

Open camera or QR reader and
scan code to access this article
and other resources online.



ORIGINAL ARTICLE

CLINICAL STUDIES

Determining the Hierarchy of Coma Recovery Scale-Revised Rating Scale Categories and Alignment with Aspen Consensus Criteria for Patients with Brain Injury: A Rasch Analysis

Jennifer A. Weaver,^{1,2,*} Alison M. Cogan,³ Katherine A. O'Brien,⁴ Piper Hansen,⁵ Joseph T. Giacino,^{6,7} John Whyte,⁸ Theresa Bender Pape,⁹⁻¹¹ Philip van der Wees,^{2,12} and Trudy Mallinson²

Abstract

This study aimed to empirically evaluate the hierarchical structure of the Coma Recovery Scale-Revised (CRS-R) rating scale categories and their alignment with the Aspen consensus criteria for determining disorders of consciousness (DoC) following a severe brain injury. CRS-R data from 262 patients with DoC following a severe brain injury were analyzed applying the partial credit Rasch Measurement Model. Rasch Analysis produced logit calibrations for each rating scale category. Twenty-eight of the 29 CRS-R rating scale categories were operationalized to the Aspen consensus criteria. We expected the hierarchical order of the calibrations to reflect Aspen consensus criteria. We also examined the association between the CRS-R Rasch person measures (indicative of performance ability) and states of consciousness as determined by the Aspen consensus criteria. Overall, the order of the 29 rating scale category calibrations reflected current literature regarding the continuum of neurobehavioral function: category 6 “Functional Object Use” of the Motor item was hardest for patients to achieve; category 0 “None” of the Oromotor/Verbal item was easiest to achieve. Of the 29 rating scale categories, six were not ordered as expected. Four rating scale categories reflecting the Vegetative State (VS)/Unresponsive Wakefulness Syndrome (UWS) had higher calibrations (reflecting greater neurobehavioral function) than the easiest Minimally Conscious State (MCS) item (category 2 “Fixation” of the Visual item). Two rating scale categories, one reflecting MCS and one not operationalized to the Aspen consensus criteria, had higher calibrations than the easiest eMCS item (category 2 “Functional: Accurate” of the Communication item). CRS-R person

¹Department of Occupational Therapy, College of Health and Human Sciences, Colorado State University, Fort Collins, Colorado, USA.

²Department of Occupational Therapy, College of Applied Health Sciences, George Washington University, Washington, DC, USA.

³VA Greater Los Angeles Healthcare System, Los Angeles, California, USA.

⁴TIRR Memorial Herman Rehabilitation Hospital, Houston, Texas, USA.

⁵College of Applied Health Sciences, University of Illinois Chicago, Chicago, Illinois, USA.

⁶Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Charlestown, Massachusetts, USA.

⁷Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, Massachusetts, USA.

⁸Moss Rehabilitation Research Institute, Albert Einstein Healthcare Network, Philadelphia, Pennsylvania, USA.

⁹Research and Development Service, ¹⁰Center for Innovation in Complex Chronic Healthcare and Research Service, Department of Veteran Affairs, Edward Hines Jr. VA Hospital, Hines, Illinois, USA.

¹¹Department of Physical Medicine and Rehabilitation, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA.

¹²Department of Rehabilitation and IQ Healthcare, Radboud University Medical Center, Nijmegen, the Netherlands.

*Address correspondence to: Jennifer A. Weaver, PhD, Department of Occupational Therapy, College of Health and Human Sciences, Colorado State University, 1573 Campus Delivery, Occupational Therapy Building, Room 222, Fort Collins, CO, 80525-1573, USA E-mail: jen.weaver@colostate.edu

measures (indicating amount of neurobehavioral function) and states of consciousness, based on Aspen consensus criteria, showed a strong correlation ($r_s = 0.86$; $p < 0.01$). Our study provides empirical evidence for revising the diagnostic criteria for MCS to also include category 2 “Localization to Sound” of the Auditory item and for Emerged from Minimally Conscious State (eMCS) to include category 4 “Consistent Movement to Command” of the Auditory item.

Keywords: brain injury; disorders of consciousness; measurement; outcome assessment

Introduction

Accurate diagnosis of state of consciousness among adults with disorders of consciousness (DoC) following brain injury is critical because it is associated with prognosis. In the United States, prognosis for recovery from DoC influences access to specialty DoC rehabilitation services.¹ DoC includes a range of states from coma (no arousal, sleep/wake cycles, or awareness), vegetative state/unresponsive wakefulness syndrome (VS/UWS; presence of wakefulness without awareness), to the minimally conscious state (MCS, inconsistent volitional behavior).² Patients who have emerged from Minimally Conscious State (eMCS) demonstrate consistent functional behavior.² To address the need for accurate diagnosis, the Aspen Neurobehavioral Conference Workgroup defined diagnostic criteria for Minimally Conscious State (MCS) and proposed criteria for eMCS based on an evidence review and expert consensus.³ The Aspen Workgroup focused on MCS and eMCS as comatose and VS/UWS were well defined.⁴ Nonetheless, differentiating DoC based on behavioral observations remains challenging because certain behaviors may occur infrequently and random movements can be interpreted as volitional behavior.⁴

The Coma Recovery Scale-Revised (CRS-R) is the reference standard for assessment of neurobehavioral function and is used to diagnose DoC.⁵ The CRS-R consists of six items (i.e., subscales): Auditory, Visual, Motor, Oromotor/Verbal, Communication, and Arousal. These six items, known as the Coma Recovery Scale, were developed by a team of multi-disciplinary professionals.⁶ The Coma Recovery Scale items and rating scale categories were revised to refine the assessment’s

clinical utility, construct validity, and diagnostic utility based on the development of diagnostic criteria by the Aspen Workgroup.³ Previous psychometric analysis of the CRS-R demonstrated unidimensionality and monotonicity, indicating all items reflect the concept of neurobehavioral function and logit values for the rating scale categories (i.e., scores on each subscale) occur in order.⁷ Analysis using the Rasch Partial Credit Model established the item hierarchy based on average item measure, from most to least neurobehavioral function, as Communication, Oromotor/Verbal, Auditory, Visual, Motor, and Arousal for individuals with a traumatic brain injury.⁸ For the purposes of applying and interpreting the Rasch model, the CRS-R subscales are treated as “items” and each score achieved within a subscale is a “rating scale category.” The prior analysis did not describe the hierarchy of the 29 rating scale categories.⁸

Prior work operationalized 28 of the 29 CRS-R rating scale categories to VS/UWS and the Aspen consensus criteria for MCS and eMCS (Table 1).⁹ More specifically, the CRS-R criteria for diagnosis of VS/UWS is delineated with 15 categories, MCS is delineated with 11 categories, and eMCS is delineated with two categories (Table 1).⁹ Diagnosis of VS/UWS requires the patient to achieve a rating scale category operationalized to VS/UWS on every item; whereas, diagnosis of MCS or eMCS requires the patient to achieve an operationalized rating scale category for only one item. Although the CRS-R item hierarchy is established, empiric evidence is required to support whether the rating scale categories are accurately operationalized to VS/UWS, MCS, and eMCS. We purposefully chose to include VS/UWS criteria in order to examine the continuum of neurobehavioral function.

Table 1. CRS-R Items and Rating Scale Categories Operationalized to Each State of Consciousness Based on the Aspen Consensus Criteria

CRS-R items	Rating scale categories operationalized to the vegetative state/unresponsive wakefulness syndrome	Rating scale categories operationalized to the minimally conscious state	Rating scale categories operationalized to the emerged from minimally conscious state
Communication	0	1	2
Auditory	0-2	3-4	–
Visual	0-1	2-5	–
Motor	0-2	3-5	6
Arousal*	0-2	–	–
Oromotor/Verbal	0-2	3	–

*Rating scale category 3 “Attention” of the Arousal item is not aligned to the Aspen criteria.

One study demonstrates that eMCS could be diagnosed with an additional CRS-R rating scale category,¹⁰ which differs from how the Aspen consensus criteria are operationalized to the CRS-R. Specifically, two rating scale categories indicative of eMCS (category 6 “Functional Object Use” of the Motor item and category 2 “Functional: Accurate” of the Communication item), were compared with the occurrence of category 4 “Consistent Movement to Command” of the Auditory item (indicative of MCS). Findings indicated these three rating scale categories occurred at the same time for 50% of the participants.¹⁰

The current study builds on this work in several important ways: the primary purpose of this study is to empirically evaluate the hierarchical ordering of the CRS-R rating scale categories and their alignment with each of the Aspen consensus criteria for VS/UWS, MCS, and eMCS. We hypothesize that the two rating scale categories indicative of eMCS will be in the same approximate upper logit region of the neurobehavioral function continuum, while the 11 rating scale categories indicative of MCS and 15 rating scale categories indicative of VS/UWS will be approximately located in markedly different and lower regions of the continuum (Table 1). A secondary purpose of this study is to examine the association between CRS-R Rasch person measures (indicative of person ability level and akin to a total raw score) and Aspen consensus criteria to substantiate that the VS/UWS, MCS, and eMCS are distinctly different. Rasch Measurement Theory is appropriate for refining existing assessments to evaluate whether the items and rating scale categories align with current knowledge, such as the Aspen consensus criteria.^{11,12}

Methods

Data sources

The CRS-R data set for this study was assembled from four cohorts: two clinical trials and two rehabilitation hospitals. One clinical trial administered amantadine or placebo and all participants received inpatient rehabilitation services.¹³ The other clinical trial administered active or placebo repetitive transcranial magnetic stimulation in a hospital setting; participants did not receive additional rehabilitation services.¹⁴ The two rehabilitation hospitals are located in metropolitan areas in the Midwest and Southern regions of the United States. Following Institutional Review Board approval from The George Washington University, data were aggregated into a single data set.

Participants ($n=262$) were included if they were >14 years old and had DoC from a brain injury. We included participants age 14-17 because the CRS-R has been administered on adolescents and young adults and is consistent with clinical practice.^{13,15} Participants had at least one CRS-R assessment and up to 37 re-assessments, resulting in a dataset of 1142 CRS-R records. The CRS-R

was administered and scored by a rehabilitation practitioner or trained researcher. Data also were collected on age, time from onset to study enrollment or rehabilitation admission, and etiology of the brain injury.

Measure

Coma Recovery Scale-Revised. Each CRS-R item (i.e., subscale) includes a hierarchical ordering of rating scale categories (i.e., scores); a higher score indicates more neurobehavioral function. The assessor begins with the highest rating scale category; if a response is observed and meets the scoring criteria, the assessor moves to the next item.¹⁶ If no response is observed or the response does not meet the scoring criteria, the assessor continues down to the next rating scale category for that item. The number of rating scale categories varies by item; for example, Communication has three rating scale categories (2, 1, 0) while Motor has seven rating scale categories (6, 5, 4, 3, 2, 1, 0). Total CRS-R raw scores range from 0-23. The CRS-R Administration and Scoring Guidelines can be found on the Rehab Measures database.¹⁷

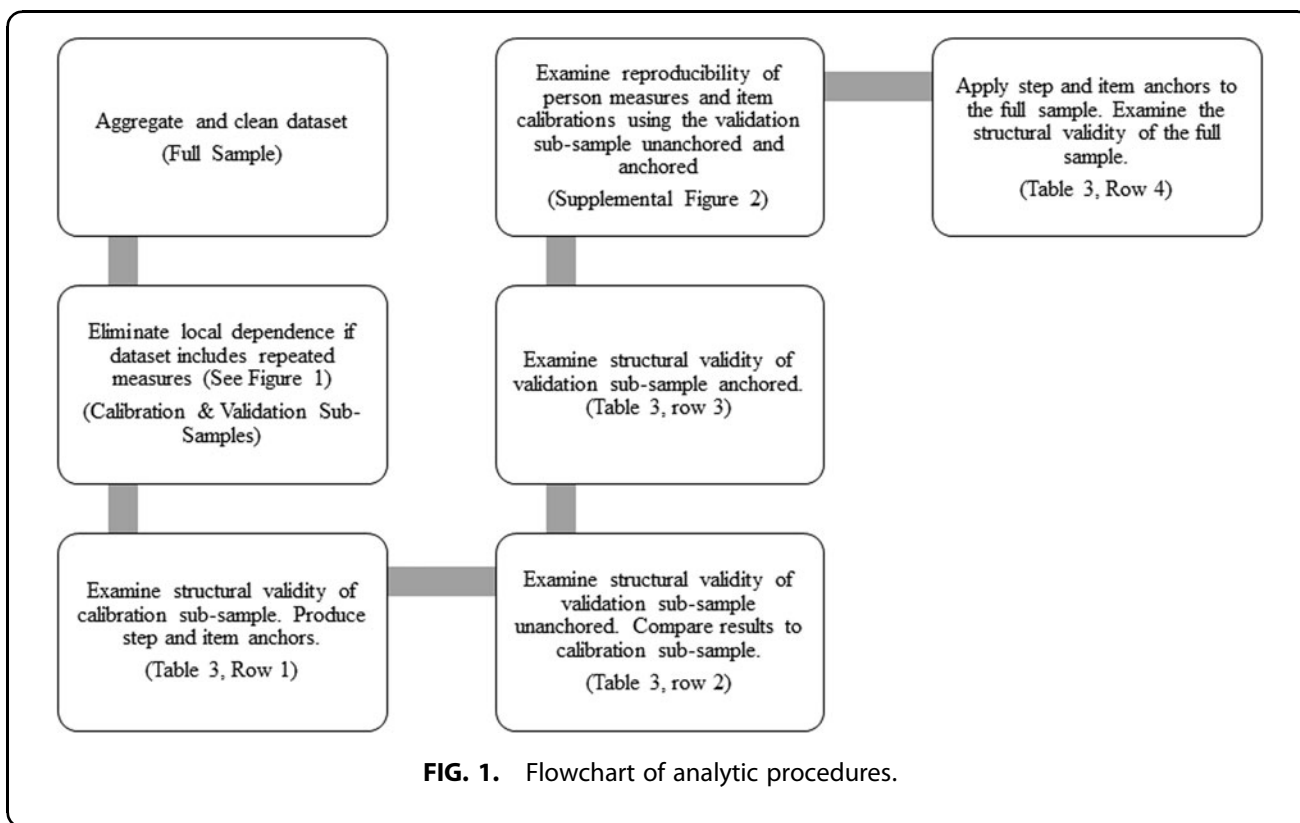
States of consciousness. The CRS-R items and rating scale categories used to diagnose VS/UWS, MCS and eMCS, based on the Aspen consensus criteria (Table 1), were applied to categorize all CRS-R records ($n=1142$) using STATA SE 14.⁹ All analytic procedures and results refers to states of consciousness, based on the Aspen consensus criteria, unless otherwise specified.

Analytic procedures

The partial credit Rasch measurement model was applied using Winsteps version 4.0.1.¹⁸ Since a score of a 1, for example, is qualitatively different for each item, the partial credit model allows for this item by item variation, enabling each item to have its own rating scale structure.¹⁹ Following the Rasch Reporting Guideline for Rehabilitation Research (RULER),^{11,12} we examined the reproducibility and structural validity of the CRS-R in order to identify the hierarchy of the CRS-R rating scale categories (Fig. 1). Reproducibility refers to whether the CRS-R assessment results are comparable across individuals. Structural validity refers to whether the items, rating scale categories, and persons cohere on the measure and reflect the requirements of the Rasch model. Second, we evaluated the extent to which Aspen consensus criteria align with the rating scale category hierarchy. Finally, we examined the association between the CRS-R Rasch person measures and states of consciousness.

Reproducibility

We addressed the potential for local dependency among persons in the full sample (since the dataset included repeated measures of the same individuals) by generating



two random subsamples—calibration and validation subsamples—in which each individual is represented once by either their first or last record.²⁰ The calibration and validation sub-samples were represented by 242 participants with either their first or last record. Twenty participants with a single record were randomly assigned for a total of 252 participants in the calibration and validation sub-sample, respectively (Fig. 2).

The calibration sub-sample was used to produce step and item anchors; these were validated with the validation sub-sample. Luppescu's method of cross-plotting person measures and item calibrations with 95% confidence intervals was used to evaluate whether there were significant deviations between the calibration and validation sub-samples.²¹ Step and item anchors were validated by evaluating item displacement >0.50 logits.²² Once validated, the step and item calibrations from the calibration sub-sample were applied as anchored values to the full sample.¹⁹

Structural validity

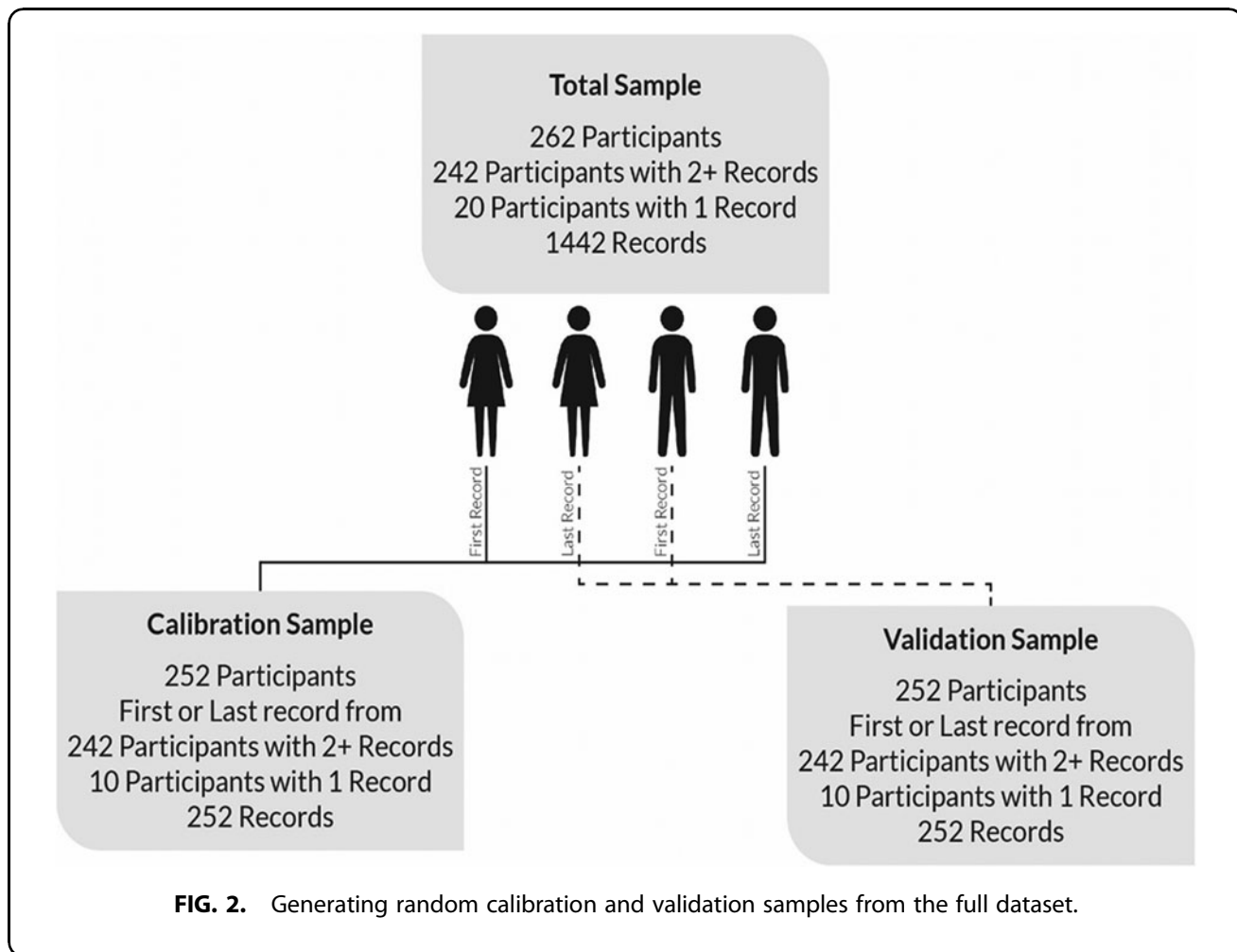
Structural validity was examined in terms of rating scale category structure, unidimensionality, hierarchical order, and measurement accuracy.

Rating scale category structure. Rating scale categories for each item were examined to ensure that each had sufficient observations and that the Andrich thresh-

olds proceeded monotonically.²³ Rating scale categories are defined by the average category difficulty measure.^{24,25} Categories with low frequencies (fewer than 10 observations) do not provide enough observations for stable category measures.¹⁹

Unidimensionality. Unidimensionality refers to the items measuring one underlying trait, neurobehavioral function, in the case of the CRS-R.¹⁹ Unidimensionality was evaluated by level of item fit, principal component analysis of residuals, and by amount of local item dependence. Items with an infit mean square >1.4 or <0.6 were considered misfitting (i.e., may not represent the same underlying trait of neurobehavioral function).²⁶ Principal component analysis of residuals (PCAR) and disattenuated correlations were also used to evaluate the extent to which items and categories share a similar underlying trait. Disattenuated correlations above 0.82 indicate items are likely measuring the same underlying trait.²⁷ We also examined the residuals of each item to determine if items are duplicative.²⁸ Local item dependence was analyzed by evaluating the inter-item correlations. Inter-item correlations >0.70 indicate local dependence which violates the assumptions of the Rasch measurement model.

To confirm item fit and PCAR, we used a more stringent technique in which we generated 10 simulated data sets based on the calibration data and fit model assumptions to identify more precise upper and lower bounds for infit mean square, Eigenvalue, and percent variance of the



first contrast.²⁹ In Winsteps, the Simulated Data File (SIFILE) output was specified based on: 1) the request for 10 data files; 2) using the data for the simulation; 3) no resampling of persons; 4) allowing for missing data to maintain the same data pattern; and 5) allowing for extreme scores. The 10 data sets were imported into STATA to calculate the more precise upper (97.5%) and lower (2.5%) bounds for infit mean square, ZSTD, Eigenvalue, and percent variance of the first contrast.

Hierarchical order. We generated logit calibrations for the average item difficulty and each rating scale category. We examined the hierarchical order of the CRS-R rating scale categories as they relate to the Aspen consensus criteria.

Measurement accuracy. The separation index and person separation reliability generated from Winsteps software program were used to examine the measurement precision and ability of the assessment to distinguish among patients with different states of consciousness. Wright's sample-independent person separation reliability (PSR) is reported for our analysis as a Shapiro Wilk test determined our data were non-normal (test

statistic=0.93; $p < 0.01$). The person strata index indicates how many statistically distinct states of consciousness the assessment can distinguish.^{30,31}

We evaluated the alignment between the distribution of persons and items by comparing the mean person measures and mean item calibrations. Ceiling and floor effects were reported to describe how well the items aligned with the range of person neurobehavioral function measures. Persons with unexpected patterns of responses were identified via infit mean squares. These unexpected patterns can often be clinically useful in identifying people with particular conditions.³²

Score-to-measure conversion. The Rasch model transforms ordinal scores into equal-interval logit measures. To enhance clinical interpretation of findings, we generated a CRS-R raw score conversion to Rasch person measures.

Statistical analysis

Alignment of Aspen consensus criteria, Rasch-based Person Measures, and CRS-R Rating Scale Categories. The distributions of the CRS-R Rasch person measures

were described by mean and standard deviation (SD) for VS/UWS, MCS, and eMCS. To confirm that the VS/UWS, MCS, and eMCS were statistically significantly different, we used a one-way analysis of variance (ANOVA). The Bartlett test indicated unequal variances; therefore, we also used a Kruskal Wallis test to describe the presence of differences in the mean CRS-R Rasch person measures across states of consciousness. The association between Rasch person measures and states of consciousness was examined via Spearman's correlation coefficient. The strength of the correlation coefficients <0.25 were interpreted as having little or no association, 0.25 to 0.50 a low to fair association, >0.50 to 0.75 a moderate to good association, and >0.75 a strong association.³³

Results

Participants by samples

Of the 262 participants, 97% ($n=254$) were receiving therapy at an intensive rehabilitation setting; seventy-three percent ($n=192$) were enrolled in a clinical trial. Participants were mostly male (70%) in a MCS (74%) after sustaining a traumatic brain injury (92%; Table 2). The average age of participants was 36.5 ± 15.2 years (range: 14-82 years).

Analytic process and reproducibility

Table 3 presents the sequence of analytic steps. During the first iteration (calibration sample; Table 3, row 1), the six CRS-R items had good precision (Wright's PSR = 0.95) and no misfitting items; inter-item correlations indicated no local item dependence.^{28,34} Items were slightly more challenging than the person ability (mean person

measure $(-0.35 \pm 1.98$ logits). Twenty-two (8.7%) of individuals reached the assessment's ceiling and there was a negligible floor effect. PCAR indicated items generally reflect the same underlying trait (Eigenvalues 1.63; percent variance of the first contrast 8.1%) and this was further confirmed via inspection of the loadings (disattenuated correlations >0.82 for all item contrasts). The second iteration, the validation sub-sample (Table 3, row 2), had comparable results to the calibration sub-sample. Therefore, for the third iteration (Table 3, row 3), we applied the step and item anchors from the calibration sub-sample to the validation sub-sample. Comparison of the person measures from the validation sample unanchored (Table 3, row 2) and anchored (Table 3, row 3) were consistent (Person $R^2=0.99$; Supplementary Fig. S1). Comparison of the item calibrations were also consistent across validation sub-samples (Items $R^2=0.98$; Supplementary Fig. S2); no displacement was greater than 0.50 logits.²² For the final iteration (Table 3, row 4), the step and item calibrations from the calibration sub-sample were applied to the full sample. All results below refer to the full sample unless otherwise specified.

Structural validity

Rating scale category structure. All rating scale categories had 10 or more responses for the calibration and full sample indicating confidence in the stability of the category measures. The validation sub-sample had less than 10 responses for rating scale category 0 on Motor ($n=6$) and Arousal ($n=8$). Rating scale categories were monotonic for all items in the calibration, validation, and full samples indicating category logits all proceeded in the same direction.

Unidimensionality. Items from the calibration, validation, and full sample each fit the measurement model with the infit mean square ranging from 0.80 to 1.36 across samples (Table 4). The 10 simulated datasets identified a more stringent infit mean square criteria range of 0.78 to 1.22 and ZSTD of -2.21 to 2.00.^{29,35} The calibration sample met this stringent infit mean square criteria with all items falling between 0.84 and 1.12. Two items misfit for the validation and full samples using the more restrictive criteria: Motor (1.31 and 1.21, respectively) and Oromotor/Verbal (1.36 and 1.27, respectively).

The PCAR Eigenvalue for the first contrast of the full sample (Table 3, row 4) was 1.61 with 8.8% unexplained variance from the first contrast, which was comparable with average values derived from the 10 simulated data sets (Eigenvalue of 1.44 and 5.4% unexplained variance in the first contrast, Supplementary Table S1).^{29,35} Disattenuated correlations were >0.85 suggesting the same underlying trait, posited to be neurobehavioral function, was being captured by the items.

Table 2. Participant Characteristics

Participant characteristic	Total n=262
Age, mean years at injury (SD)	36.5 (15.2)
Gender, n (%)	
Male	184 (70)
Female	78 (30)
Time from onset to enrollment/admission, n (%)	
Less than 28 days	3 (1)
28 to 70 days	153 (58)
71 to 112 days	41 (16)
113 to 365 days	3 (2)
366 to 730 days	7 (3)
More than 730 days	2 (1)
Missing	51 (19)
Etiology of brain injury, n (%)	
Traumatic	240 (92)
Non-traumatic	22 (8)
State of consciousness (first record), n (%)	
Emerged from minimally conscious state	21 (8)
Minimally conscious state	194 (74)
Vegetative state/unresponsive wakefulness syndrome	47 (18)

SD, standard deviation.

Table 3. Analytic Sequence Using the Rasch Partial Credit Model

Iteration	Items	Rating scale categories	Person mean (SD) logits	RMSE	Adj. SD	SI	PSR	Wright's PSR	Number of misfitting items	PCAR Eigenvalue 1 st contrast (%)	Ceiling effect n (%)	Floor effect n (%)
1. Calibration sample, 252 participants	6	29	-0.35 (1.98)	0.75	1.83	2.43	0.86	0.95	0	1.63 (8.1)	22 (8.7)	0
2. Validation sample, 252 participants	6	29	-0.15 (1.82)	0.72	1.67	2.33	0.84	0.95	0	1.71 (9.1)	18 (7.1)	0
3. Validation sample, 252 participants, anchored	6	29	-0.13 (1.89)	0.74	1.74	2.35	0.85	0.95	0	1.74 (9.5)	18 (7.1)	0
4. Full sample, 1442 records, qnchored	6	29	-0.44 (1.75)	0.72	1.59	2.20	0.83	0.95	0	1.61 (8.8)	71 (4.9)	1 (0.1)

SD, standard deviation; RMSE, root mean square standard error; SI, separation index; PSR, person separation reliability; PCAR, Principal Components Analysis of Residuals.

Hierarchical order. Item order from least to most challenging was Verbal (average item calibration -0.72), Arousal, Motor, Visual, Auditory, and then Communication (average item calibration 1.99; Table 4). The average rating scale category calibrations from least to most challenging were category 0 “None” of the Verbal item and category 6 “Functional Object Use” of the Motor item, respectively. Table 5 provides logit values (calibrations) for each rating scale category in order from least to most challenging; also indicated are the items (i.e., subscales) and state of consciousness.

Measurement accuracy. Wright’s PSR for the full sample is 0.95 and equates to 4.5 statistically different strata. The unadjusted person separation reliability was 0.83 (separation index was 2.20; Table 3).

The mean person measure was -0.44 (± 1.75) logits less than the mean item calibration. There was no appreciable floor effect (0.1%) and a minimal ceiling effect (4.9%; Table 3).^{33,36,37} Person misfit was consistent across samples at 18%, 20%, and 19% for the calibration, validation, and full samples, respectively. Upon inspection of the misfitting persons from the full samples, no pattern was found across the participant characteristics (Table 2). We examined the unexpected response patterns and the motor item was consistently reported as an unexpected response (*n* = 79, 28% of misfitting persons),

which results when the observed rating for motor (e.g., 0, 1, or 2) was lower than expected based on the participants scores on other items.

Score to measure conversion. Total CRS-R raw scores range from 0 to 23 and correspond to person measures of -6.52 to 5.85 logits. A full score-to-measure table is provided in Supplementary Table S2. Figure 3 displays a visual ruler for converting the CRS-R total raw scores into Rasch logit calibrations.

Alignment of Aspen consensus criteria with CRS-R rating scale categories and CRS-R Rasch person measures

The 15 CRS-R rating scale categories for VS/UWS have Rasch category calibrations ranging from -5.66 to 1.2 logits (Table 5). The 11 CRS-R rating scale categories for MCS have Rasch category calibrations ranging from -0.59 to 4.13 logits. The two CRS-R rating scale categories for eMCS had Rasch category calibrations ranging from 4.06 to 4.49 logits. Four rating scale categories reflecting VS/UWS had higher logit calibrations than the lowest MCS rating scale category (2 “Fixation” of the Visual item; Table 5). Two rating scale categories, 4 “Consistent Movement to Command” of the Auditory item reflecting MCS and 3 “Attention” of the Arousal item (not aligned to the Aspen consensus criteria), had higher logit calibrations than the lowest eMCS rating scale category (2 “Functional: Accurate” of the Communication item; Table 5). Rating scale category 3 “Intelligible Verbalization” of the Oromotor/Verbal item reflecting MCS was within 0.25 logits of the lowest eMCS rating scale category (2 “Functional: Accurate” of the Communication item; Table 5) indicating comparable difficulty.

CRS-R person measures were summarized for each state of consciousness: VS/UWS mean -2.0±0.87 SD, range -5.12 to 1.19 logits (raw score 0 to 16), MCS mean -0.01±1.00SD, range -2.88 to 3.63 logits (raw score 4 to 21), and eMCS mean 2.65±1.86 SD, range

Table 4. Item Calibrations and Fit Statistics Arranged in Hierarchical Order from Most to Least Challenging

Items	Measure (logits)	Std. error	Infit MnSq	Infit zstd	Outfit MnSq	Outfit zstd	Disp.
Communication	1.99	0.06	0.90	-2.0	0.68	-3.4	0.05
Auditory	-0.05	0.04	0.80	-5.5	0.75	-6.9	-0.12
Visual	-0.08	0.03	0.90	-2.6	0.86	-3.2	0.04
Motor	-0.48	0.03	1.21	4.4	1.33	6.0	0.05
Arousal	-0.66	0.05	0.99	-0.2	1.01	0.4	-0.11
Oromotor/Verbal	-0.72	0.05	1.27	6.6	1.28	6.4	0.06

*Data from full sample using anchors from calibration sample. Std., standard; MnSq, mean square; zstd, Z standard; Disp., displacement.

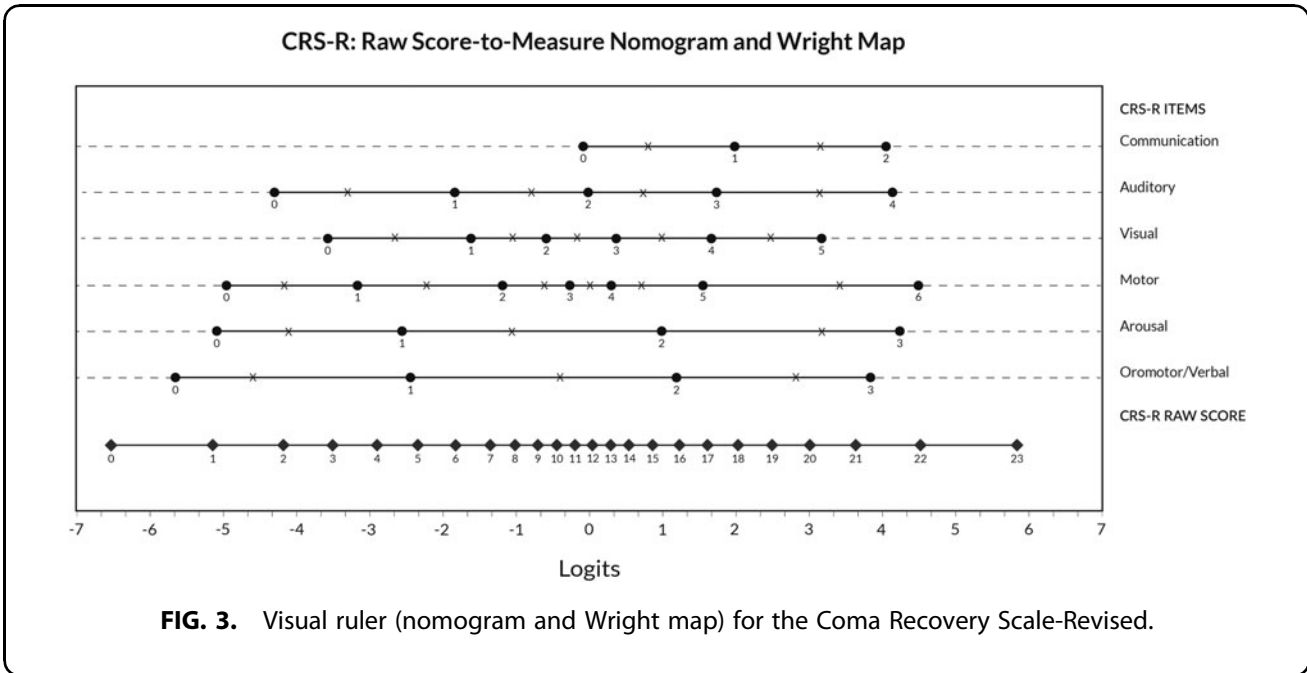


FIG. 3. Visual ruler (nomogram and Wright map) for the Coma Recovery Scale-Revised.

-0.43 to 5.84 logits (raw score 10-23; Fig. 4; Table 6). The Bartlett test from the one-way ANOVA ($F=453.5$, $df=2$; $p=0.00$) confirmed the variances of the CRS-R person measures were statistically different for each state of consciousness (VS/UWS, MCS, and eMCS; $X^2=20.43$; $p<0.001$) thus we conducted an equivalent non-parametric test; the non-parametric Kruskal-Wallis test indicated the mean ranks of CRS-R person measures were statistically different across each state of conscious-

ness ($H(2)=194.74$; $p<0.01$). Correlation between states of consciousness and CRS-R person measures indicates a strong relationship ($r_s=0.86$, $p<0.01$).³³

Discussion

The empirical evaluation of the CRS-R rating scale categories, as they are operationalized to the Aspen consensus criteria, indicated the order of the 29 rating scale category calibrations reflected current literature

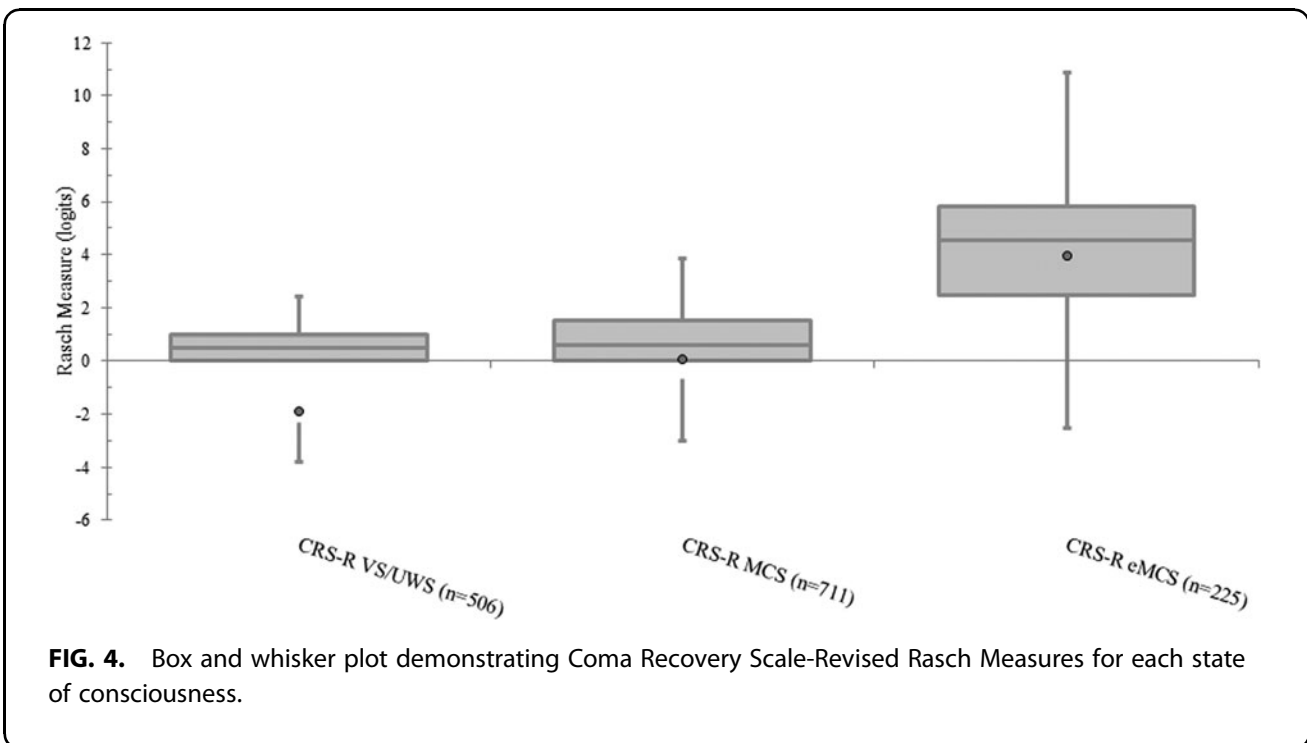


FIG. 4. Box and whisker plot demonstrating Coma Recovery Scale-Revised Rasch Measures for each state of consciousness.

regarding the continuum of neurobehavioral function: category 6 “Functional Object Use” of the Motor item was hardest for patients to achieve; category 0 “None” of the Oromotor/Verbal item was easiest to achieve. Although the pattern and sequence of CRS-R items and rating scale categories has a hierarchical order; the motor item may be more prone to unexpected lower scores (i.e., more misfitting persons) due to neuromuscular impairments such as hypertonicity and spasticity that can confound assessment of consciousness. Six categories do not occur in the expected sequential hierarchical order. Two rating scale categories, one reflecting MCS and one not operationalized to the Aspen consensus criteria, had higher calibrations than the easiest eMCS item (category 2 “Functional: Accurate” of the Communication item). A third rating scale category was within 0.25 logits of the easiest eMCS item; rating scale categories that are close together on the hierarchy are of comparable difficulty and reflect a similar level of neurobehavioral function. There are also four rating scale categories reflecting the VS/UWS that had higher calibrations (reflecting greater neurobehavioral function) than the easiest MCS item (category 2 “Fixation” of the Visual item; Fig. 3; Table 5). CRS-R person measures (indicating amount of neurobehavioral function) and states of consciousness, based on Aspen consensus criteria, showed a strong correlation ($r_s = 0.86$, $p < 0.01$).

The three rating scale categories near the two eMCS categories all had average category calibrations greater than the mean person measure for eMCS (3.65 logits; Table 6). Further, the rating scale hierarchy exhibited category 4 “Consistent Movement to Command” of the Auditory item to be slightly more challenging than category 2 “Functional: Accurate” of the Communication item and less challenging than category 6 “Functional Object Use” of the Motor item. Prior work established that category 4 “Consistent Movement to Command” of the Auditory item is a behavior that occurs approximately at the same time as the two categories reflecting eMCS.^{10,38} Our study substantiates this finding when examining average category difficulty and category 4 “Consistent Movement to Command” of the Auditory item should be included in the diagnostic criteria for eMCS.

This is the first study to demonstrate that category 3 “Attention” of the Arousal item and category 3 “Intelligible Verbalization” of the Oromotor/Verbal item may also reflect comparable ability to other eMCS categories. For a patient to achieve category 3 “Attention” of the Arousal item, the patient must respond to all but three of the verbal or gestural prompts demonstrating sustained attention and consistency throughout the administration of the CRS-R. For a patient to achieve category 3 “Intelligible Verbalization” of the Oromotor/Verbal item, the patient must be able to vocalize, write, or use an alphabet

Table 5. Describing Mean Category Difficulty of CRS-R Rating Scale Categories by Item Operationalized to States of Consciousness Based on the Aspen Consensus Criteria

<i>Item</i>	<i>Rating scale category</i>	<i>Rating scale category</i>	<i>Mean category difficulty</i>	<i>Operationalization to Aspen criteria (Schnakers et al, 2009)</i>
Oromotor/Verbal	0	None	-5.66	Vegetative State/Unresponsive Wakefulness Syndrome
Arousal	0	None	-5.09	Vegetative State/Unresponsive Wakefulness Syndrome
Motor	0	None	-4.95	Vegetative State/Unresponsive Wakefulness Syndrome
Auditory	0	None	-4.3	Vegetative State/Unresponsive Wakefulness Syndrome
Visual	0	None	-3.57	Vegetative State/Unresponsive Wakefulness Syndrome
Motor	1	Abnormal Posturing	-3.16	Vegetative State/Unresponsive Wakefulness Syndrome
Arousal	1	Eye Opening with Stimulation	-2.56	Vegetative State/Unresponsive Wakefulness Syndrome
Oromotor/Verbal	1	Oral Reflexive Movement	-2.44	Vegetative State/Unresponsive Wakefulness Syndrome
Auditory	1	Auditory Startle	-1.83	Vegetative State/Unresponsive Wakefulness Syndrome
Visual	1	Visual Startle	-1.62	Vegetative State/Unresponsive Wakefulness Syndrome
Motor	2	Flexion Withdrawal	-1.19	Vegetative State/Unresponsive Wakefulness Syndrome
Visual	2	Fixation	-0.59	Minimally Conscious State
Motor	3	Localization to Noxious Stimuli	-0.27	Minimally Conscious State
Communication	0	None	-0.08	Vegetative State/Unresponsive Wakefulness Syndrome
Auditory	2	Localization to Sound	-0.02	Vegetative State/Unresponsive Wakefulness Syndrome
Motor	4	Object Manipulation	0.3	Minimally Conscious State
Visual	3	Visual Pursuit	0.37	Minimally Conscious State
Arousal	2	Eye Opening Without Stimulation	0.99	Vegetative State/Unresponsive Wakefulness Syndrome
Oromotor/Verbal	2	Vocalization/Oral Movement	1.2	Vegetative State/Unresponsive Wakefulness Syndrome
Motor	5	Automatic Motor Response	1.57	Minimally Conscious State
Visual	4	Object Localization: Reaching	1.67	Minimally Conscious State
Auditory	3	Reproducible Movement to Command	1.74	Minimally Conscious State
Communication	1	Non-functional: Intentional	1.99	Minimally Conscious State
Visual	5	Object Recognition	3.18	Minimally Conscious State
Oromotor/Verbal	3	Intelligible Verbalization	3.83	Minimally Conscious State
Communication	2	Functional: Accurate	4.06	Emerged from Minimally Conscious State
Auditory	4	Consistent Movement to Command	4.13	Minimally Conscious State
Arousal	3	Attention	4.25	Not aligned to criteria
Motor	6	Functional Object Use	4.49	Emerged from Minimally Conscious State

Rating scale categories at the top indicate less neurobehavioral function and categories at the bottom indicate more neurobehavioral function.

Table 6. Descriptive Statistics Based on the First and Last Record for Each Participant Using the CRS-R Person Measures (Logits)

Aspen criteria	Record	N	Median	Mean (95% CI)	SE	Variance
eMCS	First	7	0.85	1.11 (-0.02, 2.24)	0.46	1.49
	Last	93	4.52	4.37 (4.04, 4.69)	0.16	2.46
	All	225	3.63	3.65 (3.40, 3.89)	.12	3.45
MCS	First	164	-0.19	-0.14 (-0.29, .01)	0.08	0.93
	Last	106	0.16	0.36 (0.14, 0.57)	0.11	1.27
	All	711	-0.19	-0.01 (-.08, .06)	0.04	1.0
VS/UWS	First	93	-2.33	-2.34 (-2.54, -2.15)	0.10	0.91
	Last	45	-1.81	-2.11 (-2.38, -1.84)	0.13	0.81
	All	506	-1.81	-2.0 (-2.08, -1.92)	0.04	0.76

CRS-R, Coma Recovery Scale-Revised; CI, confidence interval; SE, standard error; eMCS, emerged from minimally conscious state; MCS, minimally conscious state; VS, vegetative state; UWS, unresponsive wakefulness syndrome.

board to communicate two words with a consonant-vowel-consonant triad. Patients who are able to achieve these rating scale categories are demonstrating an ability level that is similar to a patient demonstrating functional communication. These empirical findings suggest these additional categories are also indicative of eMCS and warrant further substantiation in future studies.

Four rating scale categories that reflect VS/UWS covered ranges of the continuum that overlapped with the range of some MCS rating scale categories. These four rating scale categories include: 0 “None” of the Communication item; 2 “Eye Opening Without Stimulation” of the Arousal item; 2 “Vocalization/Oral Movement” of the Oromotor/Verbal item; and 2 “Localization to Sound” of the Auditory item. However, because the range of these categories were wide and most of the range aligned with other VS/UWS categories, it is likely that only category “2 Localization to Sound” of the Auditory item is really indicative of MCS; whereas, the other three rating scale categories may be indicative of VS/UWS and/or MCS dependent upon the patient’s behavior.

Category 2 “Localization to Sound” of the Auditory item is likely between -0.6 to 0.83 logits (Fig. 3), which is of similar difficulty to the range for category 2 “Fixation” of the Visual item (-0.89 to -0.09 logits; Fig. 3). “Localization to Sound” requires the patient to orient towards the auditory stimulus twice in at least one direction demonstrating awareness and a behavior in response to a specific stimuli, a key feature of MCS diagnostic criteria.³ Our study further substantiates prior evidence that found localization to sound to be reflective of higher order processing.³⁹ Thus, empirically and qualitatively category 2 “Localization to Sound” of the Auditory item should be included in the diagnostic criteria for MCS.

The other three rating scale categories reflective of VS/UWS have average calibrations within the range of both VS/UWS and MCS. Each covers a wide range of more than 3 logits: category 0 “None” of the Commu-

nication item ranges from -6.52 to 0.85 logits, category 2 “Eye Opening Without Stimulation” of the Arousal item from -1.05 to 3.32 logits, and category 2 “Vocalization/Oral Movement” of the Oromotor/Verbal item from -0.51 to 2.88 logits. These wide ranges reflect there is a range of person ability. Of note, the Communication item is only scored when there is evidence of command following on the Auditory item (e.g., rating scale categories 3 and 4, a rating scale category of 3 is achieved on the Oromotor/Verbal item, or if there is evidence of spontaneous communication).¹⁷ Therefore, while the average category measure aligns with MCS categories, category 0 “None” of the Communication item does not qualitatively describe MCS behavior. Similarly, category 2 “Eye Opening Without Stimulation” of the Arousal item may reflect patients in a VS/UWS that continuously have their eyes open, patients in MCS who are able to have their eyes open and attend to some verbal or gestural prompts, and patients eMCS. Lastly, category 2 “Vocalization/Oral Movement” of the Oromotor/Verbal item includes patients demonstrating non-reflexive oral movements and those expressing one intelligible word that is contingent or spontaneous.¹⁷ These rating scale categories require further investigation to better align them to an appropriate diagnostic category. The wide logit ranges for these three category measures suggests the need to split the category, which may help better distinguish patient behaviors reflective of VS/UWS and MCS.

Our work determined that patients categorized as VS/UWS, MCS, and eMCS are distinctly different groups when measured by Rasch analysis. We examined the CRS-R Rasch person measures relative to the states of consciousness and found a strong positive correlation, providing further empirical support for using the CRS-R for diagnostic purposes. This study provides empiric evidence that the CRS-R is useful for diagnosis and that additional CRS-R rating scale categories should be considered for diagnosing MCS and eMCS.

Limitations and future research

The present study is a retrospective analysis of CRS-R data from individuals with DoC after brain injury who were receiving inpatient rehabilitation services or participating in a clinical trial. Our analysis did not include patients in a comatose state, which is the lowest level on the continuum of DoC. The previous study that operationalized the CRS-R rating scale categories to the Aspen consensus criteria⁹ did not consider the comatose state, limiting the ability to identify patients who transition from comatose to VS/UWS. Further, we were unable to examine differential item functioning as the sample size of our demographic characteristics (e.g., etiology, gender, and state of consciousness) did not meet the minimum threshold for analysis.^{12,40} Undetected differential item functioning could influence item locations if it is

substantial; differential item functioning tends to have the greatest impact on items at the ends of the scale. Interrater reliability of each CRS-R rating scale category was not examined; it is possible rating scale categories with better interrater reliability may reflect increased confidence in scoring particular behaviors reflective of a particular state of consciousness. We did not evaluate the influence of rater severity/leniency, a rater who consistently scores more severely or more leniently, on person measures.⁴¹

Future research should examine rater severity/leniency and interrater reliability as these may impact which rating scale category is selected and therefore influence diagnosis. Future research should examine whether there is differential item functioning across etiology, gender, and state of consciousness once each subgroup has at least 100 participants.⁴⁰ Future research is also needed to examine cut-points for the transition from VS/UWS to MCS and MCS to eMCS based on the Rasch measures. The hierarchy of the rating scale categories may be impacted by the administration guidelines (e.g., administer the communication item only when certain criteria have been achieved). Therefore, a future study should replicate these findings when all rating scale categories are administered for each item.

Conclusion

Accurately diagnosing disorders of consciousness following a severe brain injury is important as it relates to clinical decision making. This Rasch analysis indicated a hierarchy of the CRS-R rating scale categories that support a unidimensional construct of neurobehavioral function. The CRS-R's high person separation reliability in this analysis indicates it is sufficiently precise for making reliable and consistent individual-level decisions. The strong association between the CRS-R person measures and states of consciousness further supports the use of the CRS-R for diagnostic purposes. Our study provides empirical evidence for revising the diagnostic criteria for MCS and eMCS. A patient achieving category 2 "Localization to Sound" of the Auditory item is empirically and qualitatively indicative of MCS and a patient achieving category 4 "Consistent Movement to Command" of the Auditory item is indicative of eMCS. The CRS-R should be used in lieu of unstructured clinical observations for evaluating disorders of consciousness when critical decisions about care are being made.

Acknowledgments

This work was completed in partial fulfillment of the requirements for a PhD in Translational Health Science by Dr. Jennifer A. Weaver.

We would like to thank Alison McGuire, MA, Instructional Technologist at the George Washington University School of Medicine and Health Sciences, for her contributions on Figures 2 and 3.

Authors' Contributions

JW: Conceptualization (lead); data curation (lead); formal analysis (lead); project administration (lead); software (lead); writing original draft (lead); writing/review and editing (lead).

AC: Formal analysis (supporting), writing/reviewing and editing (supporting).

KO: Resources (supporting); writing/reviewing and editing (supporting).

PH: Resources (supporting); writing/reviewing and editing (supporting).

JG: Resources (lead); writing/reviewing and editing (supporting).

JW: Resources (lead); writing-reviewing and editing (supporting).

TBP: Resources (supporting); writing/reviewing and editing (supporting).

PVDW: Conceptualization (supporting); supervision (supporting); writing/reviewing and editing (supporting).

TM: Conceptualization (supporting); formal analysis (supporting); methodology (lead); supervision (lead); writing/reviewing and editing (supporting).

Funding Information

This work was partially supported by the U.S. Department of Defense under Grant W81XWH-14-1-0568; U.S. Department of Defense under Grant JW150040.

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Table S1
Supplementary Table S2

References

- Giacino JT, Whyte J, Nakase-Richardson R, et al. Minimum competency recommendations for programs that provide rehabilitation services for persons with disorders of consciousness: a position statement of the American Congress of Rehabilitation Medicine and the National Institute on Disability, Independent Living and Rehabilitation Research Traumatic Brain Injury Model Systems. *Arch Phys Med Rehabil* 2020;101(6):1072–1089.
- Giacino JT, Katz DI, Schiff ND, et al. Practice guideline update recommendations summary: Disorders of consciousness: report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology* 2018;91(10):450–460.
- Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state: definition and diagnostic criteria. *Neurology* 2002;58(3):349–353.
- Giacino JT. The minimally conscious state: defining the borders of consciousness. *Prog Brain Res* 2005;150:381–395.

5. Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004;85(12):2020–2029.
6. Giacino JT, Kezmaryski MA, DeLuca J, et al. Monitoring rate of recovery to predict outcome in minimally responsive patients. *Arch Phys Med Rehabil* 1991;72(11):897–901.
7. Gerrard P, Zafonte R, Giacino JT. Coma Recovery Scale-Revised: evidentiary support for hierarchical grading of level of consciousness. *Arch Phys Med Rehabil* 2014;95(12):2335–2341.
8. La Porta F, Caselli S, Ianes AB, et al. Can we scientifically and reliably measure the level of consciousness in vegetative and minimally conscious states? Rasch analysis of the coma recovery scale-revised. *Arch Phys Med Rehabil* 2013;94(3):527–535.e521.
9. Schnakers C, Vanhaudenhuysse A, Giacino J, et al. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*. 2009;9:35.
10. Golden K, Erler KS, Wong J, et al. Should consistent command-following be added to the criteria for emergence from the minimally conscious state? *Arch Phys Med Rehabil* 2022;doi: 10.1016/j.apmr.2022.03.010. Online ahead of print.
11. Mallinson T, Kozlowski AJ, Johnston MV, et al. Rasch Reporting Guideline for Rehabilitation Research (RULER): the RULER Statement. *Arch Phys Med Rehabil* 2022;doi: 10.1016/j.apmr.2022.03.013. Online ahead of print.
12. Van de Winckel A, Kozlowski AJ, Johnston MV, et al. Reporting Guideline for RULER: Rasch reporting guideline for rehabilitation research—explanation and elaboration. *Arch Phys Med Rehabil* 2022;doi: 10.1016/j.apmr.2022.03.019. Online ahead of print.
13. Giacino JT, Whyte J, Bagiella E, et al. Placebo-controlled trial of amantadine for severe traumatic brain injury. *N Engl J Med* 2012;366(9):819–826.
14. Pape, T. National Library of Medicine. rTMS: a treatment to restore function after severe TBI. 2015. Available from: <https://clinicaltrials.gov/ct2/show/study/NCT02366754> [Last accessed May 26, 2022].
15. Wang J, Hu X, Hu Z, et al. The misdiagnosis of prolonged disorders of consciousness by a clinical consensus compared with repeated coma-recovery scale-revised assessment. 2020;20(1):343.
16. O'Dell MW, Jasin P, Lyons N, et al. Standardized assessment instruments for minimally-responsive, brain-injured patients. *NeuroRehabilitation*. 1996;6(1):45–55.
17. Bodien Y, Chatelle C, Giacino J. Spaulding Rehabilitation Network. Coma Recovery Scale-Revised Administration and Scoring Guidelines. 2020. Available from: <https://www.sralab.org/rehabilitation-measures/coma-recovery-scale-revised> [Last accessed January 18, 2022].
18. Linacre JM, Wright BD. WINSTEPS: Rasch Model Computer Program (Version 4.0.1). Mesa Press. 2017. Available from: <http://www.winsteps.com/winsteps.htm> [Last accessed May 26, 2022].
19. Bond TG, Yan Z, Heene M. *Applying the Rasch Model Fundamental Measurement in the Human Sciences*, fourth ed. Routledge: Routledge: New York, NY; 2020.
20. Mallinson T. Rasch analysis of repeated measures. *Rasch Meas Trans* 2011(251):1317.
21. Luppescu S. Comparing measures. *Rasch Meas Trans* 1995;9(1):410–411.
22. Linacre JM. Displacement measures. Help for Winsteps Rasch Measurement and Rasch Analysis Software. 2019. Available from: <https://www.winsteps.com/winman/displacement.htm> [Last accessed May 10, 2019].
23. Linacre JM. A User's Guide to WINSTEPS & MINISTEP Rasch-Model Computer Programs. Program Manual 4.3.1. 2018. Available from: <http://www.winsteps.com/winman> [Last accessed May 26, 2022].
24. Linacre JM. Rating scale conceptualization: Andrich, Thurstonian, half-point thresholds. *Winsteps Manual*. 2021. Available from: <https://www.winsteps.com/winman/ratingscale.htm> [Last accessed February 14, 2021].
25. Linacre JM. Demarcating category intervals: where are the category boundaries on the latent variable? *Rasch Meas Trans* 2006;19(3):3.
26. Wright BD, Linacre, J. M., Gustafson, et al. Reasonable mean-square fit values. *Rasch Meas Trans* 1994;8(3):370.
27. Linacre JM. Dimensionality Investigation—An Example. Help for Winsteps Rasch Measurement and Rasch Analysis Software. 2019. Available from: <https://www.winsteps.com/winman/multidimensionality.htm> [Last accessed May 10, 2019].
28. Linacre JM. Table 23.99. Largest residual correlations for items. 2020. Available from: https://www.winsteps.com/winman/table23_99.htm [Last accessed March 26, 2020].
29. Smith RM, Schumacker RE, Bush MJ. Using item mean squares to evaluate fit to the Rasch model. *J Outcome Meas* 1998;2(1):66–78.
30. Wright BD. Reliability and separation. *Rasch Meas Trans* 1996;9(4):472.
31. Wright BD, Masters GN. Number of person or item strata. *Rasch Meas Trans* 2002;16(3):888.
32. Smith RM. Person Fit in the Rasch Model. *Educ Psychol Meas* 1986;46(2):359–372.
33. Portney LG. FOUNDATIONS OF CLINICAL RESEARCH : applications to practice. Place of publication not identified: F A DAVIS; 2019.
34. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57(8):1358.
35. Linacre JM. Simulated file specifications. Help for Winsteps Rasch Measurement and Rasch Analysis Software. Available from: <https://www.winsteps.com/winman/simulated.htm> [Last accessed May 10, 2019].
36. Andresen EM, Rothenberg BM, Panzer R, et al. Selecting a generic measure of health-related quality of life for use among older adults: a comparison of candidate instruments. *Eval Health Professions* 1998;21(2):244–264.
37. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000;81:515–520.
38. Schnakers C, Giacino J. Poster 21: Criteria for emergence from the minimally conscious State: what about consistent command-following? *Arch Phys Med Rehabil* 2009;90(10):e18.
39. Carrière M, Cassol H, Aubinet C, et al. Auditory localization should be considered as a sign of minimally conscious state based on multimodal findings. *Brain Commun*. 2020;2(2):fcaa195.
40. Myers ND, Wolfe EW, Feltz DL, et al. Identifying differential item functioning of rating scale items with the Rasch Model: an introduction and an application. *Meas Phys Educ Exerc Sci* 2006;10(4):215–240.
41. Linacre JM, Engelhard G, Tatum DS, et al. Measurement with judges: many-faceted conjoint measurement. *Int J Educ Res*. 1994;21(6):569–577.