J. Phys. Ther. Sci.
24: 301–305, 2012

Reliability of the Pediatric Balance Scale in the Assessment of the Children with Cerebral Palsy

Jin-Gang Her, PhD[1)], Ji-Hea Woo, MPE[1)], Jooyeon Ko, PT, PhD[2)]

[1)] *Department of Physical Therapy, Hallym College*
[2)] *Department of Rehabilitation Medicine, CHA University, CHA Bundang Medical Center: 351 Yatap-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-712 South Korea.*
E-mail: 7806218@hanmail.net

Abstract. [Purpose] The purpose of this study was to determine the absolute reliability and relative reliability of the Pediatric Balance Scale (PBS) for children with Cerebral Palsy (CP). [Subjects] For the inter- and intra-rater reliability, 36 children with CP (20 boys, 16 girls) were recruited. For the test-retest reliability, 27 children with CP (16 boys, 11 girls) were recruited. [Methods] Seven pediatric physical therapists rated 36 video recordings of children with CP to test inter-rater reliability. Two therapists rescored each video recording to test intra-rater reliability after an interval of two weeks. To evaluate the test-retest reliability, one rater scored each of the 27 video clips on two occasions, two-weeks apart. Relative reliability was calculated using the intraclass correlation coefficient (ICC). Absolute reliability was assessed using the standard error of measurement (SEM) and the smallest real difference (SRD). [Results] The PBS showed high to very high relative reliability, and the absolute reliability was satisfactory for the inter- and intra-rater and test-retest reliability. [Conclusion] Excellent absolute reliability as well as relative reliability of the PBS was obtained, so the PBS is reliable for examining the functional balance of children with cerebral palsy.
Key words: PBS, Reliability, Cerebral palsy

(This article was submitted Oct. 7, 2011, and was accepted Nov. 9, 2011)

INTRODUCTION

Cerebral palsy (CP) is a group of disorders that affect the development of posture and movement and is caused by an irreversible insult to the developing central nervous system[1)]. Motor disorder results in loss of functional balance. As functional balance has been defined as the elements of gross motor abilities, poor balance causes difficulties with functional tasks involved in activities of daily living[2)]. Preschoolers and school-aged children with CP classified as level I of the Gross Motor Function Classification Measure (GMFCS) are able to ambulate independently without assistive devices within their homes, schools, and communities with a limited movement repertoire, but it is difficult for them to walk on uneven surfaces and walk in crowds or confined spaces. Physical therapy often includes balance training in their therapy sessions to improve their gross motor ability[3, 4)]. Therefore, a reliable and simple functional balance measure is needed to evaluate the effects of physical therapy interventions[5)].

Balance measures for children with CP have been examined through the observations of the underlying elements of the balance responses, standardized developmental measures of gross motor function, and forceplates[6, 7)]. The standardized assessment tools for children with CP include the Pediatric Evaluation of Disability Inventory[8)], the Gross Motor Function Measure[9)], and the Pediatric Reach Test[10)]. These tests, however, do not detect small changes in functional balance. Forceplates can be used only in controlled laboratory environments and are, therefore, unsuitable for use in general clinical settings[11)]. A review of balance in the literature suggested that the Pediatric Balance Scale (PBS) might be useful for assessing the functional balance of preschoolers and school-aged children with CP. The reliability of the PBS has been established[12)]. The test-retest reliability is extremely high [($ICC_{3,\ 1}$=0.998)], and inter-rater reliability is also satisfactory [($ICC_{3,\ 1}$=0.9987]. The intraclass correlation coefficient (ICC), however, doesn't detect small but clinically significant changes over time. In a reliability study, the use of the ICC and standard error of measurement (SEM) together is recommended, because ICC is a relative measure of reliability and reflects the measurement's ability to differentiate between participants, while SEM is an absolute measure, and quantifies the precise scores[13)]. Thus, our aim was to establish the absolute reliability as well as the relative reliability of the PBS in assessment of children with CP classified by the GMFCS as levels I through III.

SUBJECTS AND METHODS

The PBS includes 14 items of increasing difficulty. These items are used to assess functional skills relevant to everyday tasks, such as: "sitting to standing," "standing to

sitting," "transfers," "standing unsupported," "sitting unsupported," "standing with eyes closed," "standing with feet together," "standing with one foot in front," "standing on one foot," "turning 360 degrees," "turning to look behind," "retrieving an object from the floor," "placing alternate feet on a stool," "reaching forward with an outstretched arm." As in Franjoine's study[12], item 14, "reaching forward with an outstretched arm," was omitted from the videotape analysis because a two-dimensional videotape does not adequately record the test performance. Each item was scored on the criterion-based 0 to 4 point scale. A maximum total score of 52 could be obtained for the thirteen items measured in this study. Only one practice trial per item was allowed. Verbal, visual and physical cues were provided to ensure the child understood the requested task. If a child successfully completed the task (i.e., scored a four on the first trial), additional trials were not administered.

The subjects were a convenience sample of 36 children diagnosed with CP (20 boys and 16 girls), between 6 and 13 years of age, from two local CP clinics in South Korea. All participants were scheduled to participate in two separate test sessions with an interval of two weeks in order to determine the test-retest reliability, but only 27 of the 36 children with CP were available for the second test due to health problems or personal reasons at the time of the second test session. The inclusion criteria were: ability to follow verbal commands; absence of nerve block injection or orthopedic surgery within the previous 6 months, such as Botox injection or muscle-lengthening surgery; absence of cardiovascular disease and joint contractures; and ability to stand independently without upper extremity support for four seconds. Consent to participate in this study was obtained from the parents of all the children, and approval for this study was granted by the Hallym University Institutional Review Board.

In the present study, we translated the English version of the PBS into Korean with the permission of the original author. The translation procedures followed a forward-backward-forward method. In an effort to minimize the measurement error, the PBS assessment was administered in a pediatric physical therapy room which was comfortable and familiar to the children. In addition, for the second video-recording session, the day of the week and time of day were kept as consistent as possible. All children were assessed barefoot and without assistive devices. If the child was tired or not feeling well, the assessment was postponed. Scheduling of the test session was at the convenience of the child and their parents. All of the tests were carried out by one of the authors (JY), a pediatric physical therapist with 11 years of experience in the treatment and evaluation of children with CP; and the procedure was recorded by another pediatric physical therapist who did not participate in either scoring session. Video recordings were made of each item of the PBS (in the order they are listed above), and the process took about 15 minutes per session per child. Seven pediatric physical therapists (raters A, B, C, D, E, F and G) participated in the reliability scoring study. The raters had a mean of 5.5 years of pediatric experience, varying from 11 months (Rater G) to 9 years (Rater A). The seven raters and the one assistant who recorded the performance attended a one-day PBS training workshop on video recording, administration and scoring. The workshop was administered by the author (JY).

The reliability scoring session was conducted in three parts after all video-recording had been completed. In Part 1, to assess inter-rater reliability, the seven raters scored the same first video recordings of the 36 subjects in the therapy room. In Part 2, to determine intra-rater reliability, two of the seven raters rescored the same first 36 video recordings two weeks after the first assessment. In Part 3, to examine the test-retest reliability, rater C scored 27 video recordings in two scoring sessions, two-weeks apart. This assessment interval was used to minimize the influence of memory[14]. Each scoring session was held over two days without discussion of the scores among the raters. The GMFCS level, hand function according to the Manual Ability Classification System (MACS), tone distribution and CP type were evaluated by the author (JY) at the time of the first video recording. The definition of the GMFCS levels I − III and the MACS levels I − III[15] used in this study are as follows: GMFCS I, walking without restrictions but limitations in more advanced gross motor skills; GMFCS II, walking without restriction but with limitations in walking outdoors and in the community; GMFCS III, walking with assistive devices but with limitations in walking outdoors and in the community; MACS I, handling objects easily and successfully; MACS II, handling most objects but with somewhat reduced quality and/or speed of achievement; and MACS III, handling objects with difficulty and need helping to prepare and/or modify activities[15]. Other information about general characteristics such as epilepsy, age, weight and height were obtained through interviews with the children's parent.

Scores attained in the PBS are described using means and standard deviations. For relative reliability, the ICC with 95% confidence intervals was used to evaluate the inter-rater, intra-rater and test-retest reliability of the total scores of the PBS. The inter-rater and test-retest reliability of the total scores of the PBS were analyzed by using $ICCs_{3,1}$, a two-way mixed-effects model with absolute agreement, and 95% confidence intervals (CI). To assess intra-rater reliability, the $ICC_{1,1}$ was calculated, and thus, a one-way random effects model for a single measure was used[16]. For absolute reliability, the SEM and the smallest real difference (SRD) were calculated. SEM is described as the square root of the within participant variance: SEM = SD $\sqrt{}$ 1-r. For inter-rater reliability, a SRD value of less than 1.96 × SEM is expected to admit the results of 95% of the ratings[17] In the intra-rater reliability and test-retest reliability, the SRD should be less than $\sqrt{2} \times 1.96 \times SEM = 2.77 \times SEM$ for 95% of the pairs of ratings to be significant[18]. The statistics were performed using SPSS ver. 12.0.1 software (SPSS Inc., Chicago, IL, USA).

RESULTS

Characteristics of the subjects and descriptive statistics are shown in Tables 1 and 2. In the 7-rater trial (A1-B1-C1-D1-E1-F1-G1), high inter-rater reliability was demonstrated

Table 1. Summary of sample characteristics

Variables	Inter- and intra-rater reliability (N=36)	Test-retest reliability (n=27)
Gender		
Boy	20 (55.56)	16 (59.26)
Girl	16 (44.44)	11 (40.74)
GMFCS level		
I	23 (63.89)	18 (66.67)
II	11 (30.56)	8 (29.63)
III	2 (5.55)	1 (3.70)
Epilepsy		
Yes	5 (13.89)	4 (14.81)
No	31 (86.11)	23 (85.19)
Hand function		
I	24 (66.67)	18 (66.67)
II	9 (25.00)	6 (22.22)
III	3 (8.33)	3 (11.11)
Tone distribution		
Hemiplegia	8 (22.22)	6 (22.22)
Diplegia	27 (75.00)	21 (77.78)
Quadriplegia	1 (2.78)	0 (0.00)
CP type		
Spastic	35 (97.22)	27 (100.00)
Dyskinetic	1 (2.78)	0 (0.00)
Age (years)	8.94 ± 2.40	8.93 ± 2.32
Weight (kg)	30.17 ± 10.18	30.15 ± 10.46
Height (cm)	129.69 ± 12.33	129.11 ± 12.24

Values are n (%) or Mean (SD); GMFCS: Gross motor function classification system; CP: Cerebral palsy.

by an $ICC_{3,1}$ value of 0.901 for the total score of the PBS (Table 3). The SEM and SRD were 0.65 and 1.27, respectively (Table 3).

In the two-rater trial with an interval of two weeks, the intra-rater (A1-A2, G1-G2) reliability $ICC_{1,1}$ of the total score of the PBS was 0.988 for the experienced rater (rater A) and 0.978 for the newly trained rater (rater G) (Table 4). The SEM and SRD for the total score of the most experienced rater were 0.37 and 1.02, respectively (Table 4). For the newly trained rater, the SEM and SRD were 0.43 and 1.19, respectively (Table 4).

The test-retest (C1-C2) reliability of total score for the PBS with an interval of two weeks was assessed using $ICC_{3,1.}$, and the value was 0.958 (Table 5). The SEM and SRD were 0.61 and 1.69 for the total score, respectively (Table 5).

Table 2. Descriptive statistics of the first test for the total score of the PBS (N=36)

Rater	Mean ± SD	Min - Max
A	46.06 ± 4.60	31 - 52
B	46.64 ± 5.91	24 - 52
C	45.94 ± 4.46	34 - 51
D	45.00 ± 5.35	32 - 52
E	45.08 ± 5.22	33 - 52
F	45.22 ± 5.23	32 - 52
G	47.92 ± 4.53	36 - 52

PBS: Pediatric balance scale.

Table 3. Inter-rater reliability for the total score of the PBS: ICC, SEM, SRD for CP (N=36)

$ICC_{3,1}$	95%CI	SEM	SRD
0.901	0.849 - 0.941	0.65	1.27

ICC: Intraclass correlation coefficient; CI: Confidence interval; SEM: Standard error of measurement; SEM%: Standard error of measurement%; SRD: Smallest real difference.

Table 4. Intra-rater reliability for the total score of the PBS: ICC, SEM, SRD for CP (N=36)

Rater	1st of first test Mean ± SD	Min - Max	2nd of first test Mean ± SD	Min - Max	$ICC_{1,1}$	95%CI	SEM	SRD
A	46.06 ± 4.60	31 - 52	46.56 ± 4.57	32 - 52	0.988	0.976 - 0.994	0.37	1.02
G	47.92 ± 4.53	36 - 52	45.36 ± 5.21	33 - 52	0.978	0.957 - 0.988	0.43	1.19

PBS: Pediatric balance scale; ICC: Intraclass correlation coefficient; CI: Confidence interval; SEM: standard error of measurement; SRD: Smallest real difference.

Table 5. Test-retest reliability for total score of the PBS: ICC, SEM, SRD for CP (n=27)

Rater	First test Mean ± SD	Min-Max	Second test Mean ± SD	Min-Max	$ICC_{3,1}$	95% CI	SEM	SRD
C	46.41 ± 4.26	34 - 51	46.81 ± 4.36	35 - 52	0.958	0.910 -0.981	0.61	1.69

PBS: Pediatric balance scale; ICC: Intraclass correlation coefficient; CI: Confidence interval; SEM: Standard error of measurement; SRD: Smallest real difference

DISCUSSIONS

The objective of this study was to investigate the absolute reliability and the relative reliability of the PBS in the assessment of children with CP in order to provide therapists with useful clinical values for detecting real changes before and after interventions. The results of this study are the first to confirm not only good relative reliability but also acceptable absolute reliability for the total score of the PBS in the assessment of children with CP.

Reliability refers to the dependability, consistency, and stability of an assessment tool. Several types of reliability exist, such as inter- and inter-rater, and test-retest reliability. Inter-rater reliability estimates how consistent the test is when used by different raters; whereas intra-rater reliability is the consistency with which one rater assigns scores to a single set of responses on two or more occasions. If a tester is using videotape analysis to examine intra-rater reliability, he or she can view the same videotape on two different days. The test-retest reliability demonstrates the consistency of an assessment tool between one test occasion and another.[14]. These three kinds of reliability are necessary to establish a reliable assessment tool. Reliability is quantified as relative or absolute. Relative reliability investigates the agreement between a group of raters and is represented by the ICC. Absolute reliability examines variability in scores in repeated measurements and is represented by the SEM and the SRD[19].

In the present study, the ICC scores of the inter- and intra-rater reliability, and the test-retest reliability for the total score of the PBS were 0.901, 0.988 and 0.978, and 0.958, respectively (Tables 3, 4, and 5). A study by Franjoine et al.[12], examined the inter-rater and test-retest reliability of the PBS using the ICC. The value for inter-rater $ICC_{3,1}$ was 0.997 and the $ICC_{3,1}$ of the test-retest reliability was 0.998. Harley and Fragala-Pinkham[20] stated that values of 0.90–1.00 represent very high correlation, values of 0.70–0.89 represent high correlation, values of 0.50–0.69 indicate moderate correlation, values of 0.26–0.49 demonstrate low correlation and values of 0.00–0.25 represent little correlation. Our results were consistent with a previous study[12] (Tables 3, 4, and 5). For the intra-rater reliability, both raters showed very high reliability regardless of the extent of their clinical experience.

However, ICC alone does not provide clinical guidance, for it cannot assess real differences[21]. Bland and Altman[18] suggested that a more appropriate way of evaluating the reliability of an assessment tool for clinical use is to examine absolute reliability. To our knowledge, this is the first study of children with CP that has examined the absolute reliability, represented by the reliability indices of the SEM and the SRD, of the PBS. Liaw et al.[22] suggested that SEM scores of less than 10% of the total mean score are acceptable, and Smidt et al.[23] suggested that SRD scores less than 10% of the total range are acceptable. We found that the SEMs for the inter- and intra-rater, and test-retest reliability of the total score of the PBS were small, indicating that the error band of the observed scores was limited. The values of the SRD, which is based on the SEM, were also reasonable and were less than 10% of the total range. Because the SEM is inversely related to the ICC reliability, a relatively high ICC may result in a low SEM value; therefore, it is important to document both. The value of the SEM for assessment tools is useful for interpreting whether a change in scores is beyond measurement error in clinical setting[24]. A value higher than the SEM has been suggested[13] for determining whether a change is real. In the intra-rater reliability (Rater A), the raw score of the total score of the PBS for children with CP should change by more than 0.37 (Table 4) for the change to indicate a real change with a 95% confidence level. On the other hand, values below 0.37 cannot be interpreted as a real improvement because such a change may occur due to measurement error.

This study had some limitations. The sample was one of convenience, and we included only children with CP treated in a CP clinic. Further studies are required to examine the validity of the PBS by comparing it with other assessment tools that are frequently used in pediatric settings. In addition, other psychometric properties such as responsiveness and sensitivity to change need to be addressed with larger numbers of subjects including children with CP who attend physiotherapy outside the CP clinic.

In conclusion, we have demonstrated satisfactory absolute and relative inter- and intra-rater reliabilities of the PBS and provided a reference framework for future studies using SEM values to assess the functional balance of children with CP.

REFERENCES

1) Bax M, Goldstein M, Rosenbaum P, et al.: Proposed definition and classification of cerebral palsy, April 2005. Dev Med Child Neurol, 2005, 47: 571–576. [Medline] [CrossRef]
2) Gan SM, Tung LC, Tang YH, et al. : Psychometric properties of functional balance assessment in children with cerebral palsy. Neurorehabil Neural Repair, 2008, 22: 745–753.
3) Shumway-Cook A, Hutchinson S, Kartin D, et al.: Effect of balance training on recovery in children with cerebral palsy. Dev Med Child Neurol, 2003, 45: 591–602. [Medline] [CrossRef]
4) Liao HF, Hwang AW: Relations of balance function and gross motor ability for children with cerebral palsy. Percept Mot Skills, 2003, 96: 1173–1184. [Medline]
5) Lexell JE, Downham DY: How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil, 2005, 84: 719–723. [Medline] [CrossRef]
6) Pountney TE, Mulcahy C, Green E: Early development of postural control. Phys Can, 1990, 76: 799–802.
7) Hsue BJ, Miler F, Su FC: The dynamic balance of the children with cerebral palsy and typical developing during gait. Part I: spatial relationship between COM and COP trajectories. Gait Posture, 2009, 29: 465–470. [Medline] [CrossRef]
8) Vargus-Adams JN, Martin LK, Maignan SH, et al.: The GMFM, PEDI, and CP-QOL and perspectives on functioning from children with CP, parents, and medical professionals. J Pediatr Rehabil Med, 2011, 4: 3–12. [Medline]
9) Russell DJ, Rosenbaum PJ, Avery LM, et al.: Gross motor function measure (GMFM-66 & GMFM-88). User's manual. London: Mac Keith Press, 2002.
10) Bartlett D, Birmingham T: Validity and reliability of a pediatric reach test. Pediatr Phys Ther, 2003, 15: 84–92. [Medline] [CrossRef]
11) Gan SM, Tung LC, Tang YH, et al.: Psychometric properties of functional balance assessment in children with cerebral palsy. Neurorehabil Neural Repair, 2008, 22: 745–753. [Medline] [CrossRef]
12) Franjoine MR, Gunther JS, Taylor MJ: Pediatric balance scale: a modified version of the berg balance scale for the school-age child with mild to moderate motor impairment. Pediatr Phys Ther, 2003, 15: 114–128. [Medline] [CrossRef]
13) Polit DF, Hungler BP: Essential of nursing research: methods, appraisal,

and utilization. Philadelphia: Lippincott Williams & Wilkins; 1989.
14) Haga N, van der Heijden-Maessen HC, van Hoorn JF, et al.: Test-retest and inter- and intra-rater reliability of the quality of the upper-extremity skills test in preschool-age children with cerebral palsy. Arch Phys Med Rehabil, 2007, 88: 1686–1689. [Medline] [CrossRef]
15) Carnahan KD, Arner M, Hägglund G: Association between gross motor function (GMFCS) and manual ability (MACS) in children with cerebral palsy. A population-based study of 359 children. BMC Musculoskelet Disord, 2007, 8: 50. [Medline] [CrossRef]
16) Damstra J, Huddleston Slater JJ, Fourie Z, et al.: Reliability and the smallest detectable differences of lateral cephalometric measurements. Am J Orthod Dentofacial Orthop, 2010, 138: 546.e1-546e8.
17) Crosbie J, Kilbreath SL, Dylke E, et al.: Effects of mastectomy on shoulder and spinal kinematics during bilateral upper-limb movement. Phys Ther, 2010, 90: 679–692. [Medline] [CrossRef]
18) Bland JM, Altman DG: Measurement error and correlation coefficients. BMJ, 1996, 313: 41–42. [Medline] [CrossRef]
19) Conradsson M, Lundin-Olsson L, Lindelöf N, et al.: Berg balance scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities. Phys Ther, 2007, 87: 1155–1163. [Medline] [CrossRef]
20) Haley SM, Fragala-Pinkham MA: Interpreting change scores of tests and measures used in physical therapy. Phys Ther, 2006, 86: 735–743. [Medline]
21) Flansbjer UB, Holmbäck AM, Downham D, et al.: What change in isokinetic knee muscle strength can be detected in men and women with hemiparesis after stroke? Clin Rehabil, 2005, 19: 514–522. [Medline] [CrossRef]
22) Liaw LJ, Hsieh CL, Lo SK, et al.: The relative and absolute reliability of two balance performance measures in chronic stroke patients. Disabil Rehabil, 2008, 30: 656–661. [Medline] [CrossRef]
23) Smidt N, van der Windt DA, Assendelft WJ, et al.: Intraobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. Arch Phys Med Rehabil, 2002, 83: 1145–1150. [Medline] [CrossRef]
24) Wang HH, Liao HF, Hsieh CL: Reliability, sensitivity to change, and responsiveness of the peabody developmental motor scales–second edition for children with cerebral palsy. Phys Ther, 2006, 86: 1351–1359. [Medline] [CrossRef]