Research Report

A Simplified Stroke Rehabilitation Assessment of Movement Instrument

Background and Purpose. An efficient, reliable, and valid instrument for assessing motor function in patients with stroke is needed by both clinicians and researchers. To improve administration efficiency, we applied the multidimensional Rasch model to the 30-item, 3-subscale Stroke Rehabilitation Assessment of Movement (STREAM) instrument to produce a concise, reliable, and valid instrument (simplified STREAM [S-STREAM]) for measuring motor function in patients with stroke. **Subjects and Methods.** The STREAM (consisting of 3 subscales: upper-limb movements, lower-limb movements, and mobility) was administered to 351 subjects with first stroke occurrence and a median time after stroke of 12.5 months. The unidimensionality of each subscale of the STREAM first was verified with unidimensional Rasch analysis. Each subscale of the STREAM then was simplified by deleting redundant items on the basis of expert opinion and the results of the Rasch analysis. The Rasch reliability of the S-STREAM and the concurrent validity of the S-STREAM with the STREAM were examined with multidimensional Rasch analysis and the intraclass correlation coefficient (ICC), respectively. **Results.** After deleting the items that did not fit the Rasch model, we found that the 8-item upper-limb movement subscale, the 9-item lower-limb movement subscale, and the 10-item mobility subscale assessed single, unidimensional upper-limb movements, lower-limb movements, and mobility, respectively. We selected 5 items from each subscale to construct the S-STREAM and found that the reliability of each subscale of the resulting simplified instrument was high (Rasch reliability coefficients of ≥.91). The agreement between the subscale scores (Rasch estimates) of the S-STREAM and those of the STREAM was excellent (ICC of ≥.99, with a lower limit for the 95% confidence interval of ≥.985), indicating good concurrent validity of the S-STREAM with the STREAM. **Discussion and Conclusion.** The S-STREAM demonstrates high Rasch reliability, unidimensionality, and concurrent validity with the STREAM in patients with stroke. Furthermore, the S-STREAM is efficient to administer, as it consists of only half the number of items in the original STREAM. Additional studies to examine other psychometric properties (eg, predictive validity and responsiveness) of the S-STREAM or its psychometric properties in various recovery stages after stroke are needed to further establish its utility in both clinical and research settings. [Hsueh IP, Wang WC, Wang CH, et al. A simplified stroke rehabilitation assessment of movement instrument. *Phys Ther.* 2006;86:936–943.]

Key Words: *Motor function, Psychometrics, Rasch model, Stroke.*

I-Ping Hsueh, Wen-Chung Wang, Chun-Hou Wang, Ching-Fan Sheu, Sing-Kai Lo, Jau-Hong Lin, Ching-Lin Hsieh

Motor and mobility problems are very common after stroke.[1] For the purposes of treatment planning and outcome assessment, it is important to reliably and accurately assess motor function in patients with stroke.[2] Although a number of assessment tools are available to measure the recovery of movement after stroke, they have rarely been used in clinical practice, partly because of the lengthy administration time and complexity of scoring.[3] A reliable, valid, and efficient instrument for the assessment of motor function in patients with stroke is needed by both clinicians and researchers.

The Stroke Rehabilitation Assessment of Movement (STREAM) instrument was designed to provide a comprehensive and quantitative evaluation of voluntary movements (ie, an impairment measurement) and basic mobility (ie, a disability measurement) in patients with stroke.[4] The STREAM consists of three 10-item subscales: upper-limb movements, lower-limb movements, and mobility. The psychometric characteristics of the STREAM have been shown by classical test theory to be satisfactory.[2–6] The STREAM is preferred over other related impairment or disability measures (eg, the Box and Block Test, the Berg Balance Scale, gait speed testing, the Timed "Up & Go" Test, and the Barthel Index) for monitoring recovery from a stroke at the acute stage, as those measures appeared not to focus on the goals of immediate therapy during this period.[5] Furthermore, those measures had limited abilities to discriminate or evaluate patients with stroke because the Box and Block Test, the Berg Balance Scale, gait speed testing, and the Timed "Up & Go" Test showed floor effects in patients with severe stroke, whereas the Barthel Index showed ceiling effects in patients with mild stroke.[5,7–9]

However, the 3 subscales of the STREAM have never been tested for unidimensionality (one type of construct validity); such testing is required to justify the summation of scores to quantify motor function in patients with stroke. Only items measuring the same, unique dimension (construct) should be retained in a measure. Furthermore, the extremely high internal consistency of the STREAM (ie, the Cronbach alpha value was found to be as high as .98 for each of the subscales)[3] indicates possible redundancy among the items. These observations suggest the potential for shortening the STREAM.

Standard Rasch analysis enables the examination of whether items from a scale constitute a unidimensional construct[10,11] so as to construct a concise scale.[12] However, when an instrument consisting of more than one subscale (eg, the STREAM) is to be calibrated, it is inefficient to apply the standard unidimensional Rasch model separately to each subscale. The unidimensional approach ignores correlations between latent traits (ie, the constructs of the subscales) and thus may yield imprecise measurements of the construct (or characteristic) to be measured, especially when the subscales are short. On the other hand, the multidimensional Rasch model simultaneously calibrates all subscales and there-

IP Hsueh, OT, MA, is Assistant Professor, School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan.

WC Wang, PhD, is Professor, Department of Psychology, Chung-Cheng University, Chiayi, Taiwan.

CH Wang, PT, BS, is Professor, School of Physical Therapy, College of Medical Technology, Chung-Shan Medical University, and Department of Physical Therapy, Chung-Shan Medical University Rehabilitation Hospital, Taichung, Taiwan.

CF Sheu, PhD, is Professor, Institute of Cognitive Science, National Cheng Kung University, Tainan, Taiwan.

SK Lo, PhD, is Professor, Faculty of Health and Behavioral Sciences, Deakin University, Melbourne, Australia.

JH Lin, PT, PhD, is Professor, Faculty of Physical Therapy, Kaohsiung Medical University, Kaohsiung, Taiwan.

CL Hsieh, OT, PhD, is Professor and Chair, School of Occupational Therapy, College of Medicine, National Taiwan University, and Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, 4F, 17 Shiujou Rd, Taipei 100, Taiwan (mike26@ha.mc.ntu.edu.tw). Address all correspondence to Dr Hsieh.

Ms Hsueh and Dr Hsieh provided concept/ideas/research design, writing, and fund procurement. Dr WC Wang provided data analysis and facilities/equipment. Dr Hsieh provided project management. Professor CH Wang provided institutional liaisons. Dr WC Wang, Professor CH Wang, Dr Sheu, Dr Lo, and Dr Lin provided consultation (including review of manuscript before submission).

The research protocol was approved by local institutional review boards.

This study was supported by research grants from the National Science Council (NSC 93-2314-B-002-033 and NSC 94-2314-B-002-078) and the National Health Research Institute (NHRI-EX94-9204PP).

This article was received April 16, 2005, and was accepted January 30, 2006.

fore uses the correlations to increase measurement precision.[13,14] Theoretically, it may be difficult to conceive of constructs that are independent in the movement domains after stroke. Therefore, the multidimensional Rasch model takes into account the between-subscale correlations to increase measurement precision: the higher the correlations, the greater the measurement precision.[15,16] In other words, short subscales, if moderately correlated, still can yield precise measurements with the multidimensional approach. Because the 3 subscales of the STREAM are highly correlated with each other,[2] the multidimensional approach can be useful in simplifying the STREAM.

To improve administration efficiency, we aimed to shorten the 30-item, 3-subscale STREAM to produce a simplified STREAM (S-STREAM) by using the multidimensional Rasch model. We examined the psychometric properties of the S-STREAM (including Rasch reliability, unidimensionality, and concurrent validity with the STREAM) in subjects with stroke.

Method

Subjects

Subjects with a broad range of motor deficits were recruited from the rehabilitation departments of 5 hospitals in northern, central, southern, and eastern Taiwan between October 2003 and January 2004. These rehabilitation departments provide both inpatient and outpatient services (including physical therapy, occupational therapy, and speech therapy). Subjects were included in the study if they met the following criteria: diagnosis (*International Classification of Diseases, 9th Revision, Clinical Modification* [ICD-9] codes)[17] of cerebral hemorrhage (ICD-9 code 431) or cerebral infarction (ICD-9 code 434), absence of other major diseases (eg, tumors, arthritis) or impairments (eg, amputations, fractures) that would reduce or limit a subject's ability to perform movements, and ability to follow 2-step instructions. Only subjects who had had their first stroke and were able to give informed consent personally or by proxy were included in the study.

Procedure

The STREAM, with instructions in Chinese, was administered by the same physical therapist to all of the participants in the 5 rehabilitation departments. The intrarater reliability of data obtained by the physical therapist was satisfactory (intraclass correlation coefficient [ICC] of .94). Demographic characteristics and comorbidity data for the participants were collected from medical records.

Instrument

Items of the STREAM[3] for voluntary movements of the limbs are scored on a 3-point scale (0=unable to perform the test movement, 1=able to perform the test movement only partially, and 2=able to complete the test movement). Mobility items are scored on a 4-point scale (0=unable to perform the test movement, 1=able to perform the test movement only partially, 2=able to complete the test movement with a mobility aid, and 3=able to complete the test movement without an aid). Thus, each of the 10-item limb movement subscales was scored out of 20 points, and the 10-item mobility subscale was scored out of 30 points.

Data Analysis

The unidimensionality of the 3 subscales of the STREAM was examined with WINSTEPS.[18] The variance-covariance matrix (and the correlation matrix) for the 3 latent traits (ie, the constructs of the 3 subscales of the STREAM) was computed with ConQuest,[19] which was developed for the multidimensional random-coefficients multinomial logit model (MRCMLM).[13] A brief description of the MRCMLM is given in the Appendix.

To examine the unidimensionality of each subscale, infit and outfit statistics were used to examine whether the data fit the expectation of the Rasch rating scale model (RSM). The infit mean square (MNSQ) is sensitive to unexpected behavior affecting responses to items near the subject's proficiency measure (eg, motor status); the outfit MNSQ is sensitive to unexpected behavior on items far from the subject's motor status. Items with infit or outfit MNSQ values of greater than 1.4 indicate potential misfits.[20] The MNSQ can be transformed to a standardized z value (ZSTD) which, for large samples, follows approximately the standard normal distribution when the items fit the expectation of the model. Items with both infit and outfit ZSTD values beyond $\pm$2.58 (twice the tailed area of the normal curve above or below $\pm$2.58 is 0.01) were considered to have poor fit.

In addition, when items fit the expectation of the model, the residuals (observed scores minus expected scores) should be distributed randomly. A principal components analysis was conducted to determine whether any dominant component existed among the residuals. If dominant components were found, then the unidimensionality assumption was violated.

Rasch reliability, which can be viewed as the counterpart of classical test reliability (eg, the Cronbach alpha), was calculated.[10,20] Reliability coefficients of greater than .7 were considered good for group comparisons, whereas those greater than .9 were considered good for individual comparisons.[21]

The appropriateness of the scoring levels in each item of the STREAM was investigated with the RSM. The RSM is useful for polytomous items in a scale that share the same rating scale structure (eg, all items are rated 0, 1, or 2). Estimates of the threshold difficulty between the adjacent scoring levels can be used to examine the appropriateness of the scoring points of a test.[20] If disorderings of the step difficulty (ie, the difficulty of a higher step was lower than that of its adjacent lower step) between any 2 adjacent levels were found, then the levels of scaling of the items might be reorganized to achieve suitable scaling.

After the unidimensionality and appropriate scoring levels in each item of the STREAM were established, we attempted to reduce the length of the test further while maintaining its psychometric properties. Each of the 3 subscales of the STREAM was shortened to produce the S-STREAM on the basis of 2 criteria: content representativeness, assessed by a panel of therapists (2 physical therapists and 2 occupational therapists who each had more than 10 years of experience in stroke rehabilitation); and difficulty diversity, that is, even scattering of the difficulties of the selected items over the range of the difficulty continuum.

For each subject, the multidimensional form of the RSM can provide estimates for the 3 subscale scores for both the S-STREAM and the STREAM. We used the RSM estimates for each subscale of the STREAM as the gold standard in this study. Because the Rasch estimates for each subscale have different score ranges, all estimates were linearly transformed to a range of 0 to 100 to facilitate comparisons. The relationship and agreement among corresponding Rasch estimates for subscale scores (ie, the concurrent validity of the S-STREAM with the STREAM) were examined with the Pearson correlation coefficient (r) and the ICC(3,1),[22] respectively. Correlation coefficients of greater than .6 indicate acceptable concurrent validity.[23]

Results

A total of 351 subjects with a median time after stroke of 12.5 months met the selection criteria and agreed to participate in the study. The participants had a wide range of motor deficits, and their scores were found to be scattered throughout the entire ranges of the STREAM subscales. The clinical characteristics of the study participants are shown in Table 1.

Unidimensional Rasch analyses of the 3 subscales, separately, revealed that 2 items (scapular elevation and opposition) in the upper-limb movement subscale and 1 item (hip abduction) in the lower-limb movement subscale did not fit the expectation of the model (both infit and outfit ZSTD values of beyond ±2.58). These 3 items were removed from the instrument in subsequent analyses. Thereafter, the 8-item upper-limb movement subscale and the 9-item lower-limb movement subscale fit the expectation of the model (infit and outfit ZSTD values within the range of ±2.58). In addition, none of these items had infit and outfit MNSQs of greater than 1.4. Principal components analysis revealed that no dominant component existed among the residuals of the Rasch-transformed scores for the 8-item upper-limb movement subscale, the 9-item lower-limb movement subscale, or the 10-item mobility subscale. These results indicate that the 8-item upper-limb movement subscale, the 9-item lower-limb movement subscale, and the 10-item mobility subscale assess single, unidimensional upper-limb movements, lower-limb movements, and mobility, respectively.

A multidimensional analysis with ConQuest was performed for the remaining 27 items (ie, 8 items from the upper-limb movement subscale, 9 items from the lower-limb movement subscale, and 10 items from the mobility subscale). Table 2 shows the correlation matrix for the

Table 1.
Clinical Characteristics of Subjects With Stroke (N=351)

Characteristic[a]	Value
Sex (no. of men/women)	222/129
Age, y, median (25th–75th percentiles)	63 (53–71)
Month after onset, median (25th–75th percentiles)	12.5 (4–30)
Diagnosis, no. (%) of subjects	
Cerebral hemorrhage	113 (32)
Cerebral infarction	238 (68)
Side of paresis, no. (%) of subjects	
Right	175 (50)
Left	176 (50)
STREAM score, median (25th–75th percentiles)	
Upper-limb movement raw score	9 (0–17)
Lower-limb movement raw score	7 (3–14)
Mobility raw score	15 (8–21)
S-STREAM score,[b] mean (SD)	
Upper-limb movement score	48.8 (26.4)
Lower-limb movement score	48.3 (24.0)
Mobility score	49.2 (22.9)

[a] STREAM=Stroke Rehabilitation Assessment of Movement, S-STREAM=simplified STREAM.
[b] Rasch-transformed score ranging from 0 to 100.

Table 2.
Correlation Matrix for the Stroke Rehabilitation Assessment of Movement With the Multidimensional Approach

Subscale	Upper-Limb Movement	Lower-Limb Movement
Lower-limb movement	.90	
Mobility	.78	.84

Table 3.
Rasch Reliability for the 3 Subscales of the Stroke Rehabilitation Assessment of Movement (STREAM) and the Simplified STREAM (S-STREAM) With the Unidimensional and Multidimensional Approaches

	27-Item STREAM			S-STREAM		
	No. of	Reliability		No. of	Reliability	
Subscale	Items	Unidimensional	Multidimensional	Items	Unidimensional	Multidimensional
Upper-limb movement	8	.86	.93	5	.85	.91
Lower-limb movement	9	.91	.96	5	.88	.93
Mobility	10	.97	.98	5	.94	.95

STREAM, which revealed that the underlying latent traits of the subscales of the STREAM were highly correlated, with Pearson coefficients of between .78 and .90. Table 3 shows that the Rasch reliability for the 3 subscales was good (reliability coefficients of ≥.86).[20] Moreover, the 3 subscales of the STREAM showed better reliability when the multidimensional approach was used (reliability coefficients of ≥.93) than when the unidimensional approach was used (reliability coefficients of ≥.86).

We selected items that fit the RSM to construct the S-STREAM. To avoid possible floor or ceiling effects, we generally retained the most difficult and easiest items for each subscale. The only exception was that the most difficult item of the mobility subscale, 3 steps backward, was not selected because it apparently cannot be classified as a daily mobility activity. The second most difficult item, walking down 3 stairs, was selected instead. These 2 items were similar in difficulty (Tab. 4). Furthermore, 3 items with an intermediate degree of difficulty were selected for each subscale on the basis of expert opinion and the results of Rasch analysis. We first selected items that were scattered evenly over the range of the difficulty continuum. Some items of similar difficulty (eg, "making a fist" and "moving hand to sacrum" in the upper-limb movement subscale, "knee flexion while sitting" and "knee flexion while standing" in the lower-limb movement subscale, and "10-m walk" and "3 steps to the affected side" in the mobility subscale) were selected by the panel of therapists. "Making a fist" was used to provide 2 items, in total, measuring hand function for the upper-limb movement subscale. "Knee flexion while standing" was not used because some subjects tended to flex their hips simultaneously, making the rating difficult. "Three steps to the affected side" was not used because it is obviously not a daily activity, compared with "10-m walk." The final 5 items used for each subscale of the S-STREAM are shown in Table 4.

Unidimensional and multidimensional analyses were conducted on the 15-item version of the S-STREAM. Table 4 shows the 2 kinds of item parameter estimates (ie, difficulty logit and standard error) for the 3 subscales of the 27-item STREAM and the S-STREAM with the multidimensional approach. These 2 versions of the instrument had similar estimates for corresponding items.

The threshold difficulty estimates within each subscale were rather far apart (≥2.18 logits). In addition, the ordering of the threshold difficulty estimates was not reversed.

Table 3 shows that the use of the multidimensional approach with the S-STREAM resulted in high test reliability (Rasch reliability coefficients of ≥.91) for the 3 subscales. These results indicate that the 3 subscales of the S-STREAM can yield very precise estimates for individual subjects. When the unidimensional approach was used, the test reliability values were .85, .88, and .94 for the upper-limb movement, lower-limb movement, and mobility subscales of the S-STREAM, respectively.

The agreement between each pair of subscales was excellent (transformed scores of 0–100), with ICCs (95% confidence intervals) of .99 (.993–.995), .99 (.989–.993), and .99 (.985–.990), for the upper-limb movement, lower-limb movement, and mobility subscales, respectively. Furthermore, the Pearson correlation coefficients for the multidimensional Rasch estimates for the STREAM and the S-STREAM were all .99 for the 3 subscales. These results indicate that each subscale of the S-STREAM demonstrates high concurrent validity with the corresponding subscale of the STREAM.

Discussion

To the best of our knowledge, this study is the first to use the multidimensional approach to produce a concise measure of motor function for people with stroke. The 15-item S-STREAM was constructed on the basis of the original STREAM, expert opinion, and results of Rasch analysis. The S-STREAM contains only half the number of items in the original STREAM and shows sound reliability and validity.

There are 2 major advantages of using the S-STREAM. First, it is simple and quick to use for patients with stroke compared with the original STREAM. As the S-STREAM contains only half the number of items in the original

Table 4.
Item Parameter Estimates for the Stroke Rehabilitation Assessment of Movement (STREAM) and the Simplified STREAM (S-STREAM) With the Multidimensional Approach

	27-Item STREAM		15-Item S-STREAM	
Item[a]	Difficulty Logit	SE	Difficulty Logit	SE
Upper-limb movement				
Elbow extension while supine	−0.77	0.30	−0.78	0.32
Raising hand to touch top of head	−0.46	0.30		
Scapular protraction	−0.02	0.31	−0.10	0.30
Making a fist	1.21	0.32	1.05	0.31
Moving hand to sacrum while sitting	1.29	0.32		
Raising arm to fullest elevation	1.50	0.32	1.31	0.31
Supination and pronation	1.78	0.33		
Total extension of fingers	2.06	0.33	1.84	0.32
Threshold 1	−2.29	0.10	−2.15	0.11
Threshold 2	2.29	0.10	2.15	0.11
Lower-limb movement				
Knee extension while sitting	−2.12	0.27	−1.98	0.26
Hip flexion while sitting	−0.83	0.26	−0.75	0.27
Bending hip and knee while supine	−0.40	0.26		
Knee flexion while sitting	0.87	0.27	0.84	0.27
Knee flexion while standing	0.93	0.27		
Dorsiflexion while sitting	1.26	0.27		
Plantar flexion while sitting	1.56	0.27	1.47	0.27
Knee extension and dorsiflexion while sitting	2.12	0.28		
Dorsiflexion while standing	3.63	0.30	3.41	0.30
Threshold 1	−1.09	0.07	−1.07	0.08
Threshold 2	1.09	0.07	1.07	0.08
Mobility				
Rolling	−3.77	0.24	−3.84	0.27
Bridging (ie, raising hips off bed)	−3.37	0.23		
Moving from supine to sitting	−1.49	0.22	−1.52	0.21
Standing for 20 counts by the rater	−0.71	0.22		
Moving from sitting to standing	0.32	0.22	0.30	0.23
Placing affected foot onto first step	0.68	0.22		
10-m walk	1.32	0.22	1.28	0.22
3 steps to affected side	1.40	0.22		
Walking down 3 stairs	2.16	0.22	2.11	0.22
3 steps backward	2.26	0.22		
Threshold 1	−3.53	0.10	−3.47	0.13
Threshold 2	0.13	0.06	−0.13	0.06
Threshold 3	3.40	0.10	3.61	0.12

[a] The items selected for the S-STREAM are shown in italic type. The items are arranged in ascending order of difficulty in each subscale. Threshold indicates difficulty between the adjacent scoring levels. Note that the items of the mobility subscale have 4 levels of scaling and thus have 3 thresholds.

STREAM, the 15-item S-STREAM can be administered within 10 minutes, that is, half the time required to administer the original STREAM. Rapid assessment is a clinically important feature of this simplified version of the STREAM, as long tests can take a substantial amount of time to complete and may place unreasonable demands upon the respondents, especially in instances in which they may be seriously unwell, as in the case of stroke. Rapid and accurate assessment of functional outcomes in patients with stroke therefore will provide benefits to both clinicians and patients.

A second advantage of using the S-STREAM is that the Rasch estimates for the 3 subscales can be viewed as interval-level measurements.[10] In contrast, most measures currently used in the assessment of patients with stroke use ordinal-level measurements. For an ordinal scale, a given difference in scores at one point on the scale does not necessarily represent the same amount of functional change as an identical difference at another point on the scale.[24] Interval scores, rather than ordinal scores, can provide a more precise reflection and better resolution of disease impact, differences between individuals and groups, and treatment effects.[25] Furthermore, an ordinal scale precludes the use of standard parametric statistical inferences. Because most statistical techniques assume that the data are at least on an interval scale, the Rasch estimates for the S-STREAM are recommended for future applications.

With the multidimensional approach, the between-subscale correlations are taken into account to improve measurement precision. Patients with the same raw upper-extremity scores but with different lower-extremity scores or mobility scores would have different Rasch estimates for their upper-extremity scores. The Rasch estimates for each subscale of the S-STREAM derived from the multidimensional analysis cannot be obtained by summing the raw scores and using a simple Rasch transformation table, as in the unidimensional analysis. Because the transformation table for the multidimensional analysis of the S-STREAM is very long, we have developed a computer program to transform the raw scores for each subscale of the S-STREAM to the Rasch scores. The program is easy to run on common PC platforms. To improve the dissemination of the program and the S-STREAM,[26] the related materials can be found at http://ccms.ntu.edu.tw/~clhsieh/s-stream/. Even if some patients do not respond to all items, their Rasch scores still can be estimated and compared because, with the use of the models of the Rasch family (or item response models in general), the estimation of a patient's latent traits is based on the patient's observed item responses.[11]

In this study, multidimensional Rasch analysis was shown to be a useful tool for reducing the items of a measure while maintaining the measurement reliability and validity (eg, the Rasch reliability coefficients of the S-STREAM were above the preset criterion of .9, and the subscales of the S-STREAM were highly associated with the corresponding subscales of the STREAM). Furthermore, the multidimensional Rasch model yielded a large number of estimates of a subject's motor function (eg, 191 estimates for the upper-limb movement function in this study) compared with the raw scores of the S-STREAM (eg, 0–10 for the upper-limb movement function). These additional estimates of motor function are likely to promote the psychometric properties (eg, responsiveness and discriminative capacity) of the S-STREAM, although further validation is warranted.

It also should be noted that direct estimation of the correlation among latent traits is possible only for the multidimensional approach and not for the unidimensional one.[15,16] Rasch analysis can achieve even more efficient and precise measurements when computerized adaptive testing (CAT)[27–29] is used; CAT involves the use of a computer to administer items to respondents and allows respondents' levels of function to be estimated as precisely as desired (ie, to reach a preset reliability level). Because the impacts after stroke are multiple and a great deal of time and effort is needed to administer the measures that assess various impacts, it seems promising to combine both the multidimensional approach and CAT to simplify or elaborate on functional measurements in patients with stroke.[14]

The appropriateness of scoring levels refers to whether or not the motor functions of participants can be differentiated by their responses as clearly as the levels allow.[20] Recent studies[30,31] have shown that a larger number of scoring points may not lead to a finer differentiation of participants. The items of the subscales of the STREAM are on a 3-point or 4-point ordinal scale, but the appropriateness of scoring levels of the STREAM have rarely been examined. Our study is the first to determine the appropriateness of its scaling in subjects with stroke. We found that the threshold difficulty estimates within each subscale were rather far apart and without disorderings (ie, the ordering of the threshold difficulty of the levels was reasonable). Therefore, the rating scales of the STREAM were supported, indicating that they could differentiate the motor status of subjects very well.

Any measurement tool requires an extensive psychometric examination for the purposes of understanding its particular strengths and limitations.[32] Additional studies to examine other psychometric properties (eg, predictive validity and responsiveness) of the S-STREAM are warranted. Furthermore, patients with stroke at the acute or subacute stage receive greater intensity of motor rehabilitation and assessment than do those at the chronic stage. However, more than half of the subjects in this study had had a stroke more than 1 year before the study; therefore, the psychometric properties of the S-STREAM at the acute and subacute stages remain largely unknown. Therefore, further investigations of the psychometric properties of the S-STREAM at various recovery stages after stroke are needed to further establish its utility in both clinical and research settings. Direct psychometric and practical (utility) comparisons between the S-STREAM and other related impairment and disability measures (eg, the Fugl-Meyer Motor Test and the Rivermead Mobility Index) also are needed for prospective users to select a better measure based on empirical data.

Conclusion

Our results show that the S-STREAM has high Rasch reliability, unidimensionality, and concurrent validity with the STREAM in patients with stroke. The S-STREAM is efficient to administer, as it consists of only half the number of items in the original STREAM. Additional studies to examine the predictive validity and responsiveness of the S-STREAM or its psychometric properties in various recovery stages after stroke are needed to further establish its utility.

References

1 Kelly-Hayes M, Robertson JT, Broderick JP, et al. The American Heart Association Stroke Outcome Classification. *Stroke*. 1998;29: 1274–1280.

2 Wang CH, Hsieh CL, Dai MH, et al. Inter-rater reliability and validity of the Stroke Rehabilitation Assessment of Movement (STREAM) instrument. *J Rehabil Med*. 2002;34:20–24.

3 Daley K, Mayo N, Wood-Dauphinee S. Reliability of scores on the Stroke Rehabilitation Assessment of Movement (STREAM) measure. *Phys Ther*. 1999;79:8–19.

4 Daley K, Mayo N, Danys I, et al. The Stroke Rehabilitation Assessment of Movement (STREAM): refining and validating the content. *Physiother Can*. 1997;49:269–278.

5 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;83:617–630.

6 Hsueh IP, Wang CH, Sheu CF, Hsieh CL. Comparison of psychometric properties of three mobility measures for patients with stroke. *Stroke*. 2003;34:1741–1745.

7 Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel Activities of Daily Living Index in stroke patients. *J Formos Med Assoc*. 2001;100:526–532.

8 Mao HF, Hsueh IP, Tang PF, et al. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke*. 2002;33:1022–1027.

9 Croarkin E, Danoff J, Barnes C. Evidence-based rating of upper-extremity motor function tests used for people following a stroke. *Phys Ther*. 2004;84:62–74.

10 Wright BD, Mok M. Rasch models overview. *J Appl Meas*. 2000;1: 83–106.

11 Wright BD, Stone MH. *Best Test Design: Rasch Measurement*. Chicago, Ill: MESA; 1979.

12 Jenkinson C, Norquist JM, Fitzpatrick R. Deriving summary indices of health status from the Amyotrophic Lateral Sclerosis Assessment Questionnaires (ALSAQ-40 and ALSAQ-5). *J Neurol Neurosurg Psychiatry*. 2003;74:242–245.

13 Adams RJ, Wilson MR, Wang WC. The multidimensional random coefficients multinomial logit model. *Appl Psychol Meas*. 1997;21:1–23.

14 Wang WC, Chen PH. Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Appl Psychol Meas*. 2004;28:295–316.

15 Wang WC. Direct estimation of correlation as a measure of association strength using multidimensional item response models. *Educational and Psychological Measurement*. 2004;64:937–955.

16 Wang WC, Chen PH, Cheng YY. Improving measurement precision of test batteries using multidimensional item response models. *Psychol Methods*. 2004;9:116–136.

17 National Center for Health Statistics. *International Classification of Diseases, 9th Revision, Clinical Modification*. Hyattsville, Md: National Center for Health Statistics, Centers for Disease Control and Prevention, US Department of Health and Human Services; 1980.

18 Linacre JM. A user's guide to WINSTEPS MINISTEP Rasch-model computer programs. Available at: http://www.winsteps.com/winsteps.htm. Accessed on January 12, 2005.

19 Wu ML, Adams RJ, Wilson MR. *ConQuest*. Camberwell, Victoria, Australia: Australian Council for Educational Research; 1998.

20 Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.

21 Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11:193–205.

22 Shrout PE, Fleiss JL. Intraclass correlation: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.

23 Salter K, Jutai JW, Teasell R, et al. Issues for selection of outcome measures in stroke rehabilitation: ICF participation. *Disabil Rehabil*. 2005;27:507–528.

24 Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil*. 1989;70: 857–860.

25 Hobart J. Measuring disease impact in disabling neurological conditions: are patients' perspectives and scientific rigor compatible? *Curr Opin Neurol*. 2002;15:721–724.

26 Jette AM. Invention is hard, but dissemination is even harder. *Phys Ther*. 2005;85:390–391.

27 McHorney CA. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. *Ann Intern Med*. 1997;127:743–750.

28 McHorney CA. Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Ann Intern Med*. 2003;139:403–409.

29 van der Linden WJ, Glas CAW. *Computerized Adaptive Testing: Theory and Practice*. Boston, Mass: Kluwer Academic Publishers; 2000.

30 Hsueh IP, Wang WC, Sheu CF, Hsieh CL. Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. *Stroke*. 2004;35:721–726.

31 Wang CH, Hsueh IP, Sheu CF, et al. Psychometric properties of 2 simplified 3-level balance scales used for patients with stroke. *Phys Ther*. 2004;84:430–438.

32 Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil*. 1993;74:531–536.

Appendix.

Multidimensional Random-Coefficients Multinomial Logit Model (MRCMLM)

In the MRCMLM, let subject n's abilities on L latent traits (ie, the 3 constructs to be measured in this study: upper-limb movement, lower-limb movement, and mobility) be denoted as $\theta_n^T = (\theta_{n1}, \ldots, \theta_{nL})$, which is considered to represent a random sample from a population with multivariate normal distribution $N(\mu, \Sigma)$, where μ and Σ are the means and variance-covariance matrices of the latent traits, respectively. The probability of a response in scaling level j of item i for subject n is

$$p_{nij} = \frac{\exp(b_{ij}^T \theta_n + a_{ij}^T \xi)}{\sum_{u=1}^{K_i} \exp(b_{iu}^T \theta_n + a_{iu}^T \xi)},$$

where K_i is the number of levels in item i (in this study, K_i=3 for the items of the upper-limb and lower-limb movement subscales and K_i=4 for the items of the mobility subscale), ξ is a vector of location parameters that describe the items, b_{ij} is a score vector given to scaling level j of item i across L latent traits, and a_{ij} is a design vector given to scaling level j of item i that describes the linear relationship among the elements of ξ.