Test-retest reliability and minimal detectable change of the Balance Evaluation Systems Test and its two abbreviated versions in persons with mild to moderate spinocerebellar ataxia: A pilot study

Yuki Kondo[a,*], Kyota Bando[a], Yosuke Ariake[a], Wakana Katsuta[a], Kyoko Todoroki[a], Daisuke Nishida[a,b], Katsuhiro Mizuno[a,b] and Yuji Takahashi[c]

[a]*Department of Physical Rehabilitation, National Center Hospital, National Center of Neurology and Psychiatry, Tokyo, Japan*
[b]*Department of Rehabilitation Medicine, Keio University School of Medicine, Tokyo, Japan*
[c]*Department of Neurology, National Center Hospital, National Center of Neurology and Psychiatry, Tokyo, Japan*

Abstract.
BACKGROUND: The reliability of the evaluation of the Balance Evaluation Systems Test (BESTest) and its two abbreviated versions are confirmed for balance characteristics and reliability. However, they are not utilized in cases of spinocerebellar ataxia (SCA).
OBJECTIVE: We aimed to examine the test-retest reliability and minimal detectable change (MDC) of the BESTest and its abbreviated versions in persons with mild to moderate spinocerebellar ataxia.
METHODS: The BESTest was performed in 20 persons with SCA at baseline and one month later. The scores of the abbreviated version of the BESTest were determined from the BESTest scores. The interclass correlation coefficient (1,1) was used as a measure of relative reliability. Furthermore, we calculated the MDC in the BESTest and its abbreviated versions.
RESULTS: The intraclass correlation coefficients (1,1) and MDC at 95% confidence intervals were 0.92, 8.7(8.1%), 0.91, 4.1(14.5%), and 0.81, 5.2(21.6%) for the Balance, Mini-Balance, and Brief-Balance Evaluation Systems Tests, respectively.
CONCLUSIONS: The BESTest and its abbreviated versions had high test-retest reliability. The MDC values of the BESTest could enable clinicians and researchers to interpret changes in the balance of patients with SCA more precisely.

Keywords: Minimal detectable change, neurological rehabilitation, postural balance, rare diseases, reproducibility of results, Spinocerebellar ataxias

1. Introduction

Persons with spinocerebellar ataxia (SCA) typically present with progressive disease owing to the degeneration of neurons in the cerebellum, brainstem, and spinal cord (Ishikawa et al., 1999; Sakai et al., 2010; Yoshida et al., 2009). Warrenburg et al.

*Address for correspondence: Yuki Kondo, 4-1-1 Ogawa-higashi-cho, Kodaira-shi, Tokyo 187-8551, Japan. Tel.: +81 042 341 2711; Fax: +81 042 346 1681; E-mail: kondo.ncnp@gmail.com.

(Warrenburg et al., 2005) reported that falls occur very frequently in persons with SCA and that these falls can lead to injuries or a fear of falling. Furthermore, persons with SCA and resultant balance impairment or limb movement disorder are usually confined to a wheelchair or are bedridden within 10–20 years of onset (Ilg et al., 2010). In contrast, daily training in disease management can lead to a temporary improvement in balance and walking ability (Ilg et al., 2009; Ilg et al., 2010; Kondo et al., 2018; Miyai et al., 2011). Balance training in SCA is therefore extremely important and should be performed regularly from an early stage. Resultant function maintenance and symptom-progression delay benefits associated with activities of daily living and quality of life.

The Scale for the Assessment and Rating of Ataxia (SARA) is a tool for assessing ataxia. Ilg et al. (Ilg et al., 2009; Ilg et al., 2010) and Miyai et al. (Miyai et al., 2011) reported that balance training intervention for 4 weeks improved SARA scores in participants. The daily assessment of ataxic persons with the use of the SARA is widely used and reflects the effect of ataxia on balance; however, the SARA has limited sensitivity to change (Jacobi et al., 2011; Trouillas et al., 1997). The SARA is sufficient to assess mild to severe disease evenly. However, we adopted the Balance Evaluation Systems Test (BESTest) as an assessment that could sensitively detect changes at an early stage (i.e. mild to moderate), especially for balance disorders.

The BESTest is a relatively new multitask balance assessment tool developed to identify specific postural control problems (i.e., biomechanical constraints, stability limits, anticipatory postural adjustments, postural responses, sensory orientation, dynamic balance during gait, and cognitive effects) (Franchignoni et al., 2010; Horak et al., 2009). However, the 36-item BESTest takes 30 minutes to complete, which may be too long to use in daily clinical settings where time constraints are often a major concern. Thus, an abbreviated version of the test, the Mini-Balance Evaluation Systems Test (Mini-BESTest), has been developed. This test takes only 10 minutes to complete (Godi et al., 2012). Furthermore, the Brief-Balance Evaluation Systems Test (Brief-BESTest) (Padgett et al., 2012) was designed to assess six different aspects of postural control in standing and walking. The balance disorder caused by ataxia is a neurological sign characterized by a lack of the voluntary coordination of muscle movement. As a result, ataxia caused by SCA complicates dynamic balance

control and leads to balance and gait changes. It would be a beneficial outcome measure for intervention to use the BESTest that identify specific postural control problems. Hence, the BESTest and its abbreviated versions allow clinicians to determine the type of balance problems in order to design treatments specific for persons with SCA. However, there are limited reports on these tests in the evaluation of balance characteristics in persons with SCA.

It is important to increase the efficacy of interventions by clarifying the minimal detectable change (MDC), which can provide clinicians with useful and easy-to-understand criteria to assess change (improvement or decline) in individual performance. The MDC is the minimum amount of change in a measure that must be obtained in order to determine whether a true change has occurred between two testing occasions. The MDC is expressed as a confidence interval around the standard error of measurement (SEM), indicating values within the variability (error) range that can be attributed to the testing instrument. Therefore, we consider that it is important to calculate the MDC of the BESTest and its two abbreviated versions in persons with SCA.

The aim of this study was to examine the test-retest reliability and MDC of the BESTest and its abbreviated versions in persons with SCA.

2. Materials and methods

2.1. Participants

Persons with SCA, whose diagnoses had been established through genetic analysis, were recruited between November 2014 and May 2018. They had a gait score suggestive of SCA and a SARA score less than or equal to three points (i.e. they were capable of walking without an assistive device). In total, 20 persons with SCA (13 males and 7 females) were enrolled in the study. Informed consent was obtained from all patients. The study setting was the National Center of Neurology and Psychiatry (NCNP) in Japan. Testing was approved by the Institutional Review Board of NCNP, Japan (approval number A2016-064) in accordance with the Ethics Committee of NCNP and the Declaration of Helsinki.

2.2. Instruments

The BESTest consists of 36 items scored on an ordinal scale from 0 to 3, with "0" indicating the

lowest level of function and "3" the highest level of function. The total possible raw score is 108 points. The Mini-BESTest is an abbreviated form of the BESTest with only 14 items. It was developed with the aid of factor analysis and Rasch analysis (Franchignoni et al., 2010). The Mini-BESTest is scored on an ordinal scale from "0" to "2," and thus, the total possible raw score is 28 points. The Brief-BESTest is a six-item revised version of the BESTest, designed to improve the clinical utility and to preserve the construct validity of the BESTest (Padgett et al., 2012). The Brief-BESTest is scored on an ordinal scale from "0" to "3," and thus, the total possible raw score is 24 points (two items include a right and left component). In all cases, higher points indicate better balance and function.

2.3. Procedure

A tester (physical therapist) was trained by watching the BESTest training video provided by the developer. The BESTest score was obtained at baseline (test session 1) and the BESTest at one month (test session 2) was performed by the same evaluator for each person. Moreover, rehabilitation intervention and changes in drug therapy administration did not occur during this period. The scores of the Mini-BESTest and Brief-BESTest were extracted from the BESTest scores.

2.4. Statistical analysis

Test-retest reliability for test sessions 1 and 2 of the BESTest individual subsystem scores and total scores, Mini-BESTest, and Brief-BESTest total score were calculated using Intraclass Correlation Coefficients (ICC) (1,1) with 95% confidence intervals (CIs). For interpretation of the ICC (1,1) score, ranging from 0.00 to 1.00, values greater than 0.70 are considered to have good reliability (Koo & Li, 2016). The ICC were analyzed accordingly, using R version 2.18.1 (Ihaka & Gentleman, 1996).

Furthermore, we calculated the MDC at the 95% confidence interval (MDC_{95}) in the BESTest, Mini-BESTest, and Brief-BESTest total scores after the confirmation of the systematic error using the Bland-Altman analysis. For the Bland-Altman analysis, the differences between the two measured values and the mean of the two values were plotted on the y and x axes, respectively, to prepare a Bland-Altman plot (Bland & Altman, 2010). A systematic error represents a deviation in a specific direction, and is divided

into fixed error and proportional bias. When no systematic error is detected, only an accidental error is considered to reduce measurement reliability. An accidental error is divided into biological individual differences and error produced on measurement. To investigate this measurement error, MDC was determined. The MDC is the minimum amount of change in a given measure that must be obtained in order to determine whether a true change has occurred between two testing occasions. This minimum change is considered as a random error and is calculated at a certain level of confidence (usually 95%) (Jette et al., 2007; Lexell & Downham, 2005). Thus, the MDC_{95} can be used as a threshold to identify statistically significant individual changes (Jette et al., 2007; Lu et al., 2008).

We calculated the standard error of measurement (SEM) based on the test-retest reliability. The SEM is considered to be indicative of the range of scores that are expected on retesting and was calculated as follows: (Weir, 2005)

$$SEM = SD \times \sqrt{(1 - ICC)}$$

The calculated SEM was then used to determine the MDC_{95} as follows:

$$MDC_{95} = 1.96 \times SEM \times \sqrt{2}$$

3. Results

This study included 20 participants with SCA (13 males and 7 females), with a mean age of 63.7 years (SD = 10.1). The characteristics of the participants are shown in Table 1.

The Bland-Altman analysis revealed no systematic error for the total scores of the BESTest, Mini-BESTest, and Brief-BESTest (Fig. 1). The 95%CI of

Table 1
Characteristics of the Participants ($n = 20$)

Sex	Males: 13, Females: 7
Diagnosis (numbers)	MJD/ SCA3:4, SCA6:9, SCA31:7
Age (years)	63.7(10.1)
Age at onset (years)	53.9(10.5)
SARA (points)	9.9(3.5)

The data are presented as numbers, mean (standard deviations). MJD, Machado-Joseph Disease; SCA, Spinocerebellar ataxia; SARA, Scale for the Assessment and Rating of Ataxia; BESTest, Balance Evaluation Systems Test; Mini-BESTest, Mini-Balance Evaluation Systems Test; Brief-BESTest, Brief-Balance Evaluation Systems Test.

a) BESTest

b) Mini-BESTest
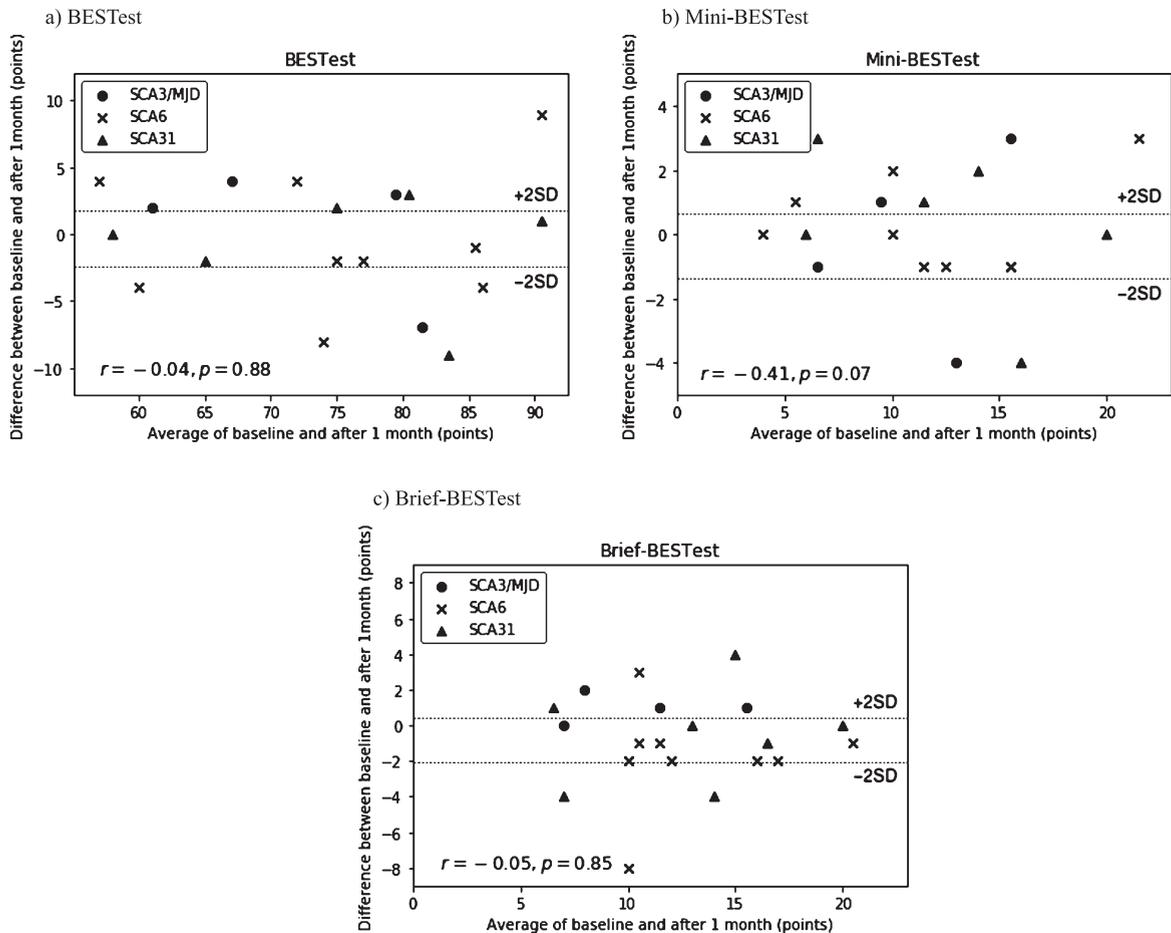
c) Brief-BESTest

Fig. 1. Bland-Altman plots showing the intrarater reliability of the (a)BESTest, (b)Mini-BESTest, and (c)Brief-BESTest. The two lines define the limits of agreement 95%CI of the mean difference. BESTest, Balance Evaluation Systems Test; Mini-BESTest, Mini-Balance Evaluation Systems Test; Brief-BESTest, Brief-Balance Evaluation Systems Test.

the mean difference of the BESTest, Mini-BESTest, and Brief-BESTest total score were 2.04 to –2.34 points, 0.66 to –1.36 points, and 0.45 to –2.05 points, respectively. These results did not indicate any fixed error. In addition, the BESTest, Mini-BESTest, and Brief-BESTest total scores did not show a statistically significant correlation in the difference between the two measured values and the mean of the two values. These results suggested that proportional bias was not present.

The BESTest total and individual subsystem scores, Mini-BESTest, and Brief-BESTest total scores and the ICC value for the interrater reliability, MDC$_{95}$ are represented in Table 2. ICC and MDC$_{95}$ for the BESTest individual subsystem scores ranged from 0.70 to 0.96, from 2.0 (13.3%) to 6.9 (38.3%) points, respectively. The ICC value for the interrater reliability of the BESTest total score was

0.92 (95% CI, 0.82–0.97), the Mini-BESTest total score was 0.91 (95% CI, 0.79–0.96), and the Brief-BESTest total score was 0.81 (95% CI, 0.58–0.92). The SEM for the total scores of BESTest, Mini-BESTest, and Brief-BESTest are 3.2 (2.9%), 1.5 (5.2%), and 1.9 (7.8%) points, respectively. The MDC for the total scores of BESTest, Mini-BESTest, and Brief-BESTest are 8.7 (8.1%), 4.1 (14.5%), and 5.2 (21.6%) points, respectively.

4. Discussion

The BESTest total and individual subsystem scores, Mini-BESTest, and Brief-BESTest total scores demonstrate a generally high test-retest reliability in persons with mild to moderate SCA. Additionally, total scores of less than or equal to

Table 2
ICC and MDC$_{95}$ for the BESTest and its abbreviated versions

Variable, (%)	Test session 1	Test session 2	ICC (1, 1)	SEM		MDC$_{95}$	
				Converted Score, %	Points	Converted Score, %	Points
BESTest	69.7(10.4)	70.0(10.5)	0.92	2.9	3.2	8.1	8.7
I Biomechanical constraints	90.3(9.0)	92.0(7.4)	0.70	5.3	0.8	15.3	2.3
II Stability limits/ verticality	88.8(7.1)	87.4(8.8)	0.77	3.8	0.8	10.5	2.2
III Anticipatory postural adjustment	64.2(12.0)	65.3(14.1)	0.85	5.0	0.9	14.4	2.6
IV Postural response	50.6(27.2)	50.0(24.4)	0.72	13.9	2.5	38.3	6.9
V Sensory orientation	78.0(13.5)	79.3(15.4)	0.90	4.7	0.7	13.3	2.0
VI Stability gait	51.2(24.1)	51.4(23.9)	0.96	4.8	1.0	13.3	2.8
Mini-BESTest	42.3(16.1)	43.6(19.2)	0.91	5.2	1.5	14.5	4.1
Brief-BESTest	50.8(18.4)	54.2(17.9)	0.81	7.8	1.9	21.6	5.2

The data are presented as numbers, mean (standard deviations). SEM, Standard error of the measurement; MDC, Minimal detectable change; BESTest, Balance Evaluation Systems Test; Mini-BESTest, Mini-Balance Evaluation Systems Test; Brief-BESTest, Brief-Balance Evaluation Systems Test.

8.7 (8.1%), 4.1 (14.5%), and 5.2 (21.6%) points were considered acceptable errors of measurement considering the calculated MDC for the BESTest, Mini-BESTest, and Brief-BESTest, respectively.

4.1. The Reliability of the BESTest, Mini-BESTest, and Brief-BESTest

This study assessed the reliability of the BESTest and abbreviated versions in capturing true changes in functional balance in persons with SCA. These results confirmed that the BESTest and two abbreviated versions had high test-retest reliability in persons with mild to moderate SCA (ICC (1,1) values for BESTest, Mini-BESTest, and Brief-BESTest were 0.92, 0.91, and 0.81 respectively). These findings are in agreement with those of previous studies on persons with Parkinson's disease, multiple sclerosis, and other conditions that resulted in balance impairments (Leddy et al., 2011; Mitchell et al., 2018; Padgett et al., 2012; Potter et al., 2018). In addition, the ICC for the individual subsystem scores ranged from 0.70 to 0.96; an ICC greater than 0.7 was considered good (Eliasziw et al., 1994; McGraw & Wong, 1996; Rankin & Stokes, 1998). In this study, the participants with mild to moderate SCA had high test-retest reliability for the BESTest and abbreviated versions, despite the one month duration between measures.

Our results indicated that the MDC$_{95}$ of the BESTest total score was 8.1% (8.7 points) and the Mini-BESTest total score was 14.5% (4.1 points), which were similar to the findings of previous studies (Jacome et al., 2016; Mitchell et al., 2018; Potter et al., 2018; Tsang et al., 2013; Wang-Hsu & Smith, 2018). The MDC for the BESTest has been reported for community-dwelling older adults (Wang-Hsu &

Smith, 2018) and persons with multiple sclerosis (Mitchell et al., 2018; Potter et al., 2018). The MDC for the Mini-BESTest has been reported in persons with vestibular disorders (Godi et al., 2012), stroke (Tsang et al., 2013), and chronic obstructive pulmonary disease (Jacome et al., 2016). In addition, our results indicated that the MDC$_{95}$ of the Brief-BESTest total score was 21.6% (5.2 points), which was similar to the results obtained by Jacome et al. who reported an MDC$_{95}$ of 26.9% (4.9 points) (Jacome et al., 2016).

It can be seen from the BESTest that the MDC$_{95}$ of Section IV (postural responses) was 38.3% higher than those of other sections. This result was in accordance with those reported by Tsang et al. (Tsang et al., 2013) and Löfgren et al. (Löfgren et al., 2014), where items associated with postural responses showed the lowest agreement. For this reason, the reactive postural responses need to be consistent in order to determine how far to allow participants to lean before suddenly releasing the support. This task appears likely to be more challenging than other tasks such as the task assessing the time a person is able to stand on a foam surface with his or her eyes closed (Section V). Accordingly, it is assumed that manipulation of the BESTest Section IV is difficult, and this had a larger variation tendency compared with other sections. In SCA, a lack of voluntary coordination of muscle movement caused by ataxia affects the postural responses. Therefore, it is considered that the BESTest Section IV showed variations in some participants because the progression of ataxia impairs postural responses.

These established MDC values are predicted to be useful for clinicians to determine whether an intervention has induced a real improvement in balance

function. Future studies in this field should aim to improve the accuracy of the interpretation of the results of the BESTest and its abbreviated versions.

4.2. The efficacy of the abbreviated versions of BESTest in persons with SCA

The results of the Mini-BESTest tend to show a lower value in persons with SCA than in persons with Parkinson's disease (Duncan et al., 2013; Wallen et al., 2016). In this study, the mean value for the total score of the Mini-BESTest and Brief-BESTest in persons with mild to moderate SCA was 11.8 points (42.3%) and 12.2 points (50.8%), respectively. The mean value for the total score of the Mini-BESTest in 112 persons with mild to moderate Parkinson's disease was 19.2 points (68.6%) (Wallen et al., 2016). The mean value for the total score of the Brief-BESTest in 80 persons with Parkinson's disease was 13.2 points (55%) (Duncan et al., 2013). The structural validity of the Mini-BESTest has been investigated in persons with Parkinson's disease using factor and Rasch analyses (Wallen et al., 2016). Similarly, the Brief-BESTest selected one item from each of the BESTest system subsections based on the highest item correlation coefficients with their respective system section in persons with and without a neurological diagnosis (e.g. Parkinson's disease, multiple sclerosis) (Padgett et al., 2012). The Mini-BESTest is a unidimensional scale that focuses on assessing "dynamic balance." As persons with SCA scored the lowest in Postural response and gait Stability, the Mini-BESTest is difficult to perform in these patients and, therefore, may not be suitable as an evaluation index. For this reason, currently, we consider that the BESTest, out of the three versions of the scale, is the best choice for the assessment of balance abilities in persons with SCA. It is necessary to create a specialized evaluation index for the BESTest using both factor and Rasch analyses of persons with SCA.

This study has a few limitations. First, the sample size was small. COSMIN guidelines recommend a sample of 100 patients as adequate or a sample of 50 as the minimum (Prinsen et al., 2018). Although the sample size was small, the participants represented a range of functional performance, with the total BESTest scores ranging from 50.9 (47.1%) to 88.0 (81.5%) points. Also, considering that SCA is a rare neurologic disease, it is an important finding that determined the test-retest reliability and MDC of the BESTest and its abbreviated versions in persons with mild-to-moderate SCA whose diagnoses had been established through genetic analysis. Second, the participants were ambulatory and therefore were classified as having mild to moderate SCA. Further studies in more persons with SCA are needed to determine whether our findings can be generalized to persons with severe SCA. Third, the imitation of the Mini-BESTest, the Brief-BESTest was not performed separately due to time constraints. Future studies are necessary to evaluate the Mini-BESTest and Brief-BESTest separately in order to comprehensively validate the Mini-BESTest, Brief-BESTest results. Finally, in our study, a distribution-based method to calculate MDC was used. Future studies should use other methods (e.g. effect size or an anchor-based method using a global rating of change) to compute responsiveness.

In conclusion, the BESTest and its abbreviated versions demonstrated high test-retest reliability in persons with mild to moderate disability as a result of SCA. More than 8.7 (8.1%), 4.1 (14.5%), and 5.2 (21.6%) points in the BESTest, Mini-BESTest, and Brief-BESTest total scores, respectively, can be judged as significant changes. The development of ICC and MDC values for the BESTest in persons with SCA will contribute to the advancement of analysis and treatment of SCA, and these results have good applicability in clinical settings. Further research into the use of these testing systems in a variety of SCA persons, with different levels of manifestations of the disease, is warranted.

Acknowledgments

Conflict of interest

No potential conflicts of interest are reported by the authors.

References

Bland, J. M., & Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*, *47*(8), 931-936.

Duncan, R. P., Leddy, A. L., Cavanaugh, J. T., Dibble, L. E., Ellis, T. D., Ford, M. P., Foreman, K. B., & Earhart, G. M. (2013). Comparative Utility of the BESTest, Mini-BESTest, and Brief-BESTest for Predicting Falls in Individuals With Parkinson Disease: A Cohort Study. *Physical Therapy*, *93*(4), 542-550.

Eliasziw, M., Young, S. L., Woodbury, M. G., & Fryday-Field, K. (1994). Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example. *Physical Therapy*, *74*(8), 777-788.

Franchignoni, F., Horak, F., Godi, M., Nardone, A., & Giordano, A. (2010). Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. *Journal of Rehabilitation Medicine*, *42*(4), 323-331.

Godi, M., Franchignoni, F., Caligari, M., Giordano, A., Turcato, A. M., & Nardone, A. (2012). Comparison of Reliability, Validity, and Responsiveness of the Mini-BESTest and Berg Balance Scale in Patients With Balance Disorders. *Physical Therapy*, *93*(2), 158-167.

Horak, F. B., Wrisley, D. M., & Frank, J. (2009). The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Physical Therapy*, *89*(5), 484-498.

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299.

Ilg, W, Synofzik, M., Brotz, D., Burkard, S., Giese, M. A., & Schols, L. (2009). Intensive coordinative training improves motor performance in degenerative cerebellar disease. *Neurology*, *73*(22), 1823-1830.

Ilg, W, Synofzik, M., Brotz, D., Burkard, S., Giese, M. A., Schöls, L., & Synofzik, M. (2010). Long-term effects of coordinative training in degenerative cerebellar disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *25*(13), 2239-2246.

Ishikawa, K., Watanabe, M., Yoshizawa, K., Fujita, T., Iwamoto, H., Yoshizawa, T., Harada, K., Nakamagoe, K., Komatsuzaki, Y., Satoh, A., Doi, M., Ogata, T., Kanazawa, I., Shoji, S., & Mizusawa, H. (1999). Clinical, neuropathological, and molecular study in two families with spinocerebellar ataxia type 6 (SCA6). *Journal of Neurology, Neurosurgery & Psychiatry*, *67*(1), 86-89.

Jacobi, H., Bauer, P., Giunti, P., Labrum, R., Sweeney, M. G., Charles, P., Dürr, A., Marelli, C., Globas, C., Linnemann, C., Schöls, L., Rakowicz, M., Rola, R., Zdzienicka, E., Schmitz-Hübsch, T., Fancellu, R., Mariotti, C., Tomasello, C., Baliko, L., & Klockgether, T. (2011). The natural history of spinocerebellar ataxia type 1, 2, 3, and 6: a 2-year follow-up study. *Neurology*, *77*(11), 1035-1041.

Jacome, C., Cruz, J., Oliveira, A., & Marques, A. (2016). Validity, Reliability, and Ability to Identify Fall Status of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-BESTest in Patients With COPD. *Physical Therapy*, *96*(11), 1807-1815.

Jette, A., Tao, W., Norweg, A., & Haley, S. (2007). Interpreting rehabilitation outcome measurements. *Journal of Rehabilitation Medicine*, *39*(8), 585-590.

Kondo, Y., Bando, K., Ariake, Y., Katsuta, W., Kobayashi, Y., Saotome, T., & Takahashi, Y. (2018). Effects of short-term intensive balance training in patients with ambulatory spinocerebellar degeneration -Balance Evaluation Systems Test (BESTest)–. *Neurological Therapeutics*, *5*(35), 628-632. (Japanese)

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Leddy, A. L., Crowner, B. E., & Earhart, G. M. (2011). Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. *Journal of Neurologic Physical Therapy: JNPT*, *35*(2), 90-97.

Lexell, J. E., & Downham, D. Y. (2005). How to Assess the Reliability of Measurements in Rehabilitation. *American Journal of Physical Medicine & Rehabilitation*, *84*(9), 719-723.

Löfgren, N., Lenholm, E., Conradsson, D., Ståhle, A., & Franzén, E. (2014). The Mini-BESTest–a clinically reproducible tool for balance evaluations in mild to moderate Parkinson's disease? *BMC Neurology*, *14*(1), 235.

Lu, W., Wang, C., Lin, J., Sheu, C., & Hsieh, C. (2008). The minimal detectable change of the simplified stroke rehabilitation assessment of movement measure. *Journal of Rehabilitation Medicine*, *40*(8), 615-619.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30-46.

Mitchell, K. D., Chen, H., & Silfies, S. P. (2018). Test-Retest Reliability, Validity, and Minimal Detectable Change of the Balance Evaluation Systems Test to Assess Balance in Persons with Multiple Sclerosis. *International Journal of MS Care*, *20*(5), 231-237.

Miyai, I., Ito, M., Hattori, N., Mihara, M., Hatakenaka, M., Yagura, H., Sobue, G., Nishizawa, M., & Collaboration, C. A. R. T. (2011). Cerebellar Ataxia Rehabilitation Trial in Degenerative Cerebellar Diseases. *Neurorehabilitation and Neural Repair*, *26*(5), 515-522.

Padgett, P. K., Jacobs, J. V., & Kasser, S. L. (2012). Is the BESTest at Its Best? A Suggested Brief Version Based on Interrater Reliability, Validity, Internal Consistency, and Theoretical Construct. *Physical Therapy*, *92*(9), 1197-1207.

Potter, K., Anderberg, L., Anderson, D., Bauer, B., Beste, M., Navrat, S., & Kohia, M. (2018). Reliability, validity, and responsiveness of the Balance Evaluation Systems Test (BESTest) in individuals with multiple sclerosis. *Physiotherapy*, *104*(1), 142-148.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., Vet, H. C. W. de, & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *27*(5), 1147-1157.

Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clinical Rehabilitation*, *12*(3), 187-199.

Sakai, H., Yoshida, K., Shimizu, Y., Morita, H., Ikeda, S., & Matsumoto, N. (2010). Analysis of an insertion mutation in a cohort of 94 patients with spinocerebellar ataxia type 31 from Nagano, Japan. *Neurogenetics*, *11*(4), 409-415.

Trouillas, P., Takayanagi, T., Hallett, M., Currier, R. D., Subramony, S. H., Wessel, K., Bryer, A., Diener, H. C., Massaquoi, S., Gomez, C. M., Coutinho, P., Hamida, M. B., Campanella, G., Filla, A., Schut, L., Timann, D., Honnorat, J., Nighoghossian, N., Manyam, B., & Neurology, T. A. N. C. of the W. F. of. (1997). International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. *Journal of the Neurological Sciences*, *145*(2), 205-211.

Tsang, C. S. L., Liao, L.-R., Chung, R. C. K., & Pang, M. Y. C. (2013). Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. *Physical Therapy*, *93*(8), 1102-1115.

Wallen, M. B., Sorjonen, K., Lofgren, N., & Franzen, E. (2016). Structural Validity of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in People With Mild to Moderate Parkinson Disease. *Physical Therapy*, *96*(11), 1799-1806.

Wang-Hsu, E., & Smith, S. S. (2018). Interrater and Test-Retest Reliability and Minimal Detectable Change of the Balance Evaluation Systems Test (BESTest) and Subsystems With Community-Dwelling Older Adults. *Journal of Geriatric Physical Therapy (2001)*, *41*(3), 173-179.

Warrenburg, B. P. C. van de, Steijns, J. A. G., Munneke, M., Kremer, B. P. H., & Bloem, B. R. (2005). Falls in degenerative cerebellar ataxias. *Movement Disorders*, *20*(4), 497-500.

Weir, J. P. (2005). Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *The Journal of Strength and Conditioning Research*, *19*(1), 231.

Yoshida, K., Shimizu, Y., Morita, H., Okano, T., Sakai, H., Ohata, T., Matsumoto, N., Nakamura, K., Tazawa, K., Ohara, S., Tabata, K., Inoue, A., Sato, S., Shimojima, Y., Hattori, T., Ushiyama, M., & Ikeda, S. (2009). Severity and progression rate of cerebellar ataxia in 16q-linked autosomal dominant cerebellar ataxia (16q-ADCA) in the endemic Nagano Area of Japan. *Cerebellum (London, England)*, *8*(1), 46-51.