

Improving the utility of the Brunnstrom recovery stages in patients with stroke

Validation and quantification

Chien-Yu Huang, PhD^a, Gong-Hong Lin, BS^b, Yi-Jing Huang, BS^b, Chen-Yi Song, PhD^c, Ya-Chen Lee, PhD^d, Mon-Jane How, MA^e, Yi-Miau Chen, MA^e, I-Ping Hsueh, MA^{b,e}, Mei-Hsiang Chen, PhD^{f,g,*}, Ching-Lin Hsieh, PhD^{b,e}

Abstract

The Brunnstrom recovery stages (the BRS) consists of 2 items assessing the poststroke motor function of the upper extremities and 1 assessing the lower extremities. The 3 items together represent overall motor function. Although the BRS efficiently assesses poststroke motor functions, a lack of rigorous examination of the psychometric properties restricts its utility. We aimed to examine the unidimensionality, Rasch reliability, and responsiveness of the BRS, and transform the raw sum scores of the BRS into Rasch logit scores once the 3 items fitted the assumptions of the Rasch model.

We retrieved medical records of the BRS (N=1180) from a medical center. We used Rasch analysis to examine the unidimensionality and Rasch reliability of both upper-extremity items and the 3 overall motor items of the BRS. In addition, to compare their responsiveness for patients (n=41) assessed with the BRS and the Stroke Rehabilitation Assessment of Movement (STREAM) on admission and at discharge, we calculated the effect size (ES) and standardized response mean (SRM).

The upper-extremity items and overall motor items fitted the assumptions of the Rasch model (infit/outfit mean square=0.57–1.40). The Rasch reliabilities of the upper-extremity items and overall motor items were high (0.91–0.92). The upper-extremity items and overall motor items had adequate responsiveness (ES=0.35–0.41, SRM=0.85–0.99), which was comparable to that of the STREAM (ES=0.43–0.44, SRM=1.00–1.13).

The results of our study support the unidimensionality, Rasch reliability, and responsiveness of the BRS. Moreover, the BRS can be transformed into an interval-level measure, which would be useful to quantify the extent of poststroke motor function, the changes of motor function, and the differences of motor functions in patients with stroke.

Abbreviations: BRS = Brunnstrom recovery stages, BRS-A = the arm item of the Brunnstrom recovery stages, BRS-H = the hand item of the Brunnstrom recovery stages, BRS-L = the leg item of the Brunnstrom recovery stages, DIF = differential item functioning, ES = effect size, MNSQ = mean square, PCA = principal component analysis, PCM = partial credit model, SRM = standardized response mean, STREAM = the stroke rehabilitation assessment of movement.

Keywords: brunnstrom recovery stages, quantification, stroke, validation

Editor: Weimin Guo.

The authors have no conflicts of interest to disclose.

^a Department of Occupational Therapy, I-Shou University, Kaohsiung, Taiwan,

^b School of Occupational Therapy, College of Medicine, National Taiwan University, ^c Department of Health Promotion and Gerontological Care, Taipei College of Maritime Technology, Taipei, ^d Department of Occupational Therapy, College of Medical and Health Science, Asia University, Taichung, ^e Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, ^f School of Occupational Therapy, Chung Shan Medical University, ^g Occupational Therapy Room, Chung Shan Medical University Hospital, Taichung, Taiwan.

* Correspondence: Mei-Hsiang Chen, School of Occupational Therapy, Chung Shan Medical University, No.110, Section 1, Jianguo N. Road., Taichung City 40201, Taiwan (e-mail: cmh@csmu.edu.tw).

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medicine (2016) 95:31(e4508)

Received: 24 November 2015 / Received in final form: 28 June 2016 /

Accepted: 13 July 2016

<http://dx.doi.org/10.1097/MD.0000000000004508>

1. Introduction

Motor recovery is one of the most important treatment goals for patients with stroke.^[1] A valid, reliable, and responsive measure assessing poststroke motor function is essential for appropriate clinical decision-making, treatment planning, and research (e.g., outcome studies). The Brunnstrom recovery stages (BRS) is a short and easily administered measure for assessing motor function.^[2] The BRS contains 3 items for the arm (BRS-A), the hand (BRS-H), and the leg (BRS-L), all of which are rated on a 6-level Likert-type scale. These items are usually used individually to describe the motor function (i.e., the arm, the hand, and the leg, respectively) of a patient. The BRS has good item-level psychometric properties.^[3–7] Therefore, the BRS may be useful in both clinical and research settings.

However, 3 weaknesses restrict the utility of the BRS. First, the unidimensionality of the BRS has not yet been investigated. It is unknown whether the 3 items of the BRS assess the same construct and whether the scores of the items can be summed to represent overall motor function.^[8] Sum scores of the BRS could quickly provide an overall impression of a patient's motor function as an alternative to inspecting the score of each item. Moreover, sum scores could be an outcome indicator because any progress made on each item by a patient could be detected, which is useful for monitoring a patient's overall change over time and

determining the effects of intervention. Therefore, validation of the unidimensionality of the BRS is warranted.

Second, it is unknown whether the BRS is as responsive as lengthier measures of motor function, such as the stroke rehabilitation assessment of movement (STREAM), which has moderate to large responsiveness.^[14,15] Theoretically, the STREAM would be more responsive than the BRS because the STREAM contains more items for detecting changes in motor function. However, previous studies have shown that the responsiveness of the short-form format of a measure can be comparable to that of the long-form format in a group of patients.^[9,10] Since the BRS has high feasibility, and if it is as responsive as the STREAM, the BRS would be appropriate as an outcome measure for a group of patients to decrease the evaluation burden on both patients and users.

Third, the BRS is rated on an ordinal scale rather than on an interval scale. An ordinal scale identifies the order of the values, but the differences between the values remain unknown.^[11] An interval scale not only identifies the order of the values but also has equal intervals between any 2 adjacent values.^[11,12] In the case of the BRS, scores rated on an interval scale would be helpful in both quantifying the changes in motor function of a patient and comparing the differences in motor function between patients.

Rasch analysis is based on a mathematical model that can estimate person ability (the motor function in our study) and item difficulty (the level of difficulty of each item) simultaneously, and then place the person ability and the item difficulty on the same interval scale.^[13] This method has 3 advantages. First, items that fit the assumptions of the Rasch model are unidimensional, which facilitates examination of the unidimensionality of a measure. Second, Rasch analysis helps users transform an ordinal-level measure into an interval measure. Rasch scores, scored on an interval scale, can be used to represent the estimated person ability and item difficulty. Third, Rasch analysis provides Rasch reliability, which is an indicator of measurement error of the BRS scores. Measurement error is useful in determining whether the motor functions measured by the BRS are close to patients' real motor functions.

Although BRS efficiently assesses motor functions, its utility is restricted because its psychometric properties have not been rigorously examined. This study had 2 purposes. First, we aimed to examine the psychometric properties of the BRS, including the unidimensionality, Rasch reliability, and responsiveness. Second, once the 3 items fitted the assumptions of the Rasch model, we aimed to transform the sum scores of the 3 items (an ordinal scale) into Rasch logit scores (an interval scale). We hypothesized that both the upper-extremity items and the overall motor items of the BRS were unidimensional, the responsiveness of the BRS was comparable with that of the STREAM, and the BRS could be transformed from an ordinal-level measure into an interval-level measure.

2. Methods

2.1. Participants

We retrospectively retrieved a set of admission and discharge data from 2012 to 2014 from medical records of the occupational therapy (OT) department of a medical center. We selected medical records according to the following inclusion criteria: patients who had diagnoses of stroke, and patients who underwent BRS evaluations at admission. Diagnoses of stroke

were based on the *International Classification of Disease, Ninth Revision, Clinical Modification Codes*, including cerebral hemorrhage (431), cerebral infarction (434), or others (430, 432, 433, 436, and 437).

This study was approved by the institutional review board of the medical center.

2.2. Measures

The BRS was designed to describe a sequence of extremity motor recovery after stroke based on the synergy pattern of movement that develops during recovery from a flaccid limb to near-normal and normal movement and coordination.^[2] The BRS contains 3 items: BRS-A, BRS-H, and BRS-L, which are scored on a 6-level Likert-type scale (level I to VI). Higher levels represent better motor function. Clinicians rate a patient's stage based on the patient's spasticity and movement. It takes less than 10 minutes to complete the evaluation. In our study, levels I to VI were respectively recoded into scores of 0 to 5 for Rasch analysis.

The STREAM was designed to provide a comprehensive, objective, and quantitative evaluation of the motor functioning of patients with stroke.^[14] It is a user-friendly measure because the content and format of the STREAM were created in an effort to minimize barriers to routine clinical use. The STREAM contains 30 items divided into 3 subscales: 10 items for voluntary motor ability of the upper-extremity, 10 items for voluntary motor ability of the lower-extremity, and 10 items for basic mobility. The upper-extremity and the lower-extremity subscales were used in our study. A 3-point ordinal scale is used for scoring voluntary movement of the extremities. The STREAM has excellent reliability (intraclass correlation coefficients ≥ 0.97), good concurrent validity with the Fugl-Meyer motor assessment, satisfactory predictive validity ($\rho = 0.69-0.75$), and moderate to large responsiveness (effect size (ES) $d = 0.47-0.51$; standardized response mean (SRM) $= 0.83-1.00$).^[15]

2.3. Statistical analyses

2.3.1. Descriptive analyses. Descriptive statistics were used to analyze the characteristics of the patients and the score distributions of the 3 BRS items. In addition, percentage floor and ceiling effects (i.e., the percentages of patients scoring at the lowest and the highest scale levels, respectively)^[16,17] of the admission BRS data were examined. We calculated the percentages of patients obtaining the highest and the lowest raw sum admission scores on the upper-extremity items and overall motor items. Floor and ceiling effects exceeding 20% were considered notable.^[17,18]

2.3.2. Unidimensionality. We examined the unidimensionality and Rasch reliability of the upper-extremity items and the overall motor items. We did not examine the lower extremity item (BRS-L) because at least 2 items were required for Rasch analysis.

Rasch analysis with the partial credit model (PCM) was used for examining data-model fitting. Unidimensionality was examined using all BRS admission data. To examine the unidimensionality of the BRS, we used infit and outfit statistics, and principal component analysis (PCA). First, the infit and outfit statistics were used to examine whether the item responses fit the expectations of the PCM model. The values of infit (weighted) mean square (MNSQ) and outfit (unweighted) MNSQ were used as the fit indicators of the model. The acceptable ranges of both infit and outfit MNSQ values for each item are from 0.6 to 1.4.^[19]

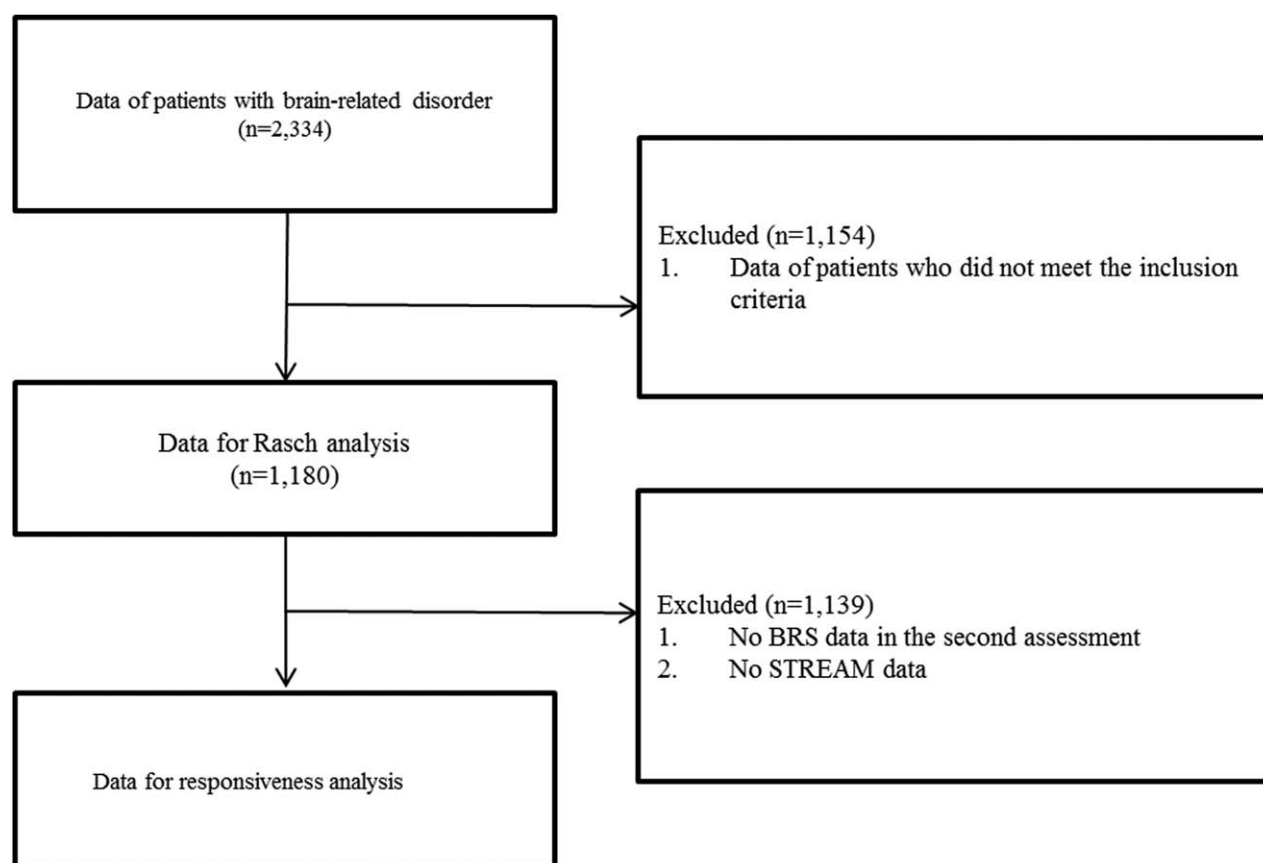


Figure 1. The procedure of data selection in this study.

PCA was further applied to examine the standardized residuals (observed BRS scores minus expected scores). The variance explained by BRS items should be large (>50%), while unexplained variance (the residual) in the first contrast should be small (eigenvalue < 2).^[19]

We also investigated the level of difficulty of the 3 items and the appropriateness of the response category of each item. The level of difficulty was calculated and expressed as a logit score along with Rasch analysis. The appropriateness of the response category was determined by the step difficulties in each item, which should be in order for the design of the response categories to be satisfactory. Disordering of the step difficulties in an item indicates the need for adjustment of the response category.^[20,21] In addition, a person-item map was provided that places the step difficulties of the items (the differences between each response category in each item) and person ability along a continuum. Notable gaps along the step difficulty continuum indicate that additional response categories or items are needed to distinguish patients falling in the gaps.

Moreover, we examined whether patients from different subgroups (age groups or sex) but at the same ability level had equal probabilities of responding positively to the three items.^[22] We performed the differential item functioning (DIF) analyses for the 3 BRS items. DIF was verified by examining the invariance of item difficulties^[23] across different demographic variables, including age (<40, 40 to 64, ≥ 65 years) and sex (male, female). Item bias existed when the DIF contrast was greater than

0.5^[19] and the statistical difference was $P < 0.016$ (0.05/3 age groups) or 0.025 (0.05/2 sex groups) for multiple comparisons using Bonferroni correction.^[25]

2.3.3. Rasch reliability. We examined the Rasch reliability for BRS admission data. We considered Rasch reliability coefficients higher than 0.7 and 0.9 as reliable for group and individual comparisons respectively.^[26]

2.3.4. The quantification of the BRS. When the upper-extremity items and overall motor items fitted the assumptions of the Rasch model, the sum scores of the upper-extremity items and overall motor items were transformed to Rasch scores, respectively.

2.3.5. Responsiveness. We examined responsiveness using data from patients who completed both BRS and STREAM evaluations on admission and at discharge. The responsiveness was examined using the Rasch scores of the BRS and the STREAM. The following 4 steps were conducted to examine and compare the responsiveness of the BRS and the STREAM. First, the sum scores of the upper-extremity and overall motor items on admission and at discharge were transformed into BRS Rasch scores using Rasch scores obtained by the quantification of the BRS. Second, the sum scores of the subscales of the upper-extremity (UE) and lower-extremity (LE) of the STREAM were transformed into STREAM Rasch scores, as suggested by the study of Hsieh.^[27] The Rasch scores of UE and LE were also

Table 1**Descriptive characteristics of the patients with stroke.**

Variable	All participants	Subjects excluded for the responsiveness analysis (n = 1139)	Subjects included for the responsiveness analysis (n = 41)	P*
Age, y: mean (SD)	67.0 (14.6)	67.1 (14.6)	62.8 (13.2)	0.061
Sex (M/F): n (%)	694 (58.8)/486 (41.2)	667 (58.6)/472 (41.4)	27 (65.8)/14 (34.2)	0.351
Diagnosis: n (%)				
Cerebral hemorrhage	345 (29.2)	327 (28.7)	18 (43.9)	0.061
Cerebral infarction	681 (57.7)	660 (57.9)	21 (51.2)	
Other	154 (13.1)	152 (13.3)	2 (4.9)	
Affected sides: n (%)				
Right side	680 (46.2)	530 (46.5)	15 (36.6)	0.442
Left side	477 (40.4)	457 (40.1)	20 (48.8)	
Both sides	158 (13.4)	152 (13.3)	6 (14.6)	
Days after onset at 1st BRS evaluation: median (minimum, maximum)			18 (7, 36)	
Length of stay (days): median (minimum, maximum)			32 (14, 56)	
STREAM (raw score): mean (SD)				
Admission				
Upper-extremity			10.24 (7.35)	
Lower-extremity			12.51 (6.95)	
Discharge				
Upper-extremity			10.44 (6.47)	
Lower-extremity			12.78 (6.08)	

BRS = Brunnstrom recovery stages, STREAM = stroke rehabilitation assessment of movement.

* The demographic variables were examined using the independent *t*-test or the χ^2 test.

summed together to represent the overall motor scores of the STREAM. Third, the responsiveness of the upper-extremity motor function and overall motor function of the BRS and the STREAM were examined using 3 indices: paired *t*-test, ES, and SRM. The paired *t*-test was used to determine the statistical significance of the change in scores. The level of statistical significance was set at $\alpha = 0.05$. The ES is a measure of change obtained by dividing the mean change in scores between assessments on admission and at discharge by the standard deviation (SD) of the assessment on admission. The SRM is the mean change in scores between two measurements divided by the SD of the changes scores. ES and SRM values of 0.20, 0.50, and 0.80 were considered to show small, moderate, and large responsiveness, respectively.^[28]

Last, to further examine if significant differences existed in the ESs and the SRMs between the BRS and the STREAM, we performed a bootstrap procedure by drawing 1000 random samples with replacement from the original data for the BRS and the STREAM. Each bootstrap sample was the same size as the original sample recruited in the responsiveness analysis. The 1000 bootstrap samples produced 1000 pairs of differences in the ESs and the SRMs of both measures. After sorting these differences from lowest to highest, we examined whether the value 0 was included in the 26th and 975th observations (i.e., 95% confidence intervals [CIs]). If yes, it was considered to indicate no significant difference in the ESs and the SRMs of the BRS and the STREAM.

The descriptive statistics and responsiveness analysis were analyzed using SPSS 17.0 software (SPSS Inc., Chicago, IL). The Rasch analysis was performed using Winstep 3.64.2 software (winsteps.com, Beaverton, OR).^[24] The bootstrap procedure was

executed using R 3.1.3 software (R Foundation for Statistical Computing, Vienna, Austria).^[29]

3. Results

3.1. Characteristics of the participants

Figure 1 shows the procedure of the data selection in this study. Admission medical records of 1180 patients with stroke were available. Forty-one subacute patients were further recruited in the responsiveness analysis of the BRS and the STREAM. Table 1 shows the demographic characteristics of the patients. The level distributions of the 3 BRS items are presented in Table 2. Patients who had better motor function (levels V and VI in the 3 BRS items) at admission tended to be lost to follow-up, since patients scoring in levels V and VI in group 1 outnumbered those in group 2. There were no significant differences in patients' age, sex, diagnosis, or affected sides between patients included and excluded from responsiveness analysis ($P = 0.061$ – 0.442).

Only upper-extremity motor function showed a notable ceiling effect, with 20.3% of all patients ($n = 239$) obtaining the highest scores on admission. There were no obvious floor effects in either upper-extremity motor function or overall motor function.

3.2. Unidimensionality

In terms of upper-extremity motor function, both the BRS-A and the BRS-H fitted the model's expectations (infit and outfit MNSQ ranged from 0.80 to 1.06) (Table 3). PCA results showed that 94.7% of the observation variance was explained by both items, and the eigenvalue of unexplained variance in the first contrast was 0.00.

Table 2
Level distributions of the BRS of participants at admission and discharge (N=1180).

	I	II	III	IV	V	VI	Missing
All participants							
<i>Admission</i>							
Arm, n	131	250	98	110	323	267	1
Hand, n	213	179	87	95	293	311	2
Leg, n	91	218	178	93	342	243	15
Subjects for responsiveness analysis							
<i>Admission</i>							
Arm, n	5	12	1	3	14	6	0
Hand, n	12	3	4	2	11	9	0
Leg, n	3	9	5	6	12	6	0
<i>Discharge</i>							
Arm, n	1	11	3	3	10	13	0
Hand, n	2	9	3	4	9	14	0
Leg, n	1	4	8	5	13	10	0

BRS =Brunnstrom recovery stages.

Regarding overall motor function, the BRS-H and BRS-L, but not the BRS-A, fitted the model’s expectations (infit and outfit MNSQ ranged from 0.83 to 1.40) (Table 3). However, we decided to retain the BRS-A in the BRS because the infit and outfit MNSQ were 0.59 and 0.57, respectively, which were very close to the lower bound of 0.6. PCA showed that 88.0% of the variance was explained by the 3 items, and the eigenvalue of unexplained variance in the first contrast was 1.7.

The mean difficulties of the 3 items were very close (Table 3). The biggest difference in difficulty among the 3 items was 0.62 logit (BRS-L = -0.35 logit and BRS-H = 0.27 logit). There was no disordering of the step difficulties in the 3 items. The step difficulties among upper-extremity items and overall motor items are shown in the left and right of Figure 2, respectively. For the upper-extremity items, there were 2 notable gaps among steps, one between step 1 of the BRS-H and step 2 of the BRS-A and between step 4 of the BRS-A and step 5 of the BRS-H. Similar results were found for the overall motor items. Examining each step of the 3 items revealed a notable gap between step 5 of the BRS-A and step 5 of the BRS-L.

No significant DIFs by age and sex were found, indicating that the difficulties of each item were the same across patients in all age groups and both sexes ($P=0.05-0.29$).

Accordingly, the 3 items of the BRS were unidimensional. In addition, the hierarchy of items was identified, which aids in the understanding of progress in the recovery of motor function. No significant DIFs due to age or sex were found.

3.3. Rasch reliability

Rasch reliability coefficients of the upper-extremity motor function were 0.91 at admission and at discharge. Rasch reliability coefficients of overall motor function were 0.92 at admission and 0.91 at discharge.

The high Rasch reliability of the BRS indicates that the Rasch scores of the BRS were precise for both individuals and groups with stroke.

3.4. The quantification of the BRS

Table 4 shows the raw sum scores of upper-extremity items and overall motor items, the corresponding Rasch-transformed scores, and standard errors. Higher scores indicated better motor function. The Rasch scores were dispersed, ranging from -13.5 to 13.1 for upper-extremity motor function, and from -9.4 to 7.6 for overall motor function, indicating that these patients had very different levels of motor function.

In addition, because the BRS contains only 3 items, the standard errors of Rasch-transformed scores were large, ranging from 1.2 to 3.7 for each score of upper-extremity motor function, and from 0.7 to 2.0 for overall motor function.

Accordingly, the raw sum scores of the BRS have been transformed into Rasch scores to represent the motor function of the upper extremities and overall motor function. Moreover, the standard errors of each Rasch score have also been identified.

Table 3
Fit MNSQ statistics, item difficulty, and step parameters of upper-extremity items and overall motor items of the BRS (N=1180).

Item	Infit	Outfit	Item difficulty ^a	Logit	Step1 Logit	Step2 Logit	Step3 Logit	Step4 Logit	Step5 Logit
Upper-extremity motor function									
Arm	1.02	1.06	-0.18		-12.48	-2.97	-0.01	3.25	12.22
Hand	0.88	0.80	0.18		-8.93	-2.93	-0.40	2.30	9.95
Overall motor function									
Arm	0.59	0.57	0.08		-6.75	-1.14	-0.09	1.03	6.95
Hand	0.91	0.83	0.27		-4.77	-1.16	-0.26	0.55	5.64
Leg	1.40	1.17	-0.35		-7.87	-2.27	-0.88	1.28	7.98

BRS =Brunnstrom recovery stages, MNSQ =mean square.

^aItem difficulty represents the average of the step difficulties for the item.

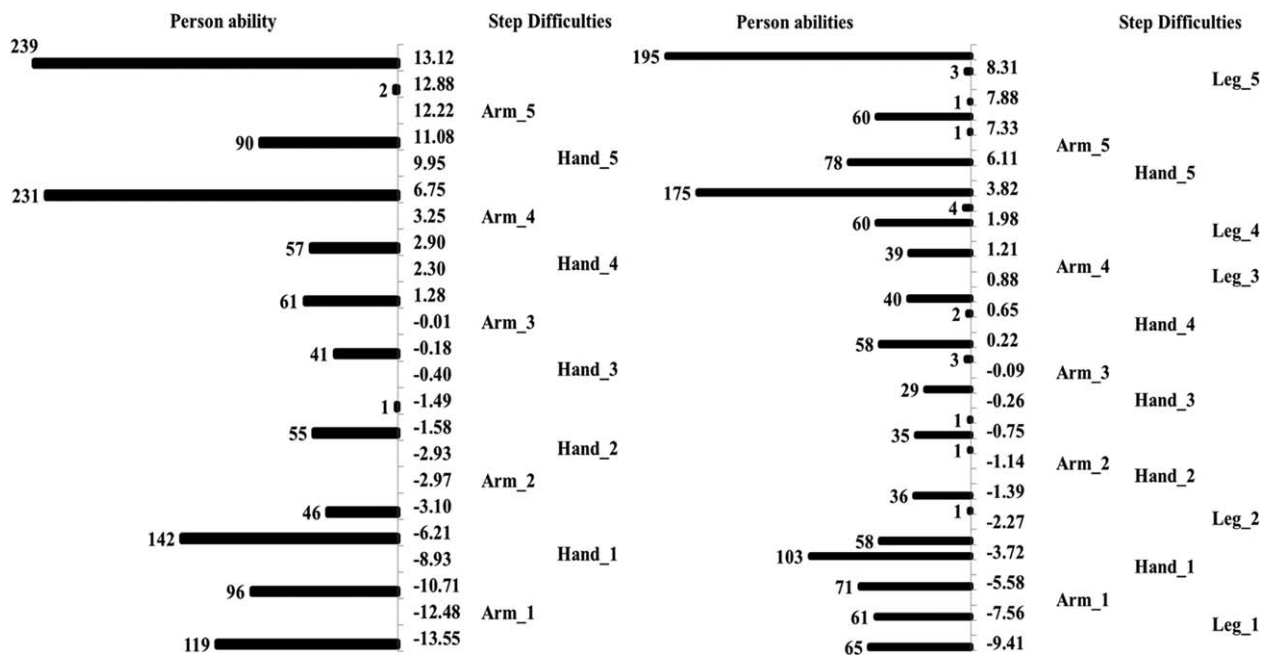


Figure 2. The person–item map of the upper-extremity items (left) and overall items (right) of the BRS. The numbers (such as 1, 2, and 3) next to the items represent the steps of each item; scores in the middle (–13.55 to 13.12 and –9.41 to 9.24) are the logit scores calculated along with Rasch analysis to reflect patients’ motor function or the difficulties of the steps of the items.

3.5. Responsiveness

In terms of responsiveness of the BRS, the change scores in both upper-extremity and overall motor functions were significant between admission and discharge ($P < 0.001$). Both upper-extremity motor function and overall motor function had small to large responsiveness ($ES = 0.35–0.41$, $SRM = 0.85–0.99$) (Table 5).

Regarding the responsiveness of the STREAM, significant differences were also found in the change scores of both upper-extremity and overall motor functions ($P < 0.001$). Both upper-extremity motor function and overall motor function had small

to large responsiveness ($ES = 0.43–0.44$, $SRM = 1.00–1.13$) (Table 5).

For the comparison of responsiveness between the BRS and the STREAM, the values of ES did not show a significant difference between the BRS and the STREAM (95% CI for difference in ES: from –0.21 to 0.09), whereas the values of SRM showed a significant difference between the BRS and the STREAM (95% CI for difference in SRM: from –1.14 to –0.22) (Table 5).

In summary, the responsiveness of the BRS was acceptable, and generally comparable to the STREAM.

4. Discussion

The upper-extremity items (BRS-A and BRS-H) and overall motor items (BRS-A, BRS-H, and BRS-L) both fitted the assumptions of the Rasch model. The results have 2 implications. First, the upper-extremity items and the overall motor items are both unidimensional, supporting the use of their raw sum scores to respectively represent upper-extremity motor function and overall motor function. Therefore, in addition to each item score describing the hand, arm, or leg motor function of a patient, users now have other choices to quantify the patient’s poststroke motor function based on their needs (such as presenting an overall motor function or investigating the treatment effect on overall motor function). The second implication is that the raw sum scores of the BRS can further be transformed into Rasch logit scores, as provided in Table 4. As shown in Table 4, all of the adjacent raw sum scores have different logit intervals, calculated by the subtraction of their corresponding Rasch scores. These findings indicate that assessing patients’ progress or comparing the differences of motor function using raw sum scores may result in over- or underestimation of patients’ progress or differences. Therefore, the Rasch scores of the upper-extremity items and overall motor items appear more appropriate for reflecting the extent of upper-extremity motor function and overall motor

Table 4
Raw sum score, logit score, and standard error of BRS.

Raw sum score	Upper-extremity motor function		Overall motor function	
	Logit score	Standard error	Logit score	Standard error
0	–13.5	2.1	–9.4	2.0
1	–10.7	2.1	–7.6	1.4
2	–6.2	2.7	–1.5	1.0
3	–3.1	1.3	–5.6	1.4
4	–1.6	1.2	–3.7	1.3
5	–0.2	1.2	–2.3	1.1
6	–1.3	1.3	–1.4	0.9
7	–2.9	1.3	–0.8	0.8
8	–6.8	3.7	–0.2	0.7
9	–11.1	1.6	–0.2	0.7
10	–13.1	2.1	–0.7	0.7
11			–1.2	0.8
12			–2.0	1.0
13			–3.8	1.8
14			–6.1	1.3
15			–7.6	1.3

BRS = Brunnstrom recovery stages.

Table 5
Responsiveness of the BRS and the STREAM.

	Upper-extremity motor function		Overall motor function	
	BRS	STREAM	BRS	STREAM
Effect size	0.35	0.43	0.41	0.44
95% Confidence interval of effect size	-0.21 to 0.09	-0.18 to 0.13		
Standardized response mean	0.85	1.00	0.99	1.13
95% Confidence interval of SRM	-1.14 to -0.22	-0.01 to 0.53		

BRS = Brunnstrom recovery stages, SRM = standardized response mean, STREAM = stroke rehabilitation assessment of movement.

function, monitoring progress, and comparing the differences of upper-extremity motor function and overall motor function. The Rasch scores of the BRS are strongly recommended for future users.

The DIF analysis revealed no significant differences in item difficulties for patients across different groups (e.g., different age groups and sexes). The results indicated the same performance patterns in the 3 items of the BRS for patients across different subgroups. Thus, the estimated item difficulties are suitable for use with patients across different subgroups.

Rasch analysis provides the Rasch reliability for each patient's score and the overall Rasch reliability (the average of all patients' reliabilities). The overall Rasch reliabilities of the upper-extremity motor function and overall motor function were above 0.9, with 80% and 77% of the patients having reliability > 0.9 on upper-extremity items and overall motor items, respectively. The Rasch reliability is estimated by the measurement error.^[30] Higher reliability represents smaller measurement error, that is, the measure is more precise. A Rasch reliability of a measure higher than 0.7 indicates that the measure is appropriate for groups of patients. That is, the measure is useful for research purposes. In research contexts, the outcomes of patients are often calculated by averaging the measurement scores, which could reduce the effect of measurement errors on the results.^[31] A Rasch reliability of a measure higher than 0.9 indicates that the measure is suitable for individual patients. Such a high reliability is especially useful for clinical practice, where measurement of an individual patient is usually followed by a specific decision for that patient.^[31] Since the BRS has small measurement error (high Rasch reliability), clinicians can have confidence in using the BRS motor scores to further plan a corresponding intervention protocol and make clinical decisions. Because the BRS has high Rasch reliability, as well as good item-level inter-rater and intra-rater reliability,^[6] as supported by previous studies, the BRS can be used to consistently and precisely assess poststroke motor function.

We examined the responsiveness of the BRS using Rasch scores and found small (indexes of ES) or large (indexes of SRM) responsiveness for both upper-extremity and overall motor function. The results indicate that the BRS has sufficient ability to detect the amount of changes in upper-extremity motor function and overall motor function in a group of subacute patients with stroke. Furthermore, we compared the responsiveness of upper-extremity and overall motor function of the BRS to that of the STREAM. No significant differences were found in 3 out of the 4 indexes of responsiveness, indicating that the responsiveness of the BRS is generally comparable to that of the STREAM. These findings are consistent with those of previous studies, which found that for a group of patients, the short-form and long-form measures presented similar responsiveness.^[9,32,33] Thus, the BRS has great potential for use as an outcome indicator.

An important finding on the estimated item difficulty and step difficulties should be noted. For the BRS items, no disordering exists in the step difficulties (i.e., higher response categories corresponded to higher level of difficulties). Absence of disordering indicates that the ordinal numbering of categories accords with their fundamental meaning. Therefore, the 6 response categories are appropriate for the BRS. However, obvious gaps exist between the step difficulties of each item. The gaps between the step difficulties indicate a lack of appropriate items or response categories to discriminate a patient's motor functions within the gaps. Therefore, we suggest that additional response categories be used to differentiate upper-extremity motor function and overall motor function in patients with stroke who fall in the gaps of the current scale. For example, an additional response category could be added to each BRS item to differentiate those patients whose motor function falls between levels V and VI.

Compared with the sample size used in Rasch analysis ($n = 1180$), the sample size used in the responsiveness analysis was relatively small ($n = 41$). Such a sample size was acceptable but may have restricted the external inference of the responsiveness results. The sample size was acceptable because, with our sample size and the large ES (≥ 0.90), the power of paired *t*-test analysis was 0.99. Moreover, the calculation of ES and SRM is free of the influence of sample size.^[28] Therefore, although the sample size looked modest, our results still provided certain evidence of the responsiveness of the BRS. However, these 41 patients were subacute patients (days after onset at 1st BRS evaluation: 7 to 36 days), and the numbers of people in the levels of each item were not evenly distributed (e.g., only 1 person was in level III in the arm item), which might have restricted the representativeness of our sample. More studies are suggested to cross-validate the responsiveness of the BRS in samples with different characteristics, such as acute or chronic patients.

Our study provided sufficient evidence on the psychometric properties of the BRS despite the retrospective design. Our study had 2 strengths, which provide robust evidence supporting the unidimensionality and Rasch reliability of the BRS. First, a large sample size ($n = 1180$) was used in Rasch analysis, providing stable estimations of item parameters, such as item and step difficulties. Second, our consecutive sampling method increased the representativeness of our sample for patients with stroke. All patients with stroke in the occupational therapy department of a medical center were recruited during the period of 2012 to 2014, except those with comorbidities affecting motor function. Nevertheless, a weakness of our study was the use of a retrospective design, which made missing data inevitable. In addition, multiple raters of the BRS might have increased the measurement error and underestimated the psychometric properties of the BRS. However, our results indicated positive findings. Thus, the aforementioned weaknesses might not be of concern.

This study has 3 contributions. First, we added an extra value to the BRS, that is, transformed the BRS from the ordinal measure into an interval measure. The BRS could preserve the original characteristics such as high accessibility and efficiently identify the motor function of patients' upper and lower extremities. Moreover, the BRS can accurately reflect changes within a patient, and difference between patients. Second, we again verified that the BRS could be a useful assessment tool of poststroke motor function. For practitioners in areas that seldom use the BRS, practitioners have one more option of assessment tool to assess poststroke motor function. Third, for practitioners in areas that widely use the BRS such as Asia, practitioners could have better interpretations of scores of the BRS.

A limitation of this study should be noted. Upper-extremity motor function contains only 2 items. The low number of items in upper-extremity motor function may have reduced the precision of estimates (greater standard errors) in transformed Rasch scores, as mentioned in the Results section, subsection "The quantification of the BRS." Therefore, the interpretations of the results of the transformed interval scores of upper-extremity motor function should be conservative.

5. Conclusion

This study had 2 aims: first, to examine the psychometric properties of the BRS, and second, to transform the ordinal scale of the BRS into an interval scale to represent a patient's motor function. Our results supported the psychometric properties of the BRS, that is, the unidimensionality, high Rasch reliability, and acceptable responsiveness. Moreover, the ordinal sum scores of the BRS could be transformed into interval Rasch scores. Therefore, prospective users could use Rasch scores of the BRS to represent a patient's overall motor function, and to precisely quantify both changes within a patient and differences between patients.

References

- [1] Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. *Lancet Neurol* 2009;8:741–54.
- [2] Brunnstrom S. Motor testing procedures in hemiplegia: based on sequential recovery stages. *Phys Ther* 1966;46:357–75.
- [3] Chang JJ, Sung YT, Lin YT. The relationship between early motor stage and hand function recovery six months after stroke. *Kaohsiung J Med Sci* 1990;6:38–44.
- [4] Safaz I, Yilmaz B, Yasar E, et al. Brunnstrom recovery stage and motricity index for the evaluation of upper extremity in stroke: analysis for correlation and responsiveness. *Int J Rehabil Res* 2009;32:228–31.
- [5] Naghdi S, Ansari NN, Mansouri K, et al. A neurophysiological and clinical study of Brunnstrom recovery stages in the upper limb following stroke. *Brain Inj* 2010;24:1372–8.
- [6] Jush SD, Wang CH, Hsieh CL, et al. The Brunnstrom recovery scale: its reliability and concurrent validity. *J Occup Ther Assoc ROC* 1996; 14:1–2.
- [7] Hashimoto K, Higuchi K, Nakayama Y, et al. Ability for basic movement as an early predictor of functioning related to activities of daily living in stroke patients. *Neurorehabil Neural Repair* 2007;21:353–7.
- [8] Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. USA:Oxford University Press; 2014.
- [9] Chou CY, Chien CW, Hsueh IP, et al. Developing a short form of the Berg Balance Scale for people with stroke. *Phys Ther* 2006;86:195–204.

- [10] Jenkinson C, Clarke C, Gray R, et al. Comparing results from long and short form versions of the Parkinson's disease questionnaire in a longitudinal study. *Parkinsonism Relat Disord* 2015;21:1312–6.
- [11] Tennant A, Geddes JML, Chamberlain MA. The Barthel Index: an ordinal score or interval level measure? *Clin Rehabil* 1996;10:301–8.
- [12] Koh CL, Hsueh I, Wang WC, et al. Validation of the action research arm test using item response theory in patients after stroke. *J Rehabil Med* 2006;38:375–80.
- [13] Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res* 2007;57:1358–62.
- [14] Ahmed S, Mayo NE, Higgins J, et al. The Stroke Rehabilitation Assessment of Movement (STREAM): a comparison with other measures used to evaluate effects of stroke and rehabilitation. *Phys Ther* 2003;83:617–30.
- [15] Hsueh IP, Hsu MJ, Sheu CF, et al. Psychometric comparisons of 2 versions of the Fugl-Meyer Motor Scale and 2 versions of the Stroke Rehabilitation Assessment of Movement. *Neurorehabil Neural Repair* 2008;22:737–44.
- [16] Aaronson NK, Muller M, Cohen PDA, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998;51:1055–68.
- [17] Van der Putten J, Hobart JC, Freeman JA, et al. Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *J Neurol Neurosurg Psychiatry* 1999;66:480–4.
- [18] Holmes WC, Shea JA. A new HIV/AIDS-targeted quality of life (HAT-QoL) instrument: development, reliability, and validity. *Med Care* 1998;36:138–54.
- [19] Linacre JM. *A User's Guide to Winsteps, Ministep: Rasch-Model Computer Programs*. Chicago:Winsteps.com; 2007.
- [20] Decruynaere C, Thonnard J-L, Plaghki L. How many response levels do children distinguish on faces scales for pain assessment? *Eur J Pain* 2009;13:641–8.
- [21] Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3:85–106.
- [22] Teresi JA, Kleinman M, Ocepek-Welickson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Stat Med* 2000;19:1651–83.
- [23] Fleishman JA, Lawrence WF. Demographic variation in SF-12 scores: true differences or differential item functioning? *Med Care* 2003;41: III75–86.
- [24] Linacre, J. M. *Winsteps® Rasch measurement computer program* (Version 3.64.2). Beaverton, Oregon: Winsteps.com 2010.
- [25] Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46:1–8.
- [26] Nunnally JC, Bernstein IH. *Psychometric Theory*. New York:McGraw-Hill; 1994.
- [27] Hsueh I-P, Wang W-C, Wang C-H, et al. A simplified stroke rehabilitation assessment of movement instrument. *Phys Ther* 2006;86:936–43.
- [28] Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- [29] R Core TeamR: A language and environment for statistical computing. Vienna:R Foundation for Statistical Computing; 2014.
- [30] Thissen D, Wainer H, Dorans NJ. *Lawrence Erlbaum Associates, Reliability and Measurement Precision*. Mahwah:2000.
- [31] De Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in Medicine: A Practical Guide*. United Kingdom:Cambridge University Press; 2011.
- [32] Huang Y-J, Chen K-L, Chou Y-T, et al. Comparison of the responsiveness of the long-form and simplified stroke rehabilitation assessment of movement: group-and individual-level analysis. *Phys Ther* 2015;95:1172–83.
- [33] Chen KL, Chen CT, Chou YT, et al. Is the long form of the Fugl-Meyer motor scale more responsive than the short form in patients with stroke? *Arch Phys Med Rehabil* 2014;95:941–9.