Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2014;95:2335-41

ORIGINAL ARTICLE

Coma Recovery Scale—Revised: Evidentiary Support for Hierarchical Grading of Level of Consciousness

Paul Gerrard, MD, Ross Zafonte, DO, Joseph T. Giacino, PhD

From Spaulding Rehabilitation Hospital and Harvard Medical School, Boston, MA.

Abstract

Objective: To investigate the neurobehavioral pattern of recovery of consciousness as reflected by performance on the subscales of the Coma Recovery Scale—Revised (CRS-R).

Design: Retrospective item response theory (IRT) and factor analysis.

Setting: Inpatient rehabilitation facilities.

Participants: Rehabilitation inpatients (N=180) with posttraumatic disturbance in consciousness who participated in a double-blinded, randomized, controlled drug trial.

Interventions: Not applicable.

Main Outcome Measures: Scores on CRS-R subscales.

Results: The CRS-R was found to fit factor analytic models adhering to the assumptions of unidimensionality and monotonicity. In addition, subscales were mutually independent based on residual correlations. Nonparametric IRT reaffirmed the finding of monotonicity. A highly constrained confirmatory factor analysis model, which imposed equal factor loadings on all items, was found to fit the data well and was used to estimate a 1-parameter IRT model.

Conclusions: This study provides evidence of the unidimensionality of the CRS-R and supports the hierarchical structure of the CRS-R subscales, suggesting that it is an effective tool for establishing diagnosis and monitoring recovery of consciousness after severe traumatic brain injury.

Archives of Physical Medicine and Rehabilitation 2014;95:2335-41

The measurement of level of consciousness is a difficult but crucial aspect of diagnostic and prognostic assessment of persons with disorders of consciousness (DOC). Estimates of misdiagnosis in this population consistently fall within the 30% to 45% range.[1-3] Diagnostic error may result from biases contributed by the examiner, patient, and environment.[1] Examiner error may arise when the range of behaviors sampled is too narrow, response-time windows are over- or underinclusive, criteria for judging purposeful responses are poorly defined or not adhered to, and examinations are conducted too infrequently to capture the full range of behavioral fluctuation. The second source of variance concerns the patient. Fluctuations in arousal level, fatigue, subclinical seizure activity, occult illness, pain, cortical sensory deficits (eg, cortical blindness/deafness), motor impairment (eg, generalized hypotonus, spasticity, or paralysis), or cognitive (eg, aphasia, apraxia, agnosia) disturbance can conspire to confound accurate diagnostic assessment, constitute a bias to the behavioral assessment, and therefore decrease the probability to observe signs of consciousness. Finally, the environment in which the patient is evaluated may bias assessment findings. Paralytic and sedating medications, restricted range of movement stemming from restraints and immobilization techniques, poor positioning, and excessive ambient noise, heat, or light can decrease or distort voluntary behavioral responses.

Accurate evaluation requires well-validated and reliable measurement tools. Since consciousness itself is a nebulous concept, efforts to develop effective assessment methods typically begin with an a priori operational definition of the construct of consciousness. Frameworks for describing consciousness have been previously proposed based on neuroanatomic, philosophical, and even computational criteria.[4-10] However, such explanations have

The data used in this article were extracted from a database developed with grant support from the National Institute on Disability and Rehabilitation Research (NIDRR), United States Department of Education (grant no. H133A031713: JFK-Johnson Rehabilitation Institute TBI Model System). Contributions were partially supported by NIDRR grant no. H133A120085 (Spaulding-Harvard TBI Model System). The contents do not necessarily represent the policy of the Department of Education, and endorsement by the Federal Government should not be assumed.

Disclosures: Giacino has served as an expert witness/consultant on 4 legal cases over the last 36 months involving patients with disorders of consciousness concerning diagnosis, prognosis, pain and suffering, and adequacy of treatment. The other authors have nothing to disclose.

0003-9993/14/$36 - see front matter © 2014 by the American Congress of Rehabilitation Medicine

http://dx.doi.org/10.1016/j.apmr.2014.06.018

limited practical use in clinical assessment. An alternative approach to characterizing the construct of consciousness involves empirically identifying a set of behaviors that represent levels of neurologic function along the continuum of consciousness. While this strategy does not have the theoretical rigor that may be seen in computational or philosophical criteria, it has the advantage of providing a clinically useful approach that can guide diagnostic decision-making.

Recently, the American Congress of Rehabilitation Medicine conducted an evidence-based review[11] of assessment scales designed specifically for use in persons with DOC. The authors concluded that among the 13 assessment scales reviewed, only the Coma Recovery Scale−Revised (CRS-R) had sufficient psychometric properties to be recommended for use in clinical practice with minor reservations. The CRS-R is a standardized measure of neurobehavioral function that has been widely used for diagnostic assessment and outcome measurement in studies involving persons with DOC.[3,12-15] It consists of 23 hierarchically arranged items that comprise 6 subscales designed to assess arousal level, audition and language comprehension, expressive speech, visuoperceptual abilities, motor functions, and communication ability. Scoring is based on the presence or absence of behavioral responses to stimuli presented in a standardized manner. The lowest item on each subscale represents reflexive behavior, while the highest item reflects cognitively mediated activity.

The examiner presents a stimulus according to standardized instructions and scores the response against predefined criteria. If an item is "failed," the examiner progresses to the next item down, continuing this process until a scorable response is obtained. For example, in an awake and fully conscious patient, only the first (ie, highest level) item on each subscale would be administered, as the corresponding behavioral response would be expected to reflect cognitively mediated activity. In contrast, in a patient with impaired brainstem function, the examiner would likely administer all items within a particular subscale, because the corresponding higher-level neurobehavioral responses would not occur. A score is assigned for each subscale based on the highest-level behavior observed. The lowest score on all subscales is 0, and the maximum ranges from 2 (communication subscale) to 6 (motor subscale). Higher scores are intended to indicate higher levels of neurologic function. Notably, some subscales include pathologic behaviors (eg, abnormal posturing) that are expected to be extinguished at higher levels of consciousness. The term *subscale* as it is used in the CRS-R refers to the item response theory (IRT) notion of an item. Thus, we refer to the subscales of the CRS-R as "items" and the individual stimulus-response pairs as "response categories." The 6 items of the CRS-R and the behavioral response categories for each item are provided in figure 1.

Identification of the underlying construct represented by the CRS-R would yield not only a quantitative measure of consciousness but also possible operational definitions for discrete levels of consciousness. The manner in which the CRS-R is administered relies on a theoretical hierarchy of neurobehavioral responses, which is in part derived from analysis of the original CRS-R.[13,16] This hierarchy rests on the assumption that behaviors considered higher level by the test do indeed correspond to a higher level of neurologic functioning and that if persons are able to demonstrate higher-level behaviors, they also either are able to demonstrate the lower-level behaviors or have progressed to a level of consciousness where such behaviors have extinguished (as is the case with pathologic behaviors). In general, evidence of construct validity is sought by determining whether the outcome measure of interest has a construct that behaves in the expected manner. Psychometricians have described 2 types of construct validity: weak validity and strong validity. Weak validity is established by a correlation with some external criterion, whereas strong validity is established by testing a well-formulated hypothesis that should explain the observed scores on the instrument.[17] To provide evidence for strong construct validity on a unidimensional assessment scale, the constituent items should demonstrate unidimensionality, monotonicity, mutual independence, and invariant item ordering.[18-20] Unidimensionality refers to the fact that a scale represents a single latent construct. Monotonicity asserts that as a respondent's score on the test increases, the expected score on any single item should increase or at least remain stable. Mutual independence holds that the only source of correlation in scores between any 2 (or more) items on a given scale should be the underlying construct that is being measured by the scale as a whole. Invariant item ordering, sometimes also referred to as the "nonintersection of the item response curves," refers to the notion that for any given ability level, the order of difficulty of items should remain the same.

To test these properties of the CRS-R individually, we applied a series of psychometric models to the CRS-R in a graded fashion from least to most restrictive as follows. We first applied kernel density smoothing IRT[21] (KSIRT) to ensure that the assumption of monotonicity was met, providing evidence of the hierarchical structure of the scale. If this assumption was not met, further analysis would not have been appropriate. Once monotonicity was established, we obtained polychoric correlations to explicitly model the ordinal CRS-R data as monotonic continuous data. We then used these polychoric correlations as input to factor analyses, further exploring construct validity and the hierarchical composition of the CRS-R items. First, exploratory factor analysis (EFA) was performed to test the adequacy of a single dimension to explain the observed data and look for evidence of local independence. Once unidimensionality and local independence were established, we tested the assumption of invariant item ordering. This was accomplished using confirmatory factor analysis, constraining item loadings to be equal.

The rationale for this approach is that each psychometric method has different constraints. For example, in the recent IRT analysis of the CRS-R by La Porta et al,[22] the Rasch model was applied. This particular psychometric approach imposes that all items have the same discrimination parameter, and the estimation algorithms are typically based on a maximum likelihood approach. This is in contrast to other methods for handling ordinal data such as nonparametric IRT models, IRT models that allow discrimination parameters to vary between items, and factor analytic methods. Because it is not possible to know for certain which psychometric model best represents any given construct,

List of abbreviations:

CFA	**confirmatory factor analysis**
CRS-R	**Coma Recovery Scale−Revised**
DOC	**disorders of consciousness**
EFA	**exploratory factor analysis**
IRT	**item response theory**
KSIRT	**kernel density smoothing item response theory**
SRMR	**standardized root mean square of the residuals**
TLI	**Tucker-Lewis Index**