. Validating the Berg Balance Scale for patients with Parkinson's disease: a key to rehabilitation evaluation. *Arch Phys Med Rehabil* 2005;86:789–792.
- Lemay JF, Nadeau S. Standing balance assessment in ASIA D paraplegic and tetraplegic participants: concurrent validity of the Berg Balance scale. *Spinal Cord* 2010;48:245–250.
- Cattaneo D, Regola A, Meotti M. Validity of six balance disorders scales in persons with multiple sclerosis. *Disabil Rehabil* 2006;30:28:789–795.
- Downs S. The Berg Balance Scale. *J Physiother* 2015;61:46.
- Walter SD, Elliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101–110.
- Maynard FM Jr, et al. International standards for neurological and functional classification of spinal cord injury american spinal injury association. *Spinal Cord*. 1997;35:266–274.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
- Wirz M, Müller R, Bastiaenen C. Falls in persons with spinal cord injury: validity and reliability of the Berg Balance Scale. *Neurorehabil Neural Repair* 2010;24:70–77.
- Quinzanos-Fresnedo Jimena, Fratini-Escobar PaolaC, Almaguer Benavides KievkaM, Aguirre-Güemez AnaValeria, Barrera-Ortiz Aída, Pérez-Zavala Ramiro, et al. Prognostic validity of a clinical trunk control test for independence and walking in individuals with spinal cord injury. *J Spinal Cord Med*. 2018;12:1–8.
- Haidet KK, Tate J, Divirgilio-Thomas D, Kolanowski A, Happ MB. Methods to improve reliability of video-recorded behavioral data. *Res Nurs Health* 2009;32:465–474.