JOURNAL OF NEUROTRAUMA 35:854–863 (March 15, 2018)

DOI: 10.1089/neu.2017.5195

# The Graded and Redefined Assessment of Strength, Sensibility, and Prehension Version 2 Provides Interval Measure Properties

Inge-Marie Velstra,[1] Carolina Fellinghauer,[2,3] Rainer Abel,[4] Sukhvinder Kalsi-Ryan,[5,6] Rüdiger Rupp,[7] and Armin Curt[8]




## Abstract

The Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) is a valid, reliable, and responsive outcome measure to evaluate upper limb function in individuals with tetraplegia. GRASSP generates ordinal total scores; therefore, applicability as an interval level measurement requires testing of its measurement properties. This study examined the metric characteristics with Rasch Analysis to derive interval level scales of the respective GRASSP subtests. The GRASSP was recorded within 10 days, and at 1, 3, 6, and 12 months after cervical spinal cord injury (SCI). Rasch analysis was performed for each GRASSP subscale to assess the following metric assumptions: absence of local item dependency (LID), unidimensionality, monotonicity, item and model fit, reliability, and absence of differential item functioning (DIF) for side (left and right) and examination stage. If these assumptions could not be met, adjustments were undertaken to achieve a good fit to the Rasch model. Seventy-seven individuals with cervical SCI were included ($n = 154$ arms). Stacking the data for the side (left and right) resulted in a total of 614 observations, which were based on the repeated measurements. With minor adjustments, the GRASSP subscales showed good reliability, item fit, and ordered response options. Local item dependencies were found in the strength and sensibility subscales. Redundancies among some measurement items allowed shortening of the subscales without reasonable loss of reliability. Absence of DIF for the examination stage supported robustness of the subscales over time. The modified GRASSP, now Version 2, subtest scores can be applied as interval level measurements, and the reduction of items within subscales allows for shorter assessment times in clinical studies without degrading metric properties.


**Keywords:** interval measurement; Rasch; SCI; tetraplegia; upper limb function

## Introduction

Individuals with cervical spinal cord injury (SCI) are a heterogeneous population[1–4] characterized by different levels of cervical cord lesion (e.g., low vs. high levels of lesion), measures of neurological deficit (such as motor and sensory scores), as well as distinct degrees of impairment (e.g., incomplete vs. complete tetraplegia). Therefore, it is important that upper limb outcome measures applied in this population are sensitive to different levels of functioning in individuals; that is, covering a broad spectrum from devastating impairment to high levels of functional independence.

The Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) is a clinical composite measurement tool assessing upper limb function in distinct domains (e.g., neurological and functional scores) in individuals with cervical SCI. Application of the GRASSP is increasing for clinical purposes and as an outcome measurement in research.[1,2,5–8]

The GRASSP version 1.0, consists of four subtests, including the domains Strength (GR-str), Sensation (GR-sens), Prehension Ability (GR-pa), and Prehension Performance (Gr-pp). All domains can be analyzed separately[1] or used together. The GRASSP has been tested to show excellent psychometric properties[1,5,8] with a classical test theory approach.[9] Metric analyses of the GRASSP

[1]Clinical Trial Unit, Swiss Paraplegic Center, Nottwil, Switzerland.
[2]Swiss Paraplegic Research (SPF), ICF Unit, Nottwil, Switzerland.
[3]Department of Health Sciences and Health Policy, University of Lucerne, Lucerne, Switzerland.
[4]Spinal Cord Injury Center, Hohe Warte, Bayreuth, Germany.
[5]Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada.
[6]Department of Physical Therapy, University of Toronto, Toronto, Ontario, Canada.
[7]Klinik für Paraplegiologie, Universitätsklinikum Heidelberg, Heidelberg, Germany.
[8]Spinal Cord Injury Center, Balgrist University Hospital, Zurich, Switzerland.

are not available to date. The GRASSP takes between 35 and 50 min to administer. Because of the combined assessment of neurological (strength and sensation) and functional outcomes (prehension), the GRASSP provides a comprehensive appreciation of upper limb function. However, the collection of different outcome measures in daily practice is challenged by time constraints. Therefore, there is a need to investigate if subtests can be reduced, to shorten the administration time of the GRASSP without jeopardizing sensitivity and metric properties.

The literature recommends[1] using the GRASSP in conjunction with other standard SCI classification or measurement tools such as the International Standards of Neurological Classification of Spinal Cord Injury[10] (ISNCSCI) and the Spinal Cord Independence Measure[11] (SCIM). The combination of these outcome tools broadens insight into the severity and burden of the disease as well as detection of subtle and clinical meaningful changes in upper limb function and independence from the very acute phase to 1 year post-injury.

Although all clinical scores so far applied in SCI (GRASSP, ISNCSCI, and SCIM) have ordered response categories, the ordinal raw scores, or sum of the items, of these measures are often used as if they are interval in nature. Sum scores assume that all response units are interval scaled; however, the equal spacing of the intervals is an assumption that needs to be tested.[12–17] A valid scale for measurement requires its items to be unidimensional and to measure a single common latent trait.[18–20] Applying Rasch analysis[21] is one way to develop a scale that fulfills the requirements for a measurement scale with good internal construct validity and reliability.

The four available GRASSP subtest sum scores may be of limited use, and only suitable for nonparametric analysis, if interval scale properties are not supported by the analysis with the Rasch model. Further, if the four subscales do not fulfill the assumptions of the Rasch model, parametric analyses of strength, sensation, and prehension that change over time could produce misleading results. Therefore, to gain knowledge about the measurement properties of GRASSP, the objectives of this study were to examine its metric characteristics with Rasch analysis, in order to obtain interval level scales for the GRASSP subtests and to identify the most relevant items for a reduced assessment of the domains.

## Methods

### Study population and design

Psychometric study was performed using data from a European prospective cohort study in individuals with cervical SCI. The data were prospectively collected between 2009 and 2011 at three SCI-specific clinics in Germany and two SCI-specific clinics in Switzerland, between 0 and 10 days, at 1 month (range 16–40 days), 3 months (range 70–98 days), 6 months (range 150–186 days) and 12 months (range 300–400 days) after SCI onset.

Inclusion criteria were traumatic or nontraumatic cervical SCI, with an American Spinal Injury Association (ASIA) Impairment Scale (AIS) grade A,B,C, or D, defined according to the ISNCSCI,[10] and age >16 years.

Exclusion criteria were other severe neurological (e.g., traumatic brain injury) or medical disorders, a diagnosis of dementia, or psychiatric disorders.

The study protocol and informed consent forms were approved by the ethics committee at all sites. Written informed consent was obtained from all participants prior to participation.

### Procedure and assessment

GRASSP assessments were performed by occupational therapists with at least 1 year of experience in working with individuals with SCI. They were trained at an investigators' workshop and received a GRASSP instruction manual.

### The GRASSP

According to Kalsi-Ryan and colleagues,[5] the GRASSP showed excellent bilateral subtest reliability and moderate to substantial concurrent validity with other functional scales such as the SCIM, SCIM Self-Care Subscale (SCIM-SS), and Capabilities of Upper Extremity Questionnaire (CUE) in individuals with chronic cervical SCI. Velstra and colleagues[1] found that the responsiveness of GRASSP was excellent (standardized response mean [SRM]) ranged from 0.79 to 1.48 for GR-str; from 0.14 to 0.64 for GR-sens, and from 0.50 to 1.03 for GR-prehension total (GR-pa and GR-pp) between all examination time points immediately after SCI onset up to 1 year post cervical SCI.

The GRASSP, version 1.0 (more details are described elsewhere[1]) includes 25 bilateral items grouped into four subtests:

The GR-str subtest counts 10 items (muscles) in the upper limb and is assessed with manual muscle testing on both sides. The items are rated from 0 to 5, with 5 indicating that the muscle has normal strength. The reliability of Gr-str is supported by Intraclass Correlation Coefficients (ICCs) 0.95–0.98,[5] and a correlation coefficient between 0.59 and 0.76 for the concurrent validity.[5]

The GR-sens subtest items measure the GRASSP palmar sensation (GR-ps) and GRASSP dorsal sensation (GR-ds) of three fingers (1: thumb; 2: middle finger; and 3: little finger) of the right and left hand, respectively. Sensation is assessed with Semmes and Weinstein monofilament testing. The measures are rated from 0 to 4, where 4 indicates a perfect sensation of the measurement point (normal sensation). The reliability of Gr-sens is supported by ICCs 0.84–0.95,[5] and a correlation coefficient between 0.57 and 0.77 for the concurrent validity.[5]

The GR-pa subtest assesses three grasp patterns: cylindrical grasp, lateral key pinch, and tip to tip pinch. The GR-pa counts three items that are rated from 0 to 4, with 4 indicating normal voluntary control of wrist and digits when generating the grasp. The reliability of GR-pa is supported by ICCs 0.95–0.98.[5]

The GR-pp subtest assesses six standardized functional tasks: pouring water from a bottle into a glass, unscrewing lids from jars, performing a pegboard task, using a key, manipulating coins, and placing nuts onto screws in both sides. The items are rated from 0 to 5, with 5 indicating that solving the task with the expected grasp has succeeded.

The reliability of GR-pp is supported by ICCs between 0.93 and 0.96.[5] GR-prehension total (GR-pa and GR-pp) also showed good concurrent validity with correlation coefficients between 0.68 and 0.83.[5]

### Statistical analysis

Earlier studies of the GRASSP do not support statistically significant differences between the left and right body sides;[1,2,7] therefore, the data were put in a long format aggregating the response options for the left and right body side into one variable. This does not mean that participants must have the same ability on both sides, but that the scale works similarly for the left and the right body sides. Also, the items were aggregated over all five measurement time points (within 10 days, 1 month, 3 months, 6 months, and 12 months post-injury) assuming an equal item difficulty hierarchy for all time points. Item invariance for the laterality and the time point was tested with a differential item functioning (DIF) analysis (see next subsection DIF) to assure the robustness of the scale to these personal and contextual factors.

### Metric analysis

The Rasch model is a probabilistic measurement model that determines if the items and the total score of a scale fulfill the essential assumptions for an interval scale. The Rasch model assumes that responses to items on a scale are a function of the person's ability (i.e., the individual's level of functioning), and the difficulty of an item. Persons can be ordered from least able to most able based on their ability, which is entirely determined by the points scored on all the items within a subtest. Similarly, subtest items can be ordered from the least to the most difficult. In that sense, the Rasch analysis expects to find a certain pattern in the data, which reflects the relationship between a person's ability level and the item difficulty on the probability of a certain response. The lower the ability of the person, the more likely it will be that an item will be considered by that person to be difficult. The residuals can be derived as the deviation of the observed from the expected responses under the Rasch model. Patterns within the residual matrix are indicative of issues within the scale (e.g., local item dependency [LID], multidimensionality, or DIF) that need to be solved.

Rasch analysis[21] was performed using the partial credit model (PCM) for polytomous ordered response options. The choice for a one parameter probabilistic model, such as the PCM, instead of a classical test theory approach, was driven by the interest in gaining knowledge about the difficulty of the GRASSP items and in obtaining person-ability parameter estimates usable in clinical practice, epidemiology, and research. Also, only one parameter models, from the family of the Rasch models, have the desirable property of score sufficiency, which allows deriving person-ability estimates directly from a person's total score if the assumptions for measurement are fulfilled. The PCM, contrary to another one parameter model, named the ''rating scale model,'' does not expect the response thresholds to be equidistant across items of a metric.[22]

When doing a Rasch analysis, important assumptions are tested that allow one to determine if an instrument fulfills fundamental measurement properties: monotonicity of the items' response options and their fit (1), absence of LID (2), the metric's unidimensionality (3), absence of DIF (4), targeting and reliability (5).[23] If the data fit the Rasch model, fulfilling its assumptions, the score can be expected to be valid for measurement and further parametric statistical analysis. If these assumptions cannot be met, adjustments have to be undertaken to achieve good fit to the Rasch model.

### *Monotonicity and item fit (1)*

Thresholds are the equal probability points between response categories of an item. Item response categories are expected to show ordered thresholds, where the number assigned to a response is a representative level for strictly increasing difficulty. In the presence of disordered thresholds, the response categories are collapsed until monotonic ordering is achieved.

Furthermore, the items of the scale must show a good fit to the model. The item fit statistics, the infit and outfit, also known as weighted and unweighted item mean squares (MnSq)[24] allow identification of the items that do not fit the Rasch model and would thus compromise the scale's construct validity. Ideally, good fitting items should present MnSq values close to 1. As the GRASSP is a clinical tool, MnSq between 0.5 and 1.7 were considered acceptable.[25] Values not within this range might be measuring a different construct or need further clarification or revision to fit the Rasch model.

### *LID (2)*

LID indicates response and trait dependencies in the assessment. LID can be determined by observing the correlation matrix of the standardized residuals. Positive correlations between item residuals with $r > 0.2$ were considered LID. Response dependencies result when items address similar aspects of the measured trait. LID may statistically bias the parameter and inflate the reliability estimates.[26] Another type of dependency is trait dependency, which is characterized by negative residual correlations, and suggests multidimensionality in the metric. Response dependency can be solved by aggregating the associated items to testlets,[27] or by deleting one of the dependent items.

### *Unidimensionality (3)*

Unidimensionality testing investigates the requirement of a measure to assess levels of difficulty in only one single latent trait (such as GR-str or GR-sens) and in several relatively independent latent constructs. The unidimensionality of the GRASSP was determined with a principal component analysis (PCA) of the residuals. The loading structure should not indicate item residuals loading strongly on different dimensions. Second eigenvalues <1.4 are expected to support the unidimensionality of the scale.[28]

### *DIF (4)*

The analysis for DIF determines the invariance of item residuals for sample characteristics with an analysis of variance (ANOVA). In this study, the GRASSP items were tested for invariance for the laterality and measurement time points. Absence of DIF for time points indicates the robustness of the scale's hierarchy of item difficulties over time, and not the stability of the participant's level of ability across time. Also, absence of DIF for the laterality assumes that the items' difficulty hierarchy is similar on both sides, but not that participants must have the same ability on the left and the right side. DIF tests the invariance at the population and not at the individual levels.

F-tests were performed on the item residuals of each respective subtest; namely, an F-test for measurement time point and laterality, respectively (Uniform DIF), an F-test for the ability level also called class-interval DIF, and the interaction of the sample characteristic and the ability level (non-uniform DIF).[29] The F-test significance level was Bonferroni corrected for repeated measurement ($\alpha = 0.05/2$ sample characteristics $\times$ 3 F-tests $\times$ number of items).[30]

### *Targeting and reliability (5)*

Finally, the targeting of each of the subtests was observed.[31] The targeting determines the difficulty of a subtest for a study population by comparing the mean person-ability with the mean item difficulty. A perfectly well-targeted subtest would have the mean item difficulty and mean person-ability close to zero. The match between difficulty of GRASSP subtest items and ability was assessed, and possible floor and ceiling effects were determined. Commonly, <15% of the participants should have higher ability than the most difficult GRASSP item in a subtest (ceiling effect). Conversely, <15% of the participants should have lower ability than the least difficult GRASSP item in a subtest (floor effect); that is, have fewer difficulties than the subtest can measure.[32] The analysis for floor and ceiling effects excluded participants with extreme values, zero, or maximum possible scores, which were also not included in the item difficulty estimations and were interpolated when determining the person-abilities.[22]

The reliability of a subtest was verified using the person separation index (PSI) reliability statistic. The interpretation of the PSI is similar to Cronbach $\alpha$[33]; PSI >0.7 is required for group use and PSI >0.85 is required for individual use.[23]

Each subtest of GRASSP was analyzed separately with a PCM analysis. Beyond testing GRASSP's fitness for measurement, one aim was to develop and test reduced versions of some of the longer GRASSP's subtests. In that sense, an informed selection of items would be undertaken in order to keep only the most relevant items, without significant loss of reliability. Items were entered and removed iteratively to analyze the changes and possible improvements in the correlational and dimensional structure as well as the changes in the item and model fit.[23] In that sense, the decision about which items to keep and which to discard was based on the results

Table 1. Demographics and Clinical Characteristics of Participants

| | Examination stage | | | | | |
|---|---|---|---|---|---|---|
| | *Overall* | *1* | *2* | *3* | *4* | *5* |
| Sample size | | | | | | |
| *n* | 77 | 43 | 77 | 72 | 58 | 57 |
| Age: Mean (SD) | | | | | | |
| At admission | 50.61 (20.24) | 54.03 (20.72) | 50.61 (20.24) | 50.17 (19.93) | 46.40 (19.36) | 45.13 (18.67) |
| Sex (%) | | | | | | |
| Male | 52 (67.5%) | 29 (67.4%) | 52 (67.5%) | 50 (69.4%) | 44 (75.9%) | 41 (71.9%) |
| Cause of injury (%) | | | | | | |
| Traumatic | 72 (93.5%) | 39 (90.7%) | 72 (93.5%) | 67 (93.1%) | 53 (91.4%) | 55 (96.5%) |
| Center location (%) | | | | | | |
| Balgrist University Hospital Zurich (CH) | 14 (18.2%) | 11 (25.6%) | 14 (18.2%) | 13 (18.1%) | 13 (22.4%) | 12 (21.1%) |
| Swiss Paraplegic Centre Nottwil (CH) | 25 (32.5%) | 11 (25.6%) | 25 (32.5%) | 25 (34.7%) | 25 (43.1%) | 23 (40.4%) |
| Unfallklinik Murnau (D) | 1 (1.3%) | 0 (0.0%) | 1 (1.3%) | 1 (1.4%) | 1 (1.7%) | 1 (1.8%) |
| Klinik Hohe Warte Bayreuth (D) | 26 (33.8%) | 13 (30.2%) | 26 (33.8%) | 24 (33.3%) | 18 (31.0%) | 19 (33.3%) |
| Orthopädische Universitätsklinik Heidelberg (D) | 11 (14.3%) | 8 (18.6%) | 11 (14.3%) | 9 (12.5%) | 1 ( 1.7%) | 2 ( 3.5%) |

CH, Switzerland; D, Germany; Examination stage 1, 0–10 days post-injury; Examination stage 2, 16–40 days post -injury; Examination stage 3, 70–98 days post-injury; Examination stage 4, 150–186 post-injury; Examination stage 5, 300–400 days post-injury; %, percent; SD, standard deviation.



of the Rasch analysis and, to a large extent, on clinical considerations as well as published findings.[1,7]

Descriptive analyses were performed with R 3.3.0.[34] Rasch analyses were conducted with the R-package eRm.[35]

## Results

Seventy seven individuals ($n = 154$ arms) with acute tetraplegia entered the study. Stacking the data for the side (left and right) resulted in total of 614 observations, which were based on the repeated measurements (five time points with dropouts). Detailed sample characteristics are shown in Table 1.

The GRASSP subtests did not show any DIF for the upper limb body side and examination stage immediately after onset up to 1 year post-SCI. In general, the subtests of the GRASSP respond well to the Rasch analysis and support fundamental measurement properties, being applicable as an interval level scaled outcome measure of upper limb function at group (PSIs >0.8) and individual levels (PSI >0.9) in tetraplegia. However, to achieve good fit, adjustments, such as collapsing of response categories and deletion of items, were necessary. The fit to the model are shown in Table 2, and detailed fit statistics for the items of the subtests are shown in Table 3.

As the aim was also to reduce the administration time of the GRASSP subtests, Tables 2 and 3 show the most relevant analysis results of the start model with all items and the best-fitting model with all items if possible, and, if available, the best-fitting reduced model.

Table 4 shows the collapsing strategy of the ordinal GRASSP subtest. Table 5 shows the transformation table, which allows translation of the total scores of the corresponding GRASSP subtest scale, taking into account the collapsing strategy of the rating scale, into a user-friendly and unbiased interval-scaled 0–100 score. Tables S1–S4 show the descriptive statistics of the four GRASSP subtests for all time points (see online supplementary material at http://www.liebertpub.com).

The GR-pa subtest was the only subtest that did not require any adjustment to fit the Rasch model. The PSI of 0.995 indicated high reliability for individual measurement. The GR-pa subtest presented good subscale targeting with no floor and ceiling effects, no

Table 2. Fit Statistics[a]

| *GRASSP subtests* | *Stage* | *Number of items* | *Item difficulty* | | *Person ability* | | *Reliability* | | *Floor (%)* | *Ceiling (%)* | *LID Yes/No* | *DIF Yes/No* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *Mean* | *SD* | *Mean* | *SD* | *PSI* | *Cronbach α* | | | | |
| GR-str | Start | 10 | 0.234 | 1.155 | 0.427 | 1.433 | 0.943 | 0.948 | 1.21% | 5.37% | Yes | No |
| | Final reduced | 4 | 0.332 | 1.296 | –0.058 | 1.647 | 0.824 | 0.858 | 0% | 6.70% | No | No |
| GR-sens | Start | 6 | 0.119 | 0.86 | 0.561 | 1.263 | 0.909 | 0.919 | 3.16% | 9.49% | Yes | No |
| | Final (ds) | 3 | 0.343 | 1.168 | 0.535 | 1.276 | 0.819 | 0.835 | 7.67% | 13.77% | No | No |
| | Final (ps) | 3 | 0.208 | 0.827 | 0.431 | 1.078 | 0.801 | 0.832 | 6.51% | 16.67% | No | No |
| GR-pp | Start | 6 | 0.232 | 1.802 | 0.845 | 1.762 | 0.956 | 0.971 | 0% | 0% | No | No |
| | Final all items | 6 | 0.597 | 1.942 | 0.814 | 2.336 | 0.956 | 0.972 | 0% | 0% | No | No |
| | Final reduced | 4 | 0.664 | 2.279 | 0.87 | 2.322 | 0.937 | 0.955 | 0% | 0% | No | No |
| GR-pa | Start and Final | 3 | 1.643 | 3.998 | 1.667 | 3.828 | 0.955 | 0.972 | 0% | 0% | No | No |

[a]Fit statistics include the item and person targeting (without the extremes), the reliability of the models as well as presence of floor or ceiling effects, local item dependencies (LID), and differential item functioning (DIF) before and after remedies were applied to each of the GRASSP subtests.

GR, GRASSP, Graded and Redefined Asssement of Strength, Sensibility and Prehension; PSI, person separation index; GR-str, GRASSP subtest strength; GR-sens, GRASSP subtest sensation; ps, palmar sensation; ds, dorsal sensation; GR-pp, GRASSP subtest prehension performance; GR-pa, GRASSP subtest prehension ability; %, percent; SD, standard deviation.

Table 3. Detailed item fit statistics[a]

| *Subtest* | *Stage* | *Item* | *Collapsing* | *Location* | *Thres.1* | *Thres.2* | *Thres.3* | *Thres.4* | *Thres.5* | *Disordered thresholds* | *Outfit* | *Infit* | *LID* | *Eigenvalue* | *PC1* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GR-str | Start | Biceps | 012345 | −1.26 | −3.30 | 0.35 | −1.57 | −1.56 | −0.19 | yes | *2.59* | *1.69* | WristExt | 4.12 | −0.41 |
| | | Delto | 012345 | −0.74 | −2.54 | −0.36 | −0.80 | −0.99 | 0.97 | yes | *2.32* | *1.79* | Biceps | 1.12 | −0.42 |
| | | WristExt | 012345 | −0.57 | −1.36 | −0.54 | −0.88 | −0.60 | 0.52 | yes | 0.87 | 0.82 | Delto | 1.08 | −0.32 |
| | | Triceps | 012345 | 0.19 | −0.41 | 0.01 | −0.25 | 0.02 | 1.57 | yes | 0.80 | 0.72 | | 0.89 | −0.14 |
| | | ExtDig | 012345 | 0.85 | 0.44 | 0.58 | −0.07 | 0.87 | 2.43 | yes | 0.59 | 0.57 | FDP | 0.69 | 0.16 |
| | | Opp | 012345 | 0.68 | 0.65 | 0.23 | 0.47 | 0.29 | 1.76 | yes | 0.57 | 0.56 | AbdDigII | 0.67 | 0.31 |
| | | FPL | 012345 | 0.47 | 0.70 | 0.52 | 0.22 | −0.36 | 1.28 | yes | 0.63 | 0.56 | Opp | 0.54 | 0.28 |
| | | FDP | 012345 | 0.51 | 0.79 | 0.46 | 0.03 | −0.16 | 1.42 | yes | 0.52 | 0.51 | FPL | 0.41 | 0.30 |
| | | AbdDigV | 012345 | 1.13 | 0.79 | 0.36 | 0.47 | 1.14 | 2.90 | yes | 0.62 | 0.67 | FDP | 0.28 | 0.34 |
| | | AbdDigII | 012345 | 1.08 | 0.58 | 0.98 | 0.06 | 1.17 | 2.62 | yes | 0.66 | 0.71 | AbdDigV | 0.21 | 0.36 |
| | Final reduced | Delto | 001112 | −0.72 | −2.39 | 0.96 | | | | no | *1.28* | *1.28* | | 1.91 | −0.56 |
| | | Triceps | 001123 | 0.26 | −0.80 | 0.08 | 1.50 | | | no | 0.56 | 0.59 | | 1.09 | −0.39 |
| | | ExtDig | 001123 | 1.17 | 0.14 | 1.02 | 2.34 | | | no | 0.46 | 0.55 | | 0.97 | 0.48 |
| | | FPL | 011123 | 0.27 | −0.67 | 0.32 | 1.17 | | | no | 0.70 | 0.71 | | 0.03 | 0.54 |
| GR-sens | Start | SWM ds1 | 01234 | −0.46 | −0.41 | −0.62 | −1.07 | 0.26 | | yes | 1.20 | 1.04 | SWM ps4 | 2.08 | −0.48 |
| | | SWM ds2 | 01234 | 0.54 | 0.07 | 0.24 | −0.20 | 2.04 | | yes | 0.68 | 0.66 | | 1.43 | 0.05 |
| | | SWM ds3 | 01234 | 0.91 | 0.52 | 0.67 | 0.19 | 2.25 | | yes | 0.76 | 0.83 | SWM ps6 | 1.01 | 0.51 |
| | | SWM ps4 | 01234 | −0.70 | −0.60 | −0.53 | −1.47 | −0.20 | | yes | 1.19 | 1.02 | SWM ds1 | 0.72 | −0.52 |
| | | SWM ps5 | 01234 | 0.01 | 0.28 | −0.43 | −0.52 | 0.72 | | yes | 0.62 | 0.67 | | 0.68 | −0.02 |
| | | SWM ps6 | 01234 | 0.42 | 0.55 | 0.28 | −0.24 | 1.09 | | yes | 0.80 | 0.81 | SWM ds3 | 0.09 | 0.49 |
| | Final variant 1 | SWM ds1 | 01123 | −0.62 | −1.53 | −0.59 | 0.27 | | | no | 0.95 | 0.86 | | 1.62 | −0.76 |
| | | SWM ds2 | 01123 | 0.62 | −0.52 | 0.49 | 1.91 | | | no | 0.59 | 0.56 | | 1.32 | 0.21 |
| | | SWM ds3 | 01123 | 1.02 | 0.10 | 0.91 | 2.06 | | | no | 0.67 | 0.67 | | 0.06 | 0.62 |
| | Final variant 2 | SWM ps4 | 01123 | −0.55 | −1.02 | −0.71 | 0.08 | | | no | 0.98 | 0.87 | | 1.66 | −0.72 |
| | | SWM ps5 | 01123 | 0.34 | −0.21 | 0.11 | 1.13 | | | no | 0.51 | 0.50 | | 1.28 | 0.04 |
| | | SWM ps6 | 01123 | 0.83 | 0.29 | 0.61 | 1.59 | | | no | 0.78 | 0.76 | | 0.06 | 0.70 |
| GR-pp | Start | Bottle | 012345 | −0.33 | 0.25 | 1.79 | −4.18 | −0.58 | 1.06 | yes | 0.67 | 0.80 | | 1.48 | −0.47 |
| | | Jars | 012345 | −0.13 | 2.31 | −2.43 | −1.76 | −0.53 | 1.78 | yes | 1.19 | 1.21 | | 1.36 | −0.66 |
| | | Pegs | 012345 | −0.49 | −0.47 | −1.81 | −2.07 | 0.56 | 1.33 | yes | 0.78 | 0.77 | | 1.13 | 0.18 |
| | | Key | 012345 | 0.18 | 0.30 | 0.54 | −1.74 | −0.32 | 2.10 | yes | 0.68 | 0.70 | | 1.00 | 0.22 |
| | | Coins | 012345 | 0.46 | 0.24 | 0.13 | −0.56 | 0.39 | 2.09 | yes | 0.77 | 0.74 | | 0.95 | 0.31 |
| | | Nuts | 012345 | 1.71 | 0.77 | −0.32 | 3.35 | 0.13 | 4.60 | yes | 0.92 | 1.04 | | 0.08 | 0.40 |
| | Final all items | Bottle | 001123 | −0.30 | −1.33 | −0.93 | 1.36 | | | no | 0.67 | 0.82 | | 1.57 | −0.52 |
| | | Jars | 001123 | −0.01 | −1.58 | −0.63 | 2.16 | | | no | 1.07 | 1.08 | | 1.24 | −0.59 |
| | | Pegs | 001123 | −0.37 | −3.35 | 0.52 | 1.72 | | | no | 0.87 | 0.90 | | 1.15 | 0.29 |
| | | Key | 001123 | 0.61 | −0.42 | −0.27 | 2.52 | | | no | 0.61 | 0.69 | | 1.03 | 0.37 |
| | | Coins | 001123 | 1.04 | −0.12 | 0.70 | 2.54 | | | no | 0.73 | 0.71 | | 0.95 | 0.36 |
| | | Nuts | 001123 | 2.62 | 0.61 | 2.23 | 5.01 | | | no | 0.83 | 0.93 | | 0.06 | 0.17 |
| | Final reduced | Bottle | 001123 | −0.28 | −1.43 | −0.81 | 1.39 | | | no | 0.66 | 0.78 | | 1.41 | −0.46 |
| | | Key | 001123 | 0.67 | −0.39 | −0.20 | 2.58 | | | no | 0.60 | 0.66 | | 1.33 | 0.51 |
| | | Pegs | 001123 | −0.46 | −3.71 | 0.57 | 1.76 | | | no | 0.76 | 0.78 | | 1.21 | 0.53 |
| | | Nuts | 001123 | 2.74 | 0.64 | 2.26 | 5.31 | | | no | 0.69 | 0.79 | | 0.06 | −0.50 |
| GR-pa | Start and final | CylGrasp | 01234 | 1.19 | −5.02 | −0.24 | 3.15 | 6.88 | | no | 0.68 | 0.73 | | 1.62 | −0.68 |
| | | LatPinch | 01234 | 0.97 | −3.47 | −0.46 | 1.91 | 5.90 | | no | 0.40 | 0.40 | | 1.32 | −0.08 |
| | | TTpinch | 01234 | 2.76 | −0.88 | 0.71 | 3.30 | 7.92 | | no | 0.88 | 0.76 | | 0.06 | 0.73 |

[a]Detailed item fit statistics including the included items, the collapsing strategy, the location, the thresholds, their ordering, outfit and infit, local dependency, and unidimensionality.

GR, Graded and Redefined asssement of Strength, Sensibility and Prehension (GRASSP); GR-str, GRASSP strength; GRASSP sens; GRASSP sensation; GR-pa, GRASSP prehension ability; GR-pp, GRASSP prehension performance; Thres, threshold; LID, local item dependency; PC1, first principal component loading; Biceps, M. biceps; Delto, M. anterior deltoid; WristExt, M. extensor carpi radialis longus/brevis; Triceps, M. triceps; ExtDig, M. extensor digitorum communis; Opp, M. opponens; FPL, M. flexor pollicis longus; FDP, flexor digitorum profundus; AbdDigV, M. abductor digiti minimi; AbdDigII, M. first dorsal interosseus; SWM, Semmes and Weinstein monofilament; ds, dorsal sensation; ps, palmar sensation; CylGrasp, cylindrical grasp; LatPinch, lateral key pinch; TTPinch, tip-to-tip pinch.

Table 4. Collapsing Strategy of the GRASSP Subtests

| | *Number of items within subtest* | *Maximum recoded raw score* | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Original GR-str subscale | 10 | | 0 | 1 | 2 | 3 | 4 | 5 |
| Rescaled GR-str subscale | 4 | | | | | | | |
| Delto | | | 0 | 1 | 1 | 1 | 1 | 2 |
| Triceps, EDC | | | 0 | 0 | 1 | 1 | 2 | 3 |
| FPL | | | 0 | 1 | 1 | 1 | 2 | 3 |
| GR-str_red subscale | | **11** | | | | | | |
| Original GR-sens subscale (ps and ds) | 6 | | 0 | 1 | 2 | 3 | 4 | |
| Rescaled GR-ps subscale | 3 | **9** | 0 | 1 | 1 | 2 | 3 | |
| Rescaled GR-ds subscale | 3 | **9** | 0 | 1 | 1 | 2 | 3 | |
| Original GR-pa subscale | 3 | | 0 | 1 | 2 | 3 | 4 | |
| Rescaling not necessary | | **12** | | | | | | |
| Original GR-pp subscale | 6 | | 0 | 1 | 2 | 3 | 4 | 5 |
| Rescaled GR-pp subscale | 6 | **18** | 0 | 0 | 1 | 1 | 2 | 3 |
| Rescaled GR-pp_red subscale | 4 | **12** | 0 | 0 | 1 | 1 | 2 | 3 |

GR, Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP); Delto, M. anterior deltoid; Triceps, M. triceps; EDC, M. extensor digitorum communis; FDP, M. flexor digitorum profundus; GR-str, GRASSP strength; GR-str_red, GRASSP strength reduced; GR-sens, GRASSP sensation; ps, palmar sensation; ds, dorsal sensation; GR-ps, GRASSP palmar sensation; GR-ds, GRASSP dorsal sensation; GR-pa, GRASSP prehension ability; GR-pp, GRASSP prehension performance; GR-pp_red, GRASSP prehension performance reduced.

LID, unidimensionality, and ordered thresholds with the original scale response categories. However, the infit and outfit statistics were <0.5 for the later key pinch item point on a high discrimination of this item. The GR-pa subtest has only three items, and, therefore, further reduction to increase efficiency in clinical assessment was not undertaken.

### *The GR-str subtest*

The analysis with the GR-str subtest items resulted in disordered response categories for all 10 items and a high number of LID items. LID's of the GR-str subscale are shown in Figure 1.

The redundancies in the assessment of the GR-str subscale supported the decision to reduce the number of items. To achieve fit, it was necessary to collapse the six response categories (0–5) of GR-str into four recoded categories (range 0–3) with exception of M. deltoid, which had to be collapsed into three recoded categories (range 0–2) to achieve a good fit. More details about the collapsing strategies are shown in Tables 3 and 5. The eigenvalues of the subscale supported unidimensionality. The M. biceps and M. deltoid muscles were the most difficult to align to the scale, showing infit and outfit values >2.0, meaning low discrimination compared with the other items. Finally, expert clinical knowledge, results of

Table 5. Transformation Table[a]

| | *Rasch_GR-str_red_100* | | *Rasch_GR-ps_100* | | *Rasch_GR_ds_100* | | *Rasch_GR-pa_100* | | *Rasch_GR-pp_100* | | *Rasch_GR-pp_red_100* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Row score* | *Rasch ability* | *0–100 score* | *Rasch ability* | *0–100 score* | *Rasch ability* | *0–100 score* | *Rasch ability* | *0–100 score* | *Rasch ability* | *0–100 score* | *Rasch ability* | *0–100 score* |
| 0 | −3.445 | 0 | −2.759 | 0 | −2.423 | 0 | −6.292 | 0 | −4.338 | 0 | −4.408 | 0 |
| 1 | −2.353 | 15 | −1.855 | 14 | −1.651 | 14 | −4.369 | 12 | −3.250 | 11 | −3.012 | 13 |
| 2 | −1.350 | 29 | −1.015 | 28 | −0.936 | 28 | −2.555 | 24 | −2.257 | 20 | −1.747 | 25 |
| 3 | −0.717 | 38 | −0.429 | 37 | −0.445 | 37 | −1.240 | 33 | −1.653 | 26 | −1.012 | 32 |
| 4 | −0.230 | 44 | 0.077 | 45 | −0.024 | 45 | −0.302 | 39 | −1.201 | 31 | −0.442 | 38 |
| 5 | 0.193 | 50 | 0.566 | 53 | 0.380 | 52 | 0.540 | 44 | −0.819 | 34 | 0.074 | 43 |
| 6 | 0.594 | 56 | 1.082 | 61 | 0.810 | 60 | 1.423 | 50 | −0.473 | 38 | 0.581 | 48 |
| 7 | 1.005 | 61 | 1.693 | 71 | 1.327 | 70 | 2.393 | 56 | −0.144 | 41 | 1.103 | 53 |
| 8 | 1.461 | 68 | 2.578 | 85 | 2.112 | 84 | 3.476 | 63 | 0.178 | 44 | 1.661 | 58 |
| **9** | 2.025 | 76 | **3.531** | **100** | **2.964** | **100** | 4.803 | 72 | 0.502 | 47 | 2.303 | 64 |
| 10 | 2.874 | 87 | | | | | 6.257 | 81 | 0.833 | 50 | 3.145 | 72 |
| **11** | **3.791** | **100** | | | | | 7.731 | 90 | 1.176 | 54 | 4.541 | 85 |
| **12** | | | | | | | **9.207** | **100** | 1.539 | 57 | **6.072** | **100** |
| 13 | | | | | | | | | 1.933 | 61 | | |
| 14 | | | | | | | | | 2.376 | 65 | | |
| 15 | | | | | | | | | 2.906 | 71 | | |
| 16 | | | | | | | | | 3.607 | 77 | | |
| 17 | | | | | | | | | 4.720 | 88 | | |
| **18** | | | | | | | | | 5.933 | **100** | | |

[a]Transformation table for each GRASSP subtest including the respective total scores, the Rasch derived logit scale and the user-friendly, unbiased 0–100 metric. GR, Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP); GR-str_red, GRASSP strength reduced; GR-ps, GRASSP palmar sensation; GR-ds, GRASSP dorsal sensation; GR-pa, GRASSP prehension ability; GR-pp, GRASSP prehension performance; GR-pp_red, GRASSP prehension performance reduced.

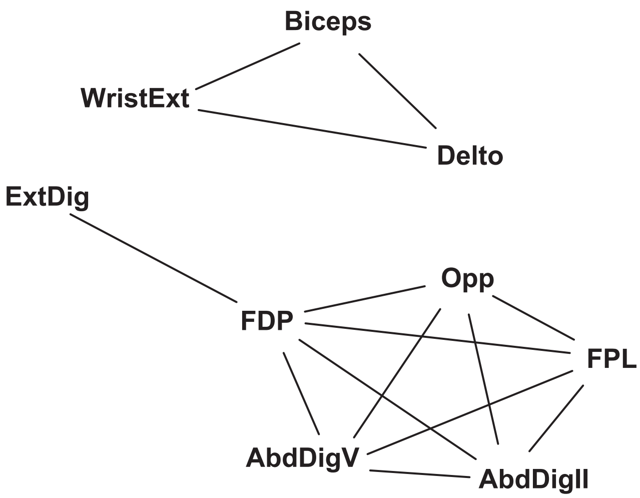


**FIG. 1.** Local item dependency >0.2 in the GR-str subscale. Biceps, M. biceps; Delto, M. anterior deltoid; WristExt, M. extensor carpi radialis longus/brevis; ExtDig, M. extensor digitorum communis; Opp, M. opponens; FPL, M. flexor pollicis longus; FDP, flexor digitorum profundus; AbdDigV, M. abductor digiti minimi; AbdDigII, M. first dorsal interosseous.

previous analyses of the items,[7] and results of the Rasch analysis guided the reduction of the number of items in the GR-str subtest. The final recoded and reduced GR-str subtest counts four items, and has still good reliability for use as measurement tool (PSI: 0.824), a slight ceiling effect, absence of LID, good item fit, and ordered thresholds.[36]

### *The GR-sens subtest*

The Rasch analysis of the GR-sens subtest showed a local dependency of the palmar and dorsal test locations, especially for the thumb and the small finger, and multidimensionality, with the LID item's loading stronger on a common dimension. When the palmar and dorsal test locations are dependent, one can assume that they are redundant, representing one and the same aspect of sensation. This supports the use of either the GR-ps or GR-ds and, for that reason, separate calibrations are shown for the dorsal and the palmar side in Tables 2 and 3. The separate analysis of GR-ps and GR-ds solved the multidimensionality issue. To accommodate for disordered thresholds, the five response categories (0–4) of GR-ps and GR-ds were collapsed into three recoded categories (range 0–3). The same collapsing strategy worked well for GR-ps and GR-ds (Tables 3 and 4). The GR-sens subtest showed the highest proportion of participants achieving the maximal number of points.

### *The GR-pp subtest*

A first Rasch analysis with all items of the GR-pp subtest supported its unidimensionality, as well as the absence of LID and DIF. This subtest presented the best reliability of all subtests, with PSI = 0.956. Ceiling and floor effects were not observed. The infit and outfit statistics were within the 0.5–1.7 range.

The main challenge of the GR-pp subtest was the disordering of the thresholds. A common collapsing strategy could be applied to all the six tasks. Collapsing of the six response categories (0–5) into four recoded categories (range 0–3) improved the item fit (Table 3). A further reduction of the tasks from six to four based on clinical expertise did not impact the reliability of the subscale (PSI = 0.937).

## Discussion

The study revealed that after minor modifications, GRASSP Version 2 subtests responded well to the Rasch analysis supporting sound measurement properties and being applicable as an interval level scaled outcome measure of upper limb function.[9] In addition, GRASSP subtests can be reliably applied in longitudinal studies (i.e., following acute SCI) because of the invariance of the GRASSP items over the assessment time points. Lastly, the assessment time of the GRASSP can be shortened by omitting some items, without losing much comprehensiveness, accuracy, and reliability.

As expected, the initial Rasch analysis showed that the thresholds for all GRASSP subtests with exception of GR-pa were disordered, affecting the applicability of the row scores of quantitative analysis.[37,38] However, after collapsing response categories (Table 4), and after Rasch adjustment, the thresholds were well ordered, whereas the Rasch analyses showed no substantial loss in reliability.[39] This means that although the response categories were reduced and some items were deleted, the continuum of levels of ability can still be discriminated in our population group. For example, Figure 2 shows that a simple task such as ''pegs'' can be performed by less able individuals, as it is on the lower range of the scale. In contrast, more difficult tasks such as ''nuts'' are on the higher range of the scale, and allow for discriminating among more capable individuals with greater dexterity to perform these tasks.

Dependency of items was observed between almost all GR-str items (Fig. 1) and between GR-ds and GR-ps items. Local

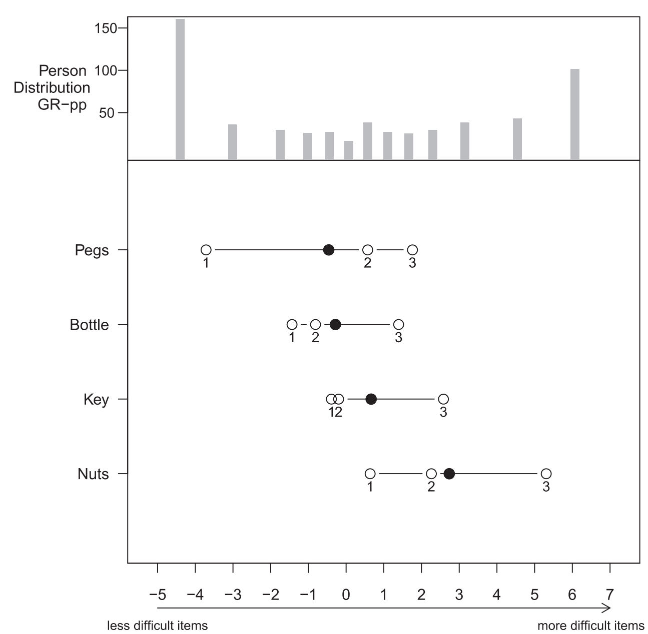


**FIG. 2.** Person item map for the prehension performance subtest. The gray bars are the number of observations. In total, 614 observations (all five examination stages, with dropouts) were included for analysis. The y axis represents the four tasks within the prehension performance subtest. The white dots are the item difficulty thresholds. The black dots are the item locations, or item difficulties, of each task and correspond to the mean of an item's difficulty thresholds. Item difficulty thresholds are the equal probability points between response options.

dependencies within GR-str were solved by discarding selected items (muscles) of the subtest, which is based on clinical judgement. A previous longitudinal study in individuals with tetraplegia revealed[7] that early assessment of a maximum of three muscle variables (e.g., M. deltoid, M. extensor digitorum communis [EDC], and M. flexor pollicis longus [FPL]) are sufficient to predict upper limb function and independence in activities of daily living (ADL) at 6 months after cervical SCI. Therefore, the decision to reduce items was also based on the results of the aforementioned prediction study[7] and clinical insights (Table 3).

The local dependencies of GR-ps and GR-ds were not surprising, as the applied thresholds on the three test locations of the palmar and dorsal surface of the hand represent the same anatomical areas (segments C6, C7, and C8, respectively). It was confirmed in this study that GR-ps and GR-ds were uniform, and, therefore, GR-ps or GR-ds can be used for clinical and research purposes in the future.[40,41] As there is some evidence that GR-ps can detect more change than GR-ds over time,[1] we would suggest using GR-ps after cervical SCI, although in the Rasch analysis the results of GR-ps and GR-ds were similar.

The level and completeness, the extent of recovery, and the individual performance in cervical SCI is highly variable.[1,3,4] The high variability influences the decision to choose appropriate outcome measures (e.g., responsive, accurate) in a clinical trial to detect an intervention effect. In our study, not including extreme values, we did not find floor and ceiling effects for GR-pa and GR-pp; however, we still found small floor and ceiling effects for GR-str (floor 0%; ceiling <7%) and GR-sens (floor <8%; ceiling <17%). Ceiling and floor effects are recognized in SCI studies,[1] and to overcome limits of manual muscle testing, it was recommended to use handheld dynamometry to identify effects of therapeutic or hand surgical interventions over time. In addition, the performance of an unbiased recursive partitioning method, called a ‘‘conditional inference tree’’ (URP-CTREE),[42] might be used to identify homogeneous outcome subgroups to improve the sensitivity of trials.[2,7]

The GR-pa subtest did not require any adjustment to fit the Rasch model, which is a unique finding. The GR-pa was designed to give a first impression of the hand through assessing the ability to perform three grasp patterns (cylindrical grasp, lateral key pinch, and tip to tip pinch) that can develop early in the recovery phase after SCI.

In SCI, there is little research regarding Rasch metric properties of upper limb conditions[43,44] or functional outcome measures.[14,17] This is in contrast to outcome measures used to evaluate upper limb function in neurological conditions[45–55] or peripheral upper limb specific disorders.[56–59] The Van Lieshout test measures upper limb functional tasks[60,61] in tetraplegia, however, it was not designed to provide information regarding changes in sensory-motor impairment. After collapsing the response options for 7 out of 10 items of the Van Lieshout Test, the Rasch analysis showed monotonicity, unidimensionality and good reliability (PSI = 0.91; Cronbach Alpha = 0.95). The robustness of the Van Lieshout scale's hierarchy of item difficulties over time was not verified. A pronounced ceiling effect (11%) is reported, which indicates a higher proportion of persons with good abilities to perform the functional tasks.

In comparison, the GRASSP revealed ordered item difficulty response thresholds, valid for measuring upper limb function in SCI over time and discriminating among persons with a broad range of abilities (from poor to good) following cervical SCI. Based on these findings, GRASSP subtests can be well applied as initially developed. Although the GRASSP subtests can remain the same, the modified GRASSP Version 2 scoring needs to be applied when the scores will be used for metric measurements. For any motor scoring a change from no innervation (score 0) to any innervation (scores 1–5) is likely relevant.[62] However, it is unclear as to whether a change from 0 to 1 is as significant as a change from 4 to 5 for the same muscle, which again is true for sum scores. Unfortunately, in research ordinal scales are often simply summed,[1,13,15,63] and automatically bias the change needed to achieve a treatment effect.[19,20,38] Therefore, when sum scores are needed, it is recommended to use the Rasch scoring algorithm to transform the person's ability to a linear interval score ranging from 0 to 100: that is, the ‘‘Rasch_GR_100 subscale.’’ The Rasch_GR_100 subscale can be derived from the GRASSP subtest ordinal scale without changing the face validity of the outcome measure (Tables 4 and 5). Also, the items included in the final models, the collapsing strategy for disordered response options, and the threshold difficulties as shown in Table 3 would allow replicating our models in an anchored Rasch analysis.

### *Limitations*

At this point, a comparison across Rasch_GR_100 subscales is not recommended, as the subscales have not been co-calibrated yet. Knowledge about the subscale's relative difficulties is yet lacking; for example, we do not know how difficult the ‘‘strength’’ subscale of the GRASSP is compared with the other subscales. A co-calibration, which would put all the GRASSP subscales on a common metric, may be undertaken in the future, but was not an aim of this study, in which the measurement reliability and validity of the separate subscales across time was the focus of interest.

We examined DIF based on examination stage and laterality, and more clinical constructs such as age and gender can be added to the DIF analysis to see how the individual items behave with the different constructs.

We have a convenience sample of 77 individuals, (measured repeatedly but with dropouts); however, in future, a larger sample is recommended to increase the power of the analysis.

It was not within the scope of this study to investigate if the final Rasch_GR_100 subscales are sensitive over time. More discussion will be needed in the future to explore the responsiveness of the final Rasch_GR_100 subscales.

## Conclusion

The Rasch analysis with the GRASSP Version 2 subtests performed well in a sample of individuals with cervical SCI with different impairment levels, from the acute phase to the chronic phase of recovery. This study supports that the GRASSP Version 2 subtests can be transformed to a metric interval using the transformation table, and that the Rasch converted scores of the GRASSP allow interval level measurement. Further, the findings of this study showed that the GRASSP Version 2 can be applied as a shorter version without losing accuracy, which is beneficial for clinical settings and research purposes.

## Acknowledgments

We thank the individuals with SCI who participated in this study, and the occupational therapists at the participating centers: Spinal Cord Injury Center, Hohe Warte, Bayreuth, Germany; Orthopädische Universitätsklinik, Heidelberg, Germany; Unfallklinik Murnau, Germany; Balgrist University Hospital, Zurich, Switzerland; and the Swiss Paraplegic Center, Nottwil, Switzerland. Further, we thank Dr. Doris Maier (Murnau) for her collaboration on this project. The study was supported by the Swiss Paraplegic Center, Nottwil, Switzerland,

and funded by the Clinical Research Priority Program NeuroRehab of the University of Zurich, Switzerland.

## Author Disclosure Statement

No competing financial interests exist.

## References


1. Velstra, I.M., Curt, A., Frotzler, A., Abel, R., Kalsi-Ryan, S., Rietman, J.S., and Bolliger, M. (2015). Changes in strength, sensation, and prehension in acute cervical spinal cord injury: European multicenter responsiveness study of the GRASSP. Neurorehabil. Neural Repair 29, 755–766.
2. Velstra, I.M., Bolliger, M., Tanadini, L.G., Baumberger, M., Abel, R., Rietman, J.S., and Curt, A. (2014). Prediction and stratification of upper limb function and self-care in acute cervical spinal cord injury with the Graded Redefined Assessment of Strength, Sensibility, and Prehension (GRASSP). Neurorehabil. Neural Repair 28, 632–642.
3. Steeves, J.D., Lammertse, D.P., Kramer, J.L., Kleitman, N., Kalsi-Ryan, S., Jones, L., Curt, A., Blight, A.R., and Anderson, K.D. (2012). Outcome measures for acute/subacute cervical sensorimotor complete (AIS-A) spinal cord injury during a phase 2 clinical trial. Top. Spinal Cord Inj. Rehabil. 18, 1–14.
4. Kramer, J.L., Lammertse, D.P., Schubert, M., Curt, A., and Steeves, J.D. (2012). Relationship between motor recovery and independence after sensorimotor-complete cervical spinal cord injury. Neurorehabil. Neural Repair 26, 1064–1071.
5. Kalsi-Ryan, S., Beaton, D., Curt, A., Duff, S., Popovic, M.R., Rudhe, C., Fehlings, M.G., and Verrier, M.C. (2012). The Graded Redefined Assessment of Strength Sensibility and Prehension: reliability and validity. J. Neurotrauma 29, 905–914.
6. Rudhe, C., and van Hedel, H.J. (2009). Upper extremity function in persons with tetraplegia: relationships between strength, capacity, and the spinal cord independence measure. Neurorehabil. Neural Repair 23, 413–421.
7. Velstra, I.M., Bolliger, M., Krebs, J., Rietman, J.S., and Curt, A. (2016). Predictive value of upper limb muscles and grasp patterns on functional outcome in cervical spinal cord injury. Neurorehabil Neural Repair 30, 295–306.
8. Kalsi-Ryan, S., Beaton, D., Ahn, H., Askes, H., Drew, B., Curt, A., Popovic, M., Wang, J., Verrier, M., and Fehlings, M. (2015). Responsiveness, sensitivity and minimally detectable difference of the Graded and Redefined Assessment of Strength, Sensibility, and Prehension, Version 1.0 (GRASSP V1). J Neurotrauma.
9. Nunnally, J.C., and Bernstein, I.H. (1994). *Psychometric Theory*, 3rd ed. McGraw-Hill: New York; London.
10. Kirshblum, S.C., Waring, W., Biering-Sorensen, F., Burns, S.P., Johansen, M., Schmidt-Read, M., Donovan, W., Graves, D., Jha, A., Jones, L., Mulcahey, M.J., and Krassioukov, A. (2011). Reference for the 2011 revision of the International Standards for Neurological Classification of Spinal Cord Injury. J. Spinal Cord Med. 34, 547–554.
11. Catz, A., Itzkovich, M., Agranov, E., Ring, H., and Tamir, A. (1997). SCIM—spinal cord independence measure: a new disability scale for patients with spinal cord lesions. Spinal Cord 35, 850–856.
12. Ravaud, J.F., Delcey, M., and Yelnik, A. (1999). Construct validity of the functional independence measure (FIM): questioning the unidimensionality of the scale and the ''value'' of FIM scores. Scand. J. Rehabil. Med. 31, 31–41.
13. Raczek, A.E., Ware, J.E., Bjorner, J.B., Gandek, B., Haley, S.M., Aaronson, N.K., Apolone, G., Bech, P., Brazier, J.E., Bullinger, M., and Sullivan, M. (1998). Comparison of Rasch and summated rating scales constructed from SF–36 physical functioning items in seven countries: results from the IQOLA project. International Quality of Life Assessment. J. Clin. Epidemiol. 51, 1203–1214.
14. Prodinger, B., Ballert, C.S., Brinkhof, M.W., Tennant, A., and Post, M.W. (2016). Metric properties of the Spinal Cord Independence Measure – self report in a community survey. J. Rehabil. Med. 48, 149–164.
15. McHorney, C.A., Haley, S.M., and Ware, J.E., Jr. (1997). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. J. Clin. Epidemiol. 50, 451–461.
16. Luther, S.L., Kromrey, J., Powell-Cope, G., Rosenberg, D., Nelson, A., Ahmed, S., and Quigley, P. (2006). A pilot study to modify the SF-36V physical functioning scale for use with veterans with spinal cord injury. Arch. Phys. Med. Rehabil. 87, 1059–1066.
17. Catz, A., Itzkovich, M., Tesio, L., Biering-Sorensen, F., Weeks, C., Laramee, M.T., Craven, B.C., Tonack, M., Hitzig, S.L., Glaser, E., Zeilig, G., Aito, S., Scivoletto, G., Mecci, M., Chadwick, R.J., El Masry, W.S., Osman, A., Glass, C.A., Silva, P., Soni, B.M., Gardner, B.P., Savic, G., Bergstrom, E.M., Bluvshtein, V., and Ronen, J. (2007). A multicenter international study on the Spinal Cord Independence Measure, version III: Rasch psychometric validation. Spinal Cord 45, 275–291.
18. Wu, X., Liu, J., Tanadini, L.G., Lammertse, D.P., Blight, A.R., Kramer, J.L., Scivoletto, G., Jones, L., Kirshblum, S., Abel, R., Fawcett, J., Field-Fote, E., Guest, J., Levinson, B., Maier, D., Tansey, K., Weidner, N., Tetzlaff, W.G., Hothorn, T., Curt, A., and Steeves, J.D. (2015). Challenges for defining minimal clinically important difference (MCID) after spinal cord injury. Spinal Cord 53, 84–91.
19. Tanadini, L.G., Steeves, J.D., Curt, A., and Hothorn, T. (2016). Autoregressive transitional ordinal model to test for treatment effect in neurological trials with complex endpoints. BMC Med. Res. Methodol. 16, 149.
20. Grimby, G., Tennant, A., and Tesio, L. (2012). The use of raw scores from ordinal scales: time to end malpractice? J. Rehabil. Med. 44, 97–98.
21. Rasch, G.W. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish National Institute for Educational Research.
22. Bond, T.G., and Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences,* 2nd ed. Lawrence Erlbaum: Mahwah, NJ; London.
23. Tennant, A., and Conaghan, P.G. (2007). 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. 57, 1358–1362.
24. Smith, R.M., Schumacker, R.E., and Bush, M.J. (1998). Using item mean squares to evaluate fit to the Rasch model. J. Outcome Meas. 2, 66–78.
25. Wright, B.D., Linacre, J.M., Gustafson, J.E., and Martin-Löf, P. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions 8, 370.
26. Smith, E.V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J. Appl. Meas. 3, 205–231.
27. Wainer, H., and Kiely, G.L. (1987). Item clusters and computerized adaptive testing: a case for testlets. J. Educ. Meas. 24, 185–201.
28. Raîche, G. (2005). Critical eigenvalue sizes (variances) in standardized residual principal components analysis. Rasch Measurement Transactions 19, 1012. https://rasch.org/rmt/contents.htm
29. Holland, P.W., and Wainger, H. (1993). *Differrential Item Functioning*. New York and London: Erlbaum.
30. Bland, J.M., and Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. BMJ 310, 170.
31. Christensen, K.B., Kreiner, S. and Mesbah, M. (2013). *Rasch Models in Health.* ISTE: Hoboken, NJ; John Wiley & Sons: London.
32. McHorney, C.A., and Tarlov, A.R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual. Life Res. 4, 293–307.
33. Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334.
34. Team, R.C. (2016). *R Core Team. R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna.
35. Mair, P., and Hatzinger, R. (2007). Extended Rasch modeling: the eRm package for the application of IRT models in R. J. Stat. Softw. 20, 1–20.
36. Linacre, J.M., and Wright, B.D. (2002). Construction of measures from many-facet data. J. Appl. Meas. 3, 486–512.
37. Tennant, A., McKenna, S.P., and Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. Value Health 7, Suppl. 1, S22–26.
38. Scott, S.C., Goldberg, M.S., and Mayo, N.E. (1997). Statistical assessment of ordinal outcomes in comparative studies. J. Clin. Epidemiol. 50, 45–55.
39. Grondin, J., and Blais, J.G. (2010). *A Rasch Analysis on Collapsing Categories in Item's Response Scales of Survey Questionnaire: Maybe It's Not One Size Fits All.* American Educational Research Association: Denver.
40. Jerosch-Herold, C. (2005). Assessment of sensibility after nerve injury and repair: a systematic review of evidence for validity, reliability and responsiveness of tests. J. Hand Surg. Br. 30, 252–264.

41. Beekhuizen, K.S., and Field–Fote, E.C. (2008). Sensory stimulation augments the effects of massed practice training in persons with tetraplegia. Arch. Phys. Med. Rehabil. 89, 602–608.
42. Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Stat. 15, 651–674.
43. Spooren, A.I., Arnould, C., Smeets, R.J., Snoek, G., and Seelen, H.A. (2013). Reference values for the transformed Van Lieshout hand function test for tetraplegia. Spinal Cord 51, 745–749.
44. Spooren, A.I., Arnould, C., Smeets, R.J., Bongers, H.M., and Seelen, H.A. (2013). Improvement of the Van Lieshout hand function test for tetraplegia using a Rasch analysis. Spinal Cord 51, 739–744.
45. Balasubramanian, C.K., Li, C.Y., Bowden, M.G., Duncan, P.W., Kautz, S.A., and Velozo, C.A. (2016). Dimensionality and item-difficulty hierarchy of the lower extremity Fugl-Meyer Assessment in individuals with subacute and chronic stroke. Arch. Phys. Med. Rehabil. 97, 582–589.
46. Chen, H.F., Lin, K.C., Wu, C.Y., and Chen, C.L. (2012). Rasch validation and predictive validity of the action research arm test in patients receiving stroke rehabilitation. Arch. Phys. Med. Rehabil. 93, 1039–1045.
47. Chen, H.F., Wu, C.Y., Lin, K.C., Chen, H.C., Chen, C.P., and Chen, C.K. (2012). Rasch validation of the streamlined Wolf Motor Function Test in people with chronic stroke and subacute stroke. Phys. Ther. 92, 1017–1026.
48. Chen, H.F., Wu, C.Y., Lin, K.C., Jang, Y., Lin, S.C., Cheng, J.W., Chung, C.Y., and Yan, Y. (2014). Measurement properties of streamlined wolf motor function test in patients at subacute to chronic stages after stroke. Neurorehabil. Neural Repair 28, 839–846.
49. Dalton, E., Lannin, N.A., Laver, K., Ross, L., Ashford, S., McCluskey, A., and shoulder and hand questionnaire in adults following stroke. Disabil. Rehabil. 39, 2504–2511.
50. Koh, C.L., Hsueh, I.P., Wang, W.C., Sheu, C.F., Yu, T.Y., Wang, C.H., and Hsieh, C.L. (2006). Validation of the action research arm test using item response theory in patients after stroke. J. Rehabil. Med. 38, 375–380.
51. Lundgren Nilsson, A., and Tennant, A. (2011). Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. J. Rehabil. Med. 43, 884–891.
52. O'Dell, M.W., Kim, G., Rivera, L., Fieo, R., Christos, P., Polistena, C., Fitzgerald, K., and Gorga, D. (2013). A psychometric evaluation of the Arm Motor Ability Test. J. Rehabil. Med. 45, 519–527.
53. Woodbury, M., Velozo, C.A., Thompson, P.A., Light, K., Uswatte, G., Taub, E., Winstein, C.J., Morris, D., Blanton, S., Nichols-Larsen, D.S., and Wolf, S.L. (2010). Measurement structure of the Wolf Motor Function Test: implications for motor control theory. Neurorehabil. Neural Repair 24, 791–801.
54. Woodbury, M.L., Velozo, C.A., Richards, L.G., and Duncan, P.W. (2013). Rasch analysis staging methodology to classify upper extremity movement impairment after stroke. Arch. Phys. Med. Rehabil. 94, 1527–1533.
55. Woodbury, M.L., Velozo, C.A., Richards, L.G., Duncan, P.W., Studenski, S., and Lai, S.M. (2007). Dimensionality and construct validity of the Fugl-Meyer Assessment of the upper extremity. Arch. Phys. Med. Rehabil. 88, 715–723.
56. Braitmayer, K., Dereskewitz, C., Oberhauser, C., Rudolf, K.D., and Coenen, M. (2016). Examination of the applicability of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire to patients with hand injuries and diseases using Rasch analysis. Patient 10, 367–376.
57. Franchignoni, F., Giordano, A., Sartorio, F., Vercelli, S., Pascariello, B., and Ferriero, G. (2010). Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch validation study. Arch. Phys. Med. Rehabil. 91, 1370–1377.
58. Prodinger, B., Stamm, T., Peterson, D., Stucki, G., and Tennant, A. (2016). Toward a standardized reporting of outcomes in hand osteoarthritis: developing a common metric of outcome measures commonly used to assess functioning. Arthritis Care Res. (Hoboken) 68, 1115–1127.
59. Taylor, W.J., and McPherson, K.M. (2007). Using Rasch analysis to compare the psychometric properties of the Short Form 36 physical function score and the Health Assessment Questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis. Arthritis Rheum. 57, 723–729.
60. Post, M.W., Van Lieshout, G., Seelen, H.A., Snoek, G.J., Ijzerman, M.J., and Pons, C. (2006). 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 44, 763–771.
61. Spooren, A.I., Janssen-Potten, Y.J., Post, M.W., Kerckhofs, E., Nene, A., and Seelen, H.A. (2006). Measuring change in arm hand skilled performance in persons with a cervical spinal cord injury: responsiveness of the Van Lieshout Test. Spinal Cord 44, 772–779.
62. Snoek, G.J., IJzerman, M.J., Hermens, H.J., Maxwell, D., and Biering-Sorensen, F. (2004). Survey of the needs of patients with spinal cord injury: impact and priority for improvement in hand function in tetraplegics. Spinal Cord 42, 526–532.
63. Hobart, J.C., Cano, S.J., Zajicek, J.P., and Thompson, A.J. (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol. 6, 1094–1105.

Address correspondence to:
*Inge-Marie Velstra, PhD*
*Clinical Trial Unit*
*Swiss Paraplegic Center, Nottwil*
*Guido A Zächstrasse 1*
*6207 CH Nottwil*
*Switzerland*

*E-mail:* inge-marie.velstra@paraplegie.ch