Framework Guiding Reporting of Rehabilitation Assessment Studies Based on Rasch Analysis

	Conceptual/Content Validity	Structural Validity	External Validity	Consequential Validity	Reproducibility & Reliability	Practical applications and clinical implementation
Scope of Psychometric Domain	Considers the extent to which items represent the underlying construct and the existing evidence that those items reflect the underlying construct. Considers the extent to which the construct is underrepresented by the content of the items or to which assessment responses reflect processes other than the underlying construct. Considers to what extent stakeholders believe the items are relevant to and representative of the domain.	Refers to the extent to which items, rating scale steps/thresholds and people cohere to form a measure that substantively reflects the assumptions of the Rasch model. The key assumptions of the Rasch model are: - Unidimensionality: items all measure the same underlying trait - Hierarchical order: harder items are harder for everyone and more able people score higher on all items	Refers to the extent that measures from the assessment of interest align with measures from similar and/or different assessments. Also refers to the ability of the measures to inform current and future outcomes. Indicates that performance on the assessment relates to skill/knowledge in the target domain. The extent to which the measures are associated with external factors that are consistent with expectations.	Refers to the extent to which assessment results are suited to the purpose for which they were intended, and to the clinical and social impact of using measures from an assessment for clinical and research decision making.	Refers to the extent to which assessment results are comparable across individuals, time, raters, or settings.	Refers to the extent to which the assessment if feasible, practical, and usable in the settings and for the purposes it is intended to be used
Measurement question addressed	Is the model underlying the construct sound? Is the underlying construct sufficiently represented by item content for the intended purpose of the assessment? Do stakeholders believe the items cover the concept of interest?	What is the internal structure of the items, the rating scale steps, and the persons?	Does the assessment align with similar (and different) assessments in ways that are predicted by theory and/or clinical experience? Do person measures from the assessment predict meaningful future clinically-relevant rehabilitation outcomes?	What is the intended purpose of the assessment and what kind of decisions are the measures intended to inform?	Does the assessment produce measures that are reproducible and consistent?	Is the assessment feasible, practical, and usable in the settings in which it is intended to be used? Is the assessment fit for the purposes for which it is intended?
Evidence	Theoretical or logical or empirical rationale for the trait As a result of Rasch Analysis, items and rating scale steps are ordered from "lowest" to "highest" in a way that aligns with the theoretical rationale or clinical understanding. Empirical evidence that stakeholders believe the items are relevant and representative of the domain, including expert opinion, Delphi testing, or cognitive testing	Rating scale step structure: Step thresholds ordering is theoretically and clinically reasonable for interpretation and intended use. Unidimensionality: Items are sufficiently unidimensional such that any dimensionality does not disrupt person measures. Measurement accuracy: Score properties and interpretations generalize to and across individuals, groups, settings, and tasks Iterations of Analysis: Successive iterations of Rasch Analyses improves assessment psychometrics	Item calibrations from current assessment align with other assessments sharing similar underlying trait. Person measures demonstrate an association with measures (or raw scores) from assessments sharing similar (or different) underlying trait. Measures from the assessment predict meaningful future outcomes.	The measures obtained from the assessment are useful for helping clinicians, patients, and caregivers (and other stakeholders) make good treatment/clinical decisions that inform rehabilitation services, research, and policy The measures do not result in unintended consequences that negatively impact the lives of persons with disability including but not limited to reducing timely access to appropriate rehabilitation services	Reproducibility: Step thresholds and item calibrations are reproducible across samples/groups/ settings, time, raters, and forms Reliability: The degree of association among item scores and/or total raw scores obtained at different times, by different raters, or different assessment forms	The assessment can be used as intended (not shortened or altered) in the rehabilitation setting for which it was intended (e.g., clinical practice, research, policy). Training to administer assessment is available, affordable, practicable. The assessment fits in clinical workflow. The assessment is in languages that are appropriate for the intended uses.