

ORIGINAL ARTICLE

Improvement of the Van Lieshout hand function test for Tetraplegia using a Rasch analysis

AIF Spooren^{1,2}, C Arnould³, RJEM Smeets^{1,2}, HMH Bongers⁴ and HAM Seelen^{1,2}

Study design: Cross-sectional study

Objective: The Van Lieshout hand function test for tetraplegia (VLT) measures the quality of arm–hand functioning in persons with tetraplegia. It is valid, reliable and responsive. However, it does not satisfy all the criteria for interval level measurement. The present study aims to apply the Rasch model to the VLT short form (VLT-SF) to upgrade its scale type towards interval level, and to verify if the requirements of an objective measure are satisfied in the revised version.

Setting: Eight Dutch Rehabilitation centres.

Methods: The VLT-SF data of 73 tetraplegic patients were Rasch-analysed (RUMM2030 software, RUMM Laboratory Pty Ltd, Perth, Australia) to verify the order of response categories, unidimensionality and reliability of the VLT-SF, and to assess its applicability regardless of (motor) lesion completeness.

Results: Seven of the ten VLT-SF items showed disordered response categories. The six original response categories were therefore recoded into three or four categories. After recoding, all items satisfied the model requirement of unidimensionality. The items were relatively well-targeted on the subjects' arm–hand skilled performance measures, leading to a good person separation index ($R=0.91$). The difficulty hierarchy of the VLT-SF items was invariant across patient subgroups of (motor) lesion completeness.

Conclusions: Provided that response categories are recoded, VLT-SF Rasch analysis showed that the requirements of an objective measure were satisfied. This allows to compare the measurements of different patients quantitatively, and to follow their results over time.

Spinal Cord advance online publication 18 June 2013; doi:10.1038/sc.2013.54

Keywords: hand function test; tetraplegia; Rasch analysis; upper extremity

INTRODUCTION

Voluntary control over the arm and hand is an important issue during and after the rehabilitation of persons with a cervical spinal cord injury (C-SCI).^{1,2} Patients are more interested in what they actually will be able to do with their hands (ICF activity level³), rather than (for example) the amount of force they can generate (ICF body function level). Moreover, therapists are particularly interested in the way patients perform daily activities. In order to rate the quality of upper limb (arm and hand) performance regarding activities that are performed commonly or daily, the Van Lieshout hand function test for Tetraplegia (VLT) has been developed^{4,5} (www.vlt-test.eu). Next to evaluation of arm–hand performance, and given the ordered structure of the test score levels per test item, the VLT may also provide insight to the therapist about the next higher performance level on a test item a specific patient may attain. The clinical version of the VLT consists of 19 items divided into 5 areas of interest, that is, arm ability to transfer the body, arm positioning and stabilizing, hand opening and closing, grasping and releasing, and manipulating (www.vlt-test.eu). Each item has six ordinal response categories ranging from 0 (unable to perform the task) to 5 (maximum performance level). An example of one task item is given in Figure 1.

A short, less time consuming 10-item VLT research version (VLT short form or VLT-SF) (www.vlt-test.eu) has been developed and proved to be valid, reliable and responsive.^{5,6} The 10 VLT-SF are listed in Table 1. However, both VLT and VLT-SF generate ordinal total scores while relying on counts of potentially unequal units. Ordinal scores of the item response categories are separated by unknown distances. Moreover, obtaining the same score to different items does not necessarily represent the same amount of the measured variable.^{7,8} As the VLT-SF does not fully satisfy the criteria of an objective scientific measure,⁹ its interpretation and use are somewhat restricted. To correct this flaw, the ordinal scores should be converted into linear interval level scores before quantitative comparisons across patients and across time can be done. Of the measurement models that can be used to estimate linear measures from raw total scores, the Rasch model is the most promising.¹⁰ This model prescribes that only the item difficulty, the person's ability (in terms of the VLT-SF the arm–hand performance) and the threshold difficulty (thresholds are the ability levels required to have an equal probability of endorsing one response rather than the previous one (in terms of the VLT the test score level (www.scionn.nl) determine the probabilities of endorsing any category to an item), regardless of other subjects'

¹Rehabilitation Sciences, Adelante Centre of Expertise in Rehabilitation and Audiology, Hoensbroek, The Netherlands; ²Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University, Maastricht, The Netherlands; ³Departments of Paramedical Category, Physical and Occupational Therapy, Haute Ecole Louvain en Hainaut, Montignies-sur-Sambre, Belgium and ⁴Adelante Rehabilitation Centre, Hoensbroek, The Netherlands
Correspondence: Dr AIF Spooren, Rehabilitation Sciences, Adelante Centre of Expertise in Rehabilitation and Audiology, Zandbergsweg 111, Hoensbroek, Limburg 6432 CC, The Netherlands.

E-mail: a.spooren@adelante-zorggroep.nl

Received 23 July 2012; revised 17 December 2012; accepted 27 December 2012

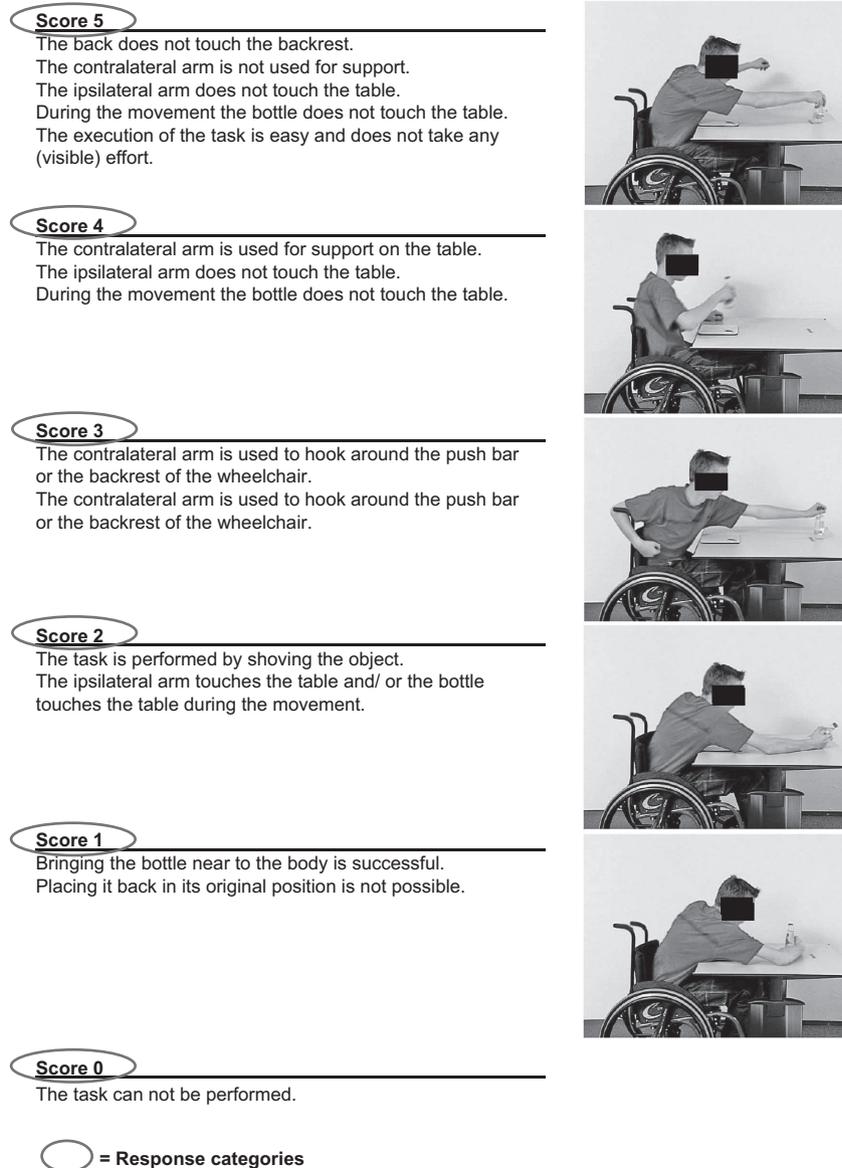


Figure 1 Example of VLT-SF test item.

attributes like age and gender. In case of the VLT-SF, a Rasch analysis¹⁰ can be used to mathematically model the relationship between a C-SCI person's ability and VLT-SF test item difficulty as a function of response probabilities. This model can be used to verify the following:

1. order of response categories (Do successive response categories of each item represent increasing levels of a person's ability?);
2. unidimensionality (Does a (sub-)scale or variable solely represent a single property or attribute under investigation?);
3. reliability of the scale;
4. scale invariance (Is the difficulty of items stable in different subgroups?).

The aim of the study is to apply the Rasch model to the VLT-SF to upgrade its ordinal scale towards an interval level scale and to verify if the requirements of an objective measure are satisfied.

MATERIALS AND METHODS

Subjects

Data from a large prospective cohort study in SCI patients (www.scienn.nl) across eight Dutch Rehabilitation centres and a research project at Adelante Rehabilitation Centre called ToCUEST (Task-oriented Client-centred Upper Extremity Skill Training)¹¹ were used. Data stem from patients admitted for initial rehabilitation. Inclusion criteria were: complete or incomplete C-SCI (including lesion at T1), and age between 18 and 70 years. Exclusion criteria were: severe additional neurological, orthopaedic or rheumatologic disease, hampering the upper limb performance, and total inability to perform upper extremity measurements. Three months after the patients were able to sit in a wheelchair for at least 3 consecutive hours, VLT-SF data (of the least affected arm-hand) were collected.

Data analysis

A Rasch analysis, featuring the Partial Credit Model, was performed using RUMM2030 software. The Partial Credit Model was preferred, as it accommodates the use of scales with different response categories (number of response categories and/or content of one category) in which the location of

the thresholds (relative to the item difficulty) is not the same from one item to the other. Based on a (experimental) data set provided (in our case the VLT-SF test data set), the Rasch Partial Credit Model (A) mathematically determines the position of each response category of each test item along a (new) linearised interval scale (common to all test items), representing item difficulty, and (B) mathematically determines the position of the thresholds between adjacent response categories per test item, based on threshold position probability. As a mathematical consequence of (A) and (B), each response category of each test item is positioned relative to all other response categories of all test items. Once the observed responses are found to fit the unidimensionality requirement (by comparing the observed scores (responses given by the patients) with the expected scores (see also next paragraphs)), the Rasch model can be used to determine the level of arm and hand skilled performance of each patient and the difficulty of each item or threshold on a common linear interval scale.¹² The Rasch model uses a logistic transformation to convert the ordinal scores into linear measures expressed in 'logits' (that is, log-odds units). The logit is defined as the probability unit that expresses the natural logarithm of the odds of success (that is, pass to fail probability ratio of a subject to an item).¹³ A 'logit' is, in effect, the unit in which the new measure (= *r_VLT-SF*) is expressed in.

To evaluate whether the VLT-SF satisfies the requirements of an objective measurement, the following steps were performed:

The *order of response categories (and their thresholds)* were verified to determine whether the multiple response categories reflect an increasing continuum of arm and hand skilled performance. This means that subjects with a higher level of performance always select a higher response to any given item, and subjects selecting a higher response for a given item indeed present a higher level of performance.¹⁴ Categories not discriminated can be collapsed (= combined) with an adjacent category into a single category,¹⁵ thus necessitating the Partial Credit Model mentioned above. Two guidelines have been used to decide on collapsing/combining of categories.¹⁶ Firstly, the collapsing/combining should make sense from a clinical point of view, for which the opinion of four experts was gathered. Secondly, when the collapsing is clinically relevant, the response category frequencies were investigated (that is, number of responses per categories and shape of the frequency distribution).

The *unidimensionality* was assessed by χ^2 fit statistics and a principal component analysis on the residuals (that is, the differences between observed and expected responses). After dividing the total sample into three class intervals of patients with increasing ability, the item χ^2 fit statistic was computed as the sum of the squared standardized residuals of each of the three class interval.¹⁷ Items with a *P*-value <0.05 indicate a threat to the fit requirement.¹³ The principal component analysis on the residuals gives the percentage of variance attributable to the Rasch factor and the first residual

factor. Independent *t*-tests were used to compare the estimates for each subject, which were derived from the highest positive and negative loading items (correlated at 0.3 and above with the component) on the first principal component of the residuals. The scale is considered unidimensional when the percentage of tests outside the range of ± 1.96 is <5%.^{18,19}

The *reliability* of the scale, that is, the internal consistency was assessed using the Pearson's Separation Index (PSI).⁹ The PSI should be 0.70 for research purposes and 0.85 for clinical use.²⁰ The traditional Cronbach alpha coefficient was also used to assess the reliability.

The *invariance of item difficulty hierarchy* across groups of patients with either a motor complete or incomplete lesion (two subgroups) was tested using Differential Item Functioning (DIF) tests.⁹ If an item measures the same ability in the same way across groups then, except for random variations, the same success rate should be found, irrespective of the nature of the group. Items that give different success rates for two or more groups, at the same ability level, are said to display DIF. A two-way analysis of variance was computed on the standardized residuals.²¹ In case of a significant motor lesion completeness main effect, the relative difficulty of the item is different between the patients' subgroups (uniform DIF). A significant interaction effect between the patients' subgroups and the class intervals of patients with increasing ability indicates a non-uniform DIF.

RESULTS

Patient characteristics are given in Table 2.

Verification of the order of response categories

Although Rasch analysis of the original VLT-SF data indicated a good capacity of the scale to discriminate persons' ability levels (PSI 0.91), 7 out of 10 items demonstrated disordered response categories, that is, the order of response categories within a test item did not match successive increase of response difficulty. Therefore, the original six response categories were collapsed/combined into either three categories (range: 0–2) for four items, or four categories (range: 0–3) for six items. The resulting recoded VLT-SF response scale is shown in Table 1. Note that its maximum recoded (raw) score is 26 (= 4 items \times 2 and 6 items \times 3).

Verification of unidimensionality requirement

After recoding the response scale, and Rasch transformation of the data into logits (*r_VLT-SF*), all items showed ordered response categories and contributed to the definition of an unidimensional measure of arm–hand functioning (total and individual item χ^2 probabilities >0.05). The average item fit residual was equal to -0.22 (s.d. 1.33), indicating a slight overfit (that is, more proficient persons perform better and less proficient persons perform worse than expected by the model). After the principal component analysis, the percentage of individual *t*-tests outside the range of ± 1.96 (95% confidence interval)

Table 1 Recoded VLT-SF response scale and example of patient score

Original VLT-SF scale	0	1	2	3	4	5
Recoded VLT-SF response scale						
Item 1: forward reaching	0	1	1	2	2	3
Item 2: arch task	0	1	1	1	2	2
Item 3: opening thumb	0	1	2	2	3	3
Item 4: thumb grip	0	1	1	2	2	3
Item 5: strength thumb	0	1	1	1	2	2
Item 6: opening fingers	0	1	1	1	2	2
Item 7: strength fingers	0	1	1	1	2	2
Item 8: pen grip	0	1	1	2	3	3
Item 9: lighting match	0	1	2	2	3	3
Item 10: opening bottle	0	1	2	2	3	3

Abbreviation: VLT-SF, Van Lieshout hand function test for tetraplegia short form. The scores in bold indicate an example of a patient's scores: the score on the recoded VLT-SF is 6; the score on the original VLT-SF score is 7 (an item score of 2 of the original VLT-SF on the item 'arch task' e.g. is recoded to an item score 1 og the recoded VLT-SF. Note: The maximal attainable score on the recoded VLT-SF is 26.

Table 2 Patient characteristics

Patient characteristics	
<i>n</i>	73
Age (years) (mean (s.d.))	40 (14,5)
Days since injury (mean (s.d.))	196 (87)
Gender	53 Male; 20 female
Lesion level	45 High (C0–C6); 28 low (C7–T1)
	1 C0; 2 C4; 17 C5; 25 C6; 14 C7; 6 C8; 8 T1
Motor completeness	35 (AB); 38 (CD)

Abbreviations: AB, motor complete lesions according to ASIA Impairment Scale; CD, motor incomplete lesion according to ASIA Impairment Scale.

was 4%, which is significant ($P < 0.05$), indicating unidimensionality of the $r_{\text{VLT-SF}}$.¹⁸

Targeting of the $r_{\text{VLT-SF}}$ test relative to the patients' performance and verification of the scale reliability

The distributions of the persons' ability values and the items' threshold are displayed in figure 2. The average person's ability was 1.49 logits (s.d. 2.85), deviating from the average difficulty of the test that was arbitrarily set at 0.0 logits by the Rasch model. The difficulty of the test items are relatively well-targeted on the persons' ability range, although some gaps are present (for example, around 0 logits) (Figure 2). Persons' ability values cover a range of 11.1 logits (maximal person's ability – minimal person's ability, excluding extreme (floor and ceiling effect) values). A slight floor effect (4% of the sample), as well as a more pronounced ceiling effect (11% of the sample) exist. A PSI of 0.91 and a Cronbach alpha of 0.95 indicate that the $r_{\text{VLT-SF}}$ scale is reliable and allows for the patients to be discriminated.

The item difficulty hierarchy of each item relative to the other items in the total test difficulty continuum

In Table 3 the item difficulties are sorted, from top to bottom, in order of decreasing difficulty (range: from 1.341 to -1.110 logits). 'Thumb grip' was the item requiring the highest ability level to be performed, whereas 'Arch task' was the easiest item. Table 3 also reports the standard error associated with each item difficulty, as well as each corresponding χ^2 probability.

Description of the $r_{\text{VLT-SF}}$ scale

The definition and use of the $r_{\text{VLT-SF}}$ scale is depicted in Figure 3. The top panel shows the distribution of arm-hand performance measures of the patients across the entire $r_{\text{VLT-SF}}$ range. The bottom panel illustrates the s-shaped relationship between the recoded VLT-SF raw total scores (y axis) ranging from 0 to 26, and the arm-hand functioning linear measures expressed in logits (x axis) ranging from -4.78 to 6.32 logits. This relationship is approximately linear between total scores of -2 and 2 logits. Outside this central range, however, a unitary progression in total score accounts for an increasing number of $r_{\text{VLT-SF}}$ measures. In the central range, the change in $r_{\text{VLT-SF}}$ measures corresponding to an unitary increment in total score from, for example, 12 to 13 is equivalent to 0.24 logits. Outside this central range, it increases to 1.44 logits for the same

increment in total score from, for example, 25 to 26. This six-fold difference denotes the nonlinearity of the total score. The middle panel demonstrates the expected score to a given item as a function of the $r_{\text{VLT-SF}}$ arm-hand performance measure. By comparing the ability of a given patient to the difficulty of each item, it is possible to determine the expected score of the patient to that item. For example, a patient with a total score of 2 logits would be expected to obtain a score of 2 at the easiest item (item 2 arch task) and a score of 1 at item 5 (strength thumb), a more difficult item (see Figure 3).

Table 3 $r_{\text{VLT-SF}}$ calibration for patient with C-SCI

Item	Location	s.e.	Prob
Item 4: thumb grip	1.341	0.255	0.461
Item 10: opening bottle	0.649	0.213	0.589
Item 5: strength thumb	0.540	0.317	0.086
Item 8: pen grip	0.415	0.218	0.255
Item 1: forward reaching	0.007	0.208	0.280
Item 7: strength fingers	-0.123	0.256	0.406
Item 9: lighting match	-0.220	0.206	0.161
Item 6: opening fingers	-0.549	0.263	0.395
Item 3: opening thumb	-0.950	0.224	0.727
Item 2: arch task	-1.110	0.288	0.294

Abbreviations: C-SCI, cervical spinal cord injury; Prob, χ^2 probability; s.e., standard error.

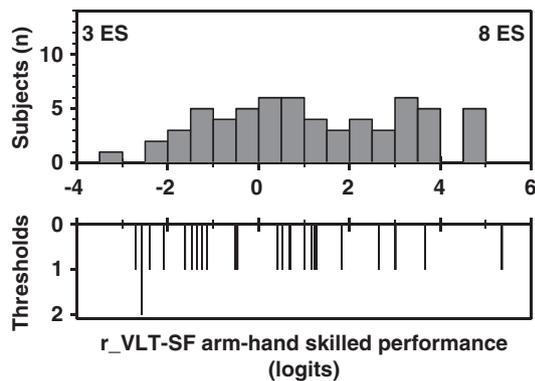


Figure 2 Distribution of persons' ability and response category difficulty. ES (Extreme subject, showing either a ceiling effect or a floor effect); n (number). Please note: in the bottom panel, each vertical line represents a threshold between adjacent response categories (difficulty) within each test item. Two threshold locations overlap. (See also Figure 3, middle panel).

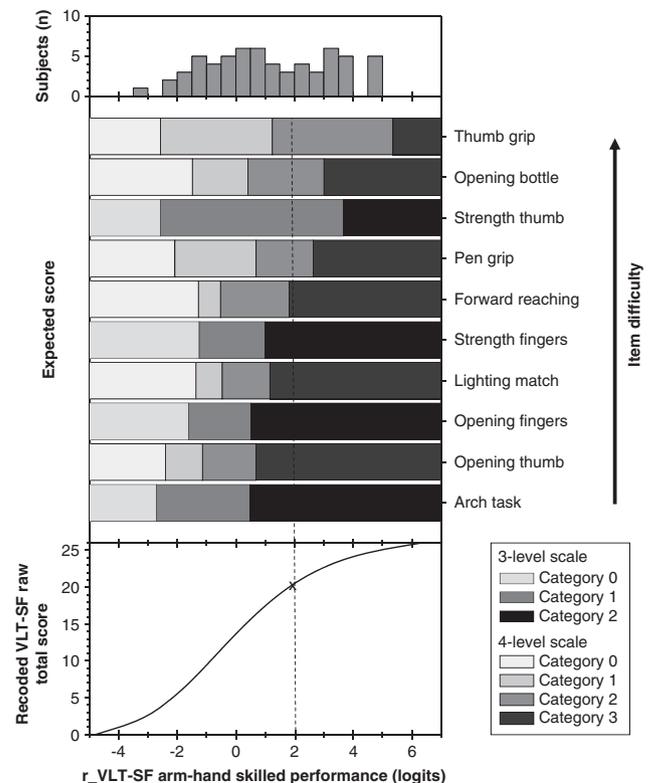


Figure 3 Distribution of arm-hand skilled performance scores of patients (top panel); expected response categories (and thresholds) within each test item in relation to the recoded VLT-SF in logits (middle panel); relationship between recoded VLT-SF raw total scores and $r_{\text{VLT-SF}}$ expressed in logits (bottom panel). n (number); X (hypothetical patient scoring 2 logits on the $r_{\text{VLT-SF}}$); dotted line (expected scores on each test item); colour coding in middle panel (order of response categories per test item). A full colour version of this figure is available at the *Spinal Cord* journal online.

Verification of the invariance scale requirements

The invariance of r_VLT-SF was validated in our sample across motor complete or incomplete lesions, as neither uniform nor non-uniform DIF was detected (using Bonferroni adjustment). This means that the difficulty hierarchy of the VLT-SF activities was invariant across patients' with motor complete or incomplete lesions. Therefore, r_VLT-SF can be used to measure arm–hand skilled performance in patients with C-SCI, regardless the motor completeness of the lesion.

DISCUSSION

The purpose of the study was to apply the Rasch measurement model to the VLT-SF to verify whether the requirements of an objective measurement were satisfied.

The Rasch analysis of the original VLT-SF indicated that although the overall power of fit was good, there were some shortcomings regarding the ordering of the thresholds and the unidimensionality of the scale. As the number of response categories may be largely responsible for disordered thresholds, it was decided to reduce the number of response categories from six to three or four. The problem that may arise by reducing the response categories is that the scale loses the ability to discriminate between persons. However, the original PSI of 0.91 did not change after the collapsing of response categories, indicating that the scale allows for patients to be discriminated. Another option to obtain ordered response categories and unidimensionality would be to discard items of the original VLT-SF. However, disordered thresholds were present in most items (7 out of 10), and Rasch analysis of the original VLT-SF demonstrated that person–item distribution was well targeted and that the items were locally independent. As discarding items would influence the comprehensiveness with which the VLT-SF scale covers the clinically relevant areas, the latter option was not applied. The reducing and recoding of the response categories from six to three or four led to ordered thresholds and improved the item fit to the model. The same observation was made in a study of Tennant *et al.*²² investigating the response categories of the Functional Independence Measure. In r_VLT-SF, both total and individual item χ^2 probabilities were higher than 0.05, indicating that the scale and all items meet the criteria for unidimensionality. This is confirmed by the results of the principal component analysis on the residuals. The misfit observed in the original VLT-SF scale was probably caused by the inability of clinicians to discriminate the original six response categories. Actually, when two categories are indistinguishable for assessors, there is no threshold between them. However, when assessors try to estimate a hypothetical one, this may introduce bias, which is highlighted by the fit statistics.²³

Considering the advantage to recode the VLT-SF response categories into the r_VLT-SF ones, one might want to use only the reduced response categories for clinical purposes. However, clinicians indicate that besides the purpose of the upper extremity assessment, the different response categories of the original version are also a relevant tool to guide them in their clinical reasoning and therapy planning. Therefore, it was decided to not replace the VLT-SF by the r_VLT-SF for daily practice, but to provide a recalculation table (software package) for the therapists (see also (www.vlt-test.eu)). This will enable the clinicians to transform the original item scores into r_VLT item scores. The latter allows interval level measures to be generated, from which it is possible to derive the expected individual item responses, as it is presented in Figure 3.

As persons with motor complete and incomplete lesions may have different levels of arm–hand skilled performance, it was necessary to

investigate whether the different items are invariable, regardless of the completeness of the lesion. The Rasch analysis has demonstrated that the item difficulty hierarchy of the r_VLT items does not change according to the motor completeness of the lesion. This is in accordance with a previous study showing that the VLT can be used both in persons with motor complete and incomplete lesions.⁶ However, our small sample size may have failed to detect DIF. According to Scott *et al.*,²⁴ a sample size of 200 is required to detect DIF in a scale with more than two items, with adequate power (>80%).

Performing a Rasch analysis on the VLT-SF provides additional clinical value to the instrument. It proved to be a unique high precision instrument (PSI = 0.91; Cronbach alpha = 0.95) to quantify the quality of arm–hand skill performance in persons with C-SCI, supporting the clinicians in assessment and prediction of outcome. In practice, the conversion of the VLT-SF ordinal total scores into linear measures r_VLT-SF will permit a better quantification of individual patient's progress regarding arm–hand skill performance. The hierarchical nature of the Rasch-improved scale can be used to identify patients' pattern of improvement, given the current arm–hand skill performance measurement. Furthermore, the r_VLT-SF scale has the ability to detect discrepancies between the observed score to each item and the score expected by the model, given the overall patients' measure. For instance, it will be possible to identify on which items the person's performance lags behind expectations. This in turn may lead to a more client-centered approach, in which therapy will be customised to the patient's specific needs.

The Rasch investigation and calibration of the r_VLT-SF scale were based on data from the assessment performed 3 months after the start of the active rehabilitation. This assessment time was chosen because: (1) a greater number of data were available and (2) this moment was considered the most representative in the rehabilitation process by omitting major floor or ceiling effects in patients who were either unable to perform any task (at the start of the rehabilitation) or patients who obtained the maximum score (at the end of the rehabilitation). Maximum and minimum total scores (that is, floor and ceiling scores) are per definition (statistically) discarded from any Rasch procedures, as they do not contribute to the sensitivity of the scaling.⁹ However, some floor and ceiling effects are present even at 3 months after the start of active rehabilitation (see Figure 2). This may be due to the fact that data of the best hand were used. Gathering data of both hands may be more representative of patients' actual arm–hand ability. This, however, should be investigated in future research.

As clinicians are generally unfamiliar with the concept of logits, transforming the linear measures expressed in logits into 0–100 scores may improve the clinical application of the VLT-SF scale and the communication amongst its users. Moreover, in order to improve the prediction of arm–hand skilled performance outcome with regard to the extent of the lesion and the time since injury, it is essential to obtain reference values. The accompanying paper by Spooren *et al.*²⁵ describes the transformation from linear measures expressed in logits into 0–100 total scores and provides reference values of r_VLT-SF.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The project was funded by ZonMw, the Netherlands Organization for Health Research and Development (Grant 335020023), The Hague, the Netherlands.

- 1 Snoek GJ, IJzerman MJ, Hermens HJ, Maxwell D, Biering-Sorensen F. Survey of the needs of patients with spinal cord injury: impact and priority for improvement in hand function in tetraplegics. *Spinal Cord* 2004; **42**: 526–532.
- 2 Spooren AIF, Janssen-Potten YJM, Snoek GJ, IJzerman MJ, Kerckhofs E, Seelen H. Rehabilitation outcome of upper extremity skilled performance in persons with cervical spinal cord injuries. *J Rehabil Med* 2008; **40**: 637–644.
- 3 World Health Organization. *ICF: International Classification of Functioning, Disability and Health*. Geneva, 2001.
- 4 Post MWM. Pilot-onderzoek Van Lieshout Test Hoensbroek: iRv (in Dutch) 2000.
- 5 Post MW, Van Lieshout G, Seelen HA, Snoek GJ, IJzerman MJ, Pons C. Measurement properties of the short version of the Van Lieshout test for arm/hand function of persons with tetraplegia after spinal cord injury. *Spinal Cord* 2006; **44**: 763–771.
- 6 Spooren AI, Janssen-Potten YJ, Post MW, Kerckhofs E, Nene A, Seelen HA. Measuring change in arm hand skilled performance in persons with a cervical spinal cord injury: responsiveness of the Van Lieshout Test. *Spinal Cord* 2006; **44**: 772–779.
- 7 Merbitz C, LMorris J, Grip JC. Ordinal scales and foundation of misinference. *Arch Phys Med Rehabil* 1989; **70**: 308–312.
- 8 Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989; **70**: 857–860.
- 9 Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd edn Routledge, Taylor & Francis group: Abingdon, 2007.
- 10 Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Messa Press: Chicago, 1980.
- 11 Spooren AIF, Janssen-Potten YJM, Kerckhofs E, Seelen HAM. Evaluation of a task-oriented client-centred training program to improve arm hand skilled performance in cervical spinal cord injury. *Spinal Cord* 2011; **49**: 1049–1054.
- 12 Rasch G. *Probabilistic Models for Some Intelligence and Attainment Test*. Danish Institutions for Educational Research: Copenhagen, 1960.
- 13 Caty GD, Arnould C, Thonnard JL, Lejeune TM. ABILOCO-Kids: a Rasch-built 10-item questionnaire for assessing locomotion ability in children with cerebral palsy. *J Rehabil Med* 2008; **40**: 823–830.
- 14 Arnould C, Penta M, Renders A, Thonnard JL. ABILHAND-Kids: a measure of manual ability in children with cerebral palsy. *Neurology* 2004; **63**: 1045–1052.
- 15 Tennant A, Conaghan AG. 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 Rheum* 2007; **57**: 1358–1362.
- 16 Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002; **3**: 85–106.
- 17 Andrich D. *Rasch Models for Measurement*. Sage Publications: London, 1988.
- 18 Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002; **3**: 206–231.
- 19 Boufflioux E, Arnould C, Thonnard JL. SATIS-Stroke: a satisfaction measure of activities and participation in the actual environment experienced by patients with chronic stroke. *J Rehabil Med* 2008; **40**: 836–843.
- 20 Fisher W. Reliability statistics. *Rasch Meas Transact* 1992; **6**: 238.
- 21 Andrich D, Sheridan BE, Luo G. *Rasch Unidimensional Measurement Models (RUMM): A Windows Based Computer Program*. Murdoch University: Perth, 2004.
- 22 Tennant A, Penta M, Tesio L. Structure of a combination of Functional Independence Measure and Instrumental Activity Measure items in community-living persons: a study of individuals with cerebral palsy and spina bifida. *Arch Phys Med Rehabil* 1996; **77**: 11.
- 23 Andrich D. Category ordering and their utility. *Rasch Meas Transact* 1996; **9**: 464–465.
- 24 Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M *et al*. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *J Clin Epidemiol* 2009; **623**: 288–295.
- 25 Spooren AIF, Arnould C, Smeets RJEM, Snoek G, Seelen HAM. Reference values for the transformed Van Lieshout handfunction test for Tetraplegia. *Spinal Cord* 2013, (this issue).