



Establishing severity levels for patient-reported measures of functional communication, participation, and perceived cognitive function for adults with acquired cognitive and language disorders

Matthew L. Cohen^{1,2} · Stacy M. Harnish³ · Alyssa M. Lanzi¹ · Jennifer Brello³ · William D. Hula⁴ · David Victorson⁵ · Ratna Nandakumar⁶ · Pamela A. Kisala^{2,7} · David S. Tulskey^{2,7}

Accepted: 21 December 2022 / Published online: 27 December 2022
© The Author(s) 2022

Abstract

Purpose To empirically assign severity levels (e.g., *mild*, *moderate*) to four relatively new patient-reported outcome measures (PROMs) for adults with acquired cognitive/language disorders. They include the Communicative Participation Item Bank, the Aphasia Communication Outcome Measure, and Neuro-QoL’s item banks of Cognitive Function (v2.0) and Ability to Participate in Social Roles and Activities (v1.0).

Method We conducted 17 focus groups that comprised 22 adults with an acquired cognitive/language disorder from stroke, Parkinson’s disease, or traumatic brain injury; 30 care partners of an adult with an acquired cognitive/language disorder; and 42 speech-language pathologists who had experience assessing/treating individuals with those and other cognitive/language disorders. In a small, moderated focus-group format, participants completed “PROM-bookmarking” procedures: They discussed hypothetical vignettes based on PROM item responses about people with cognitive/language disorders and had to reach consensus regarding whether their symptoms/function should be categorized as *within normal limits* or *mild*, *moderate*, or *severe* challenges.

Results There was generally good agreement among the stakeholder groups about how to classify vignettes, particularly when they reflected very high or low functioning. People with aphasia described a larger range of functional communication challenges as “mild” compared to other stakeholder types. Based on a consensus across groups, we present severity levels for specific score ranges for each PROM.

Conclusion Standardized, stakeholder-informed severity levels that aid interpretation of PROM scores can help clinicians and researchers derive better clinical meaning from those scores, for example, by identifying important clinical windows of opportunity and assessing when symptoms have returned to a “normal” range.

Keywords Patient-reported outcomes · Patient-reported outcome measures · Reference values · Reference standards · Cognitive impairments · Acquired communication disorders

✉ Matthew L. Cohen
mlcohen@udel.edu

¹ Department of Communication Sciences and Disorders, University of Delaware, 100 Discovery Blvd 6th Floor, Newark, DE 19713, USA

² Center for Health Assessment Research and Translation, University of Delaware, Newark, DE 19713, USA

³ Department of Speech and Hearing Science, The Ohio State University, Columbus, OH 43210, USA

⁴ Geriatric Research, Education, and Clinical Center, VA Health Care System, and Department of Communication Sciences and Disorders, University of Pittsburgh, Pittsburgh, PA, USA

⁵ School of Medicine Department of Medical Social Science, Northwestern University, Chicago, IL 60611, USA

⁶ University of Delaware School of Education, Newark, DE 19713, USA

⁷ Department of Physical Therapy, University of Delaware, Newark, DE 19713, USA

Patient-reported outcome measures (PROMs) are standardized scales that assess health outcomes from the patient's perspective [1], including health-related quality of life, functional status, symptoms, and health behaviors [2, 3]. In clinical settings, PROMs may be used to assist with screening and referral, diagnosis and estimating prognosis, monitoring symptoms over time, goal setting, monitoring treatment progress, and facilitating shared decision-making [4–9]. The use of PROMs in clinical practice supports the missions of evidence-based practice and person-centered care by quantifying health outcomes that matter to an individual client but that may not be fully observable to clinicians [7]. For example, people with cognitive/language disorders, their care partners, and speech-language pathologists (SLPs) often identify communicative *participation* (e.g., ordering at a restaurant) as an important therapeutic target in addition to or instead of communication *skills* (e.g., word-level repetition) [10–14]. However, participation-focused therapy requires assessment of communicative participation [10], which SLPs do less often [11]. Assessment of participation is inherently complex because “participation” measures need to reflect the convergence of communication skills, the environment, and personal perspectives to document whether the client is meeting their communication demands and preferences successfully and satisfactorily (Baylor & Darling-White, 2020, p.5). This makes communicative participation especially well-suited to be assessed by PROMs, and the Communicative Participation Item Bank (CPIB) was published by Baylor et al. (2013) to meet this need.

Within the last 20 years, several PROM systems have been developed using rigorous development standards and item-response theory (IRT). These include the Patient-Reported Outcomes Measurement Information System (PROMIS) [15], Neuro-QoL measurement system, Traumatic Brain Injury (TBI)-QOL measurement system [16], and Spinal Cord Injury (SCI)-QOL [17] measurement system. These systems were primarily designed for clinical research [18], but are increasingly being used for clinical purposes as well [4, 5, 8, 9, 19–21]. However, because they were not originally designed for clinical purposes, clinicians may question their applicability and usefulness [22]. For example, for many clinical purposes, clinicians find it useful to have simple guides for interpreting how “good” or “bad” the client's score is [23] with terms such as *mild*, *moderate*, and *severe* [24]. Established score ranges can help facilitate and simplify screening, referring, and monitoring change, for example, by identifying important clinical windows of opportunity and assessing when symptoms have returned to a “normal” range [9, 23, 25]. Establishing and assigning descriptors to score ranges are typically not part of the psychometric development of new PROMs, so additional work is needed to make the measures optimally useful for clinicians.

One method for assigning severity levels to PROM score ranges is called bookmarking, a procedure originally used in the field of education whereby content experts set criteria for mastery of learned material [26]. Applied to PROMs, bookmarking engages stakeholders (e.g., patients, care partners, and clinicians) to determine how PROM scores should be interpreted [23, 24, 27–30]. This procedure involves groups of stakeholders reviewing hypothetical vignettes based on PROM scores about a person's experience of a particular health construct. Vignettes are carefully written to correspond with actual T-score levels from a PROM. The stakeholder groups discuss and order the vignettes and then achieve consensus on where bookmarks should be placed between them to denote distinct severity levels such as normal, mild, moderate, or severe [6].

In the present paper, we describe some of the main findings of a study that aimed to set clinical cut points for a set of relatively new, IRT-based PROMs that might benefit clinical practice with adults with acquired cognitive/language disorders [6, 7, 31], particularly speech-language therapy. Specifically, we report the results of bookmarking groups that set clinical cut points for two measures of participation, the CPIB [32] and the Neuro-QoL Item Bank v1.0—Ability to Participate in Social Roles and Activities (NQ-SRA) [33, 34], a measure of functional communication, the Aphasia Communication Outcome Measure (ACOM) [35], and a measure of perceived cognitive function, the Neuro-QoL Item Bank v2.0—Cognitive Function (NQ-Cog) [33, 34]. These constructs have broad relevance to clinical practice with adults with acquired cognitive/language disorders and are well captured by PROMs.

Methods

Participants

This study aimed to develop cut points that were broadly applicable to adults with acquired cognitive/language conditions, based on input from different stakeholders and representative conditions. Stakeholders included patients, care partners, and SLPs. Conditions that caused cognitive/language challenges included traumatic brain injury (TBI), stroke, and Parkinson's disease (PD)—some of the most common conditions that lead a person to seek cognitive/language treatment from an SLP. In line with previous PRO-Bookmarking studies [27, 29, 30, 36], we aimed to enroll a single group for each stakeholder and condition type (e.g., 1 TBI client group, 1 TBI care partner group, 1 stroke client group, 1 stroke care partner group, etc.). We ended up exceeding the planned enrollment for some groups but not all.

In the end, we conducted 17 focus groups—4 stroke client groups ($n = 13$), 4 stroke care partner groups ($n = 19$), 1 PD client group ($n = 6$), 1 PD care partner group ($n = 7$), 1 TBI client group ($n = 3$), and 1 TBI care partner group ($n = 4$). These groups approached the PRO-Bookmarking task based on shared lived experiences with their condition. We also enrolled 5 SLP groups ($n = 42$) that set bookmark locations that they thought were applicable to any adult client with a cognitive/language condition. Details of these participants and groups can be found in Table 1.

The presence of an acquired cognitive/language condition was determined by clients and care partners endorsing one or more of the following challenges: Expressing oneself; other people understanding what one wants to say; understanding what other people say; reading or writing because of language difficulties (i.e., not because of visual or motor difficulties); thinking of the right word to say; moving one's mouth, lips, or tongue to pronounce words; or thinking or memory. Participants with TBI or stroke were at least

6 months post-onset. Participants with these conditions were familiar to the research team by virtue of having participated in other research studies. They were invited to participate in this study because their cognitive/language challenges were thought to be representative of their condition but not too severe to preclude participation in the PRO-bookmarking tasks. For descriptive purposes, the severity of language impairment for people with aphasia was assessed with the Quick Aphasia Battery [37] or Comprehensive Aphasia Test [38] and was characterized by these performance-based measures as being mild or moderate (for more details, see [39]). People with PD or TBI produced Montreal Cognitive Assessment (MoCA) [40] total scores between 19–27 (median = 23.5), which are in the range of scores produced by people with mild neurocognitive disorder, overlapping with the upper end of scores produced by people with major neurocognitive disorder [40, 41].

Care partners were spouses of a person with a cognitive/language condition who had at least weekly contact with

Table 1 Participants and groups

ID	Participants	Age: range (median)	Sex	Race	Ethnicity	Experience: range (median)	PROMs book-marked
						Years as SLP	
Clinicians							
A	7 SLPs	30–62 (40)	7 w, 0 m	7 White	7 Non-Hispanic	6–34 (10)	1,2,3,4
B	10 SLPs	25–54, 3 did not report (33)	9 w, 1 m	9 White, 1 Black	10 Non-Hispanic	1–20 (6.5)	1,2,3,4
C	8 SLPs	27–40 (33.5)	8 w, 0 m	6 White, 1 Asian, 1 Other	8 Non-Hispanic	3.5–14 (9.5)	1,2,3,4
D	6 SLPs	26–65 (55.5)	6 w, 0 m;	6 White	6 Non-Hispanic	3–42 (22.5)	1,3,4
E	11 SLPs	27–67 (48)	11 w, 0 m	10 White, 1 Black	11 Non-Hispanic	3–30 (25)	1,2,4
						Years as care partner	
Care partners							
F	4 partners of PWA	60–76 (69)	3 w, 1 m	3 White, 1 Asian	1 Hispanic, 3 non-Hispanic	1–7.5 (1.5)	1,2,3,4
G	6 partners of PWA	35–69 (61.5)	6 w, 0 m	4 White, 2 Black	6 Non-Hispanic	2–10 (5.25)	1,2,3,4
H	4 partners of PWA	41–68, one did not report (62)	4 w, 0 m	4 White	4 Non-Hispanic	1–5.5 (3.25)	1,2,3,4
I	5 partners of PWA	48–75 (62)	4 w, 1 m	5 White	5 Non-Hispanic	1–18 (3)	1,2,3,4
J	7 partners of PwPD	55–76 (67)	6 w, 1 m	7 White	7 Non-Hispanic	2–15 (7)	1,3
K	4 partners of PwTBI	57–77 (64.5)	3 w, 1 m	4 White	4 Non-Hispanic	2–46 (14)	1,2,3
						Years with condition	
People with communication disorders							
L	6 PwPD	58–76 (68)	1 w, 5 m	6 White	6 Non-Hispanic	2–15 (8.5)	1,3
M	3 PWA	49–76 (62)	0 w, 3 m	2 White, 1 Black	3 Non-Hispanic	1.5–4.5 (3.5)	1
N	2 PWA	49–76 (62.5)	0 w, 2 m	2 White	2 Non-Hispanic	1.5–3.5 (2.5)	2,3,4
O	3 PWA	38–69 (53)	0 w, 3 m	2 White, 1 Black	3 Non-Hispanic	1.5–7 (4)	1,2,3,4
P	5 PWA	38–73 (55)	1 w, 4 m	4 White, 1 Black	5 Non-Hispanic	1–6 (4)	1,2,3,4
Q	3 PwTBI	30–77 (46)	0 w, 3 m	3 White	3 Non-Hispanic	3–23 (5)	1,2,3

SLP = speech-language pathologist, PWA = person with aphasia; PwPD = person with Parkinson's disease; PwTBI = person with TBI. PROMs Bookmarked: 1 = Communicative Participation Item Bank, 2 = Neuro-QoL Cognition, 3 = Neuro-QoL Ability to Participate in Social Roles and Activities, 4 = Aphasia Communication Outcome Measure. Group IDs here do not match those presented in Cohen, Harnish et al. (2021). All demographic variables were self-identified by participants

the person both before and after the onset of the condition. The inclusion criteria permitted other close relations, but the care partners in our sample were all spouses. SLPs were certified by the American Speech-Language-Hearing Association and had experience treating at least 50 adults with acquired cognitive/language conditions. All participants were at least 18 years old and could read and understand spoken English. All study procedures were approved by the Institutional Review Boards at the University of Delaware and The Ohio State University. Participants provided written informed consent, and participants with cognitive/language disorders were required to pass a series of yes/no comprehension questions to ensure their cognitive/linguistic capacity to provide informed consent. Participants were compensated for their time.

Measures

Communicative Participation Item Bank (CPIB) [32]

The CPIB is a 46-item PROM that assesses communicative participation, defined as “taking part in life situations in which knowledge, information, ideas, or feelings are exchanged” [32, 42]. A 10-item short form is also available. It was initially calibrated on a sample of adults with multiple sclerosis, Parkinson’s disease, amyotrophic lateral sclerosis, and head and neck cancer [32], and was later found to have evidence for validity with people with aphasia as well [43]. Examples of items on the CPIB include “Does your condition interfere with getting your turn in a fast-moving conversation?” and “Does your condition interfere with asking questions in a conversation?”.

Aphasia Communication Outcome Measure (ACOM) [35]

The ACOM consists of 59 items that assess post-stroke functional communication, defined as “the ability to engage in common, everyday behaviors, tasks, activities, and life situations that involve understanding and/or producing spoken, written, and/or non-verbal messages, signs, and symbols” (Doyle et al., 2008, p720 as cited in Hula et al., 2015). ACOM items assess spoken language expression and comprehension, reading, writing, and number use. Examples of ACOM items include “How effectively can you find the words you want to say during communication?” and “How effectively can you say the name of common objects (e.g., bed, lamp, pencil)?”.

Neuro-QoL Cognitive Function Item Bank (v2.0) (NQ-Cog) [33, 34]

The NQ-Cog consists of 28 items that assess “perceived difficulties in cognitive abilities (e.g., memory, attention,

and decision-making) or in the application of such abilities to everyday tasks (e.g., planning, organizing, calculating, remembering, and learning)” [44]. Short forms are also available. Examples of NQ-Cog items include “How much difficulty do you currently have checking the accuracy of financial documents, (e.g., bills, checkbook, or bank statements)?” and “In the past 7 days, I had trouble keeping track of what I was doing if I was interrupted.”

Neuro-QoL Ability to Participate in Social Roles and Activities (v1.0) (NQ-SRA) [33, 34]

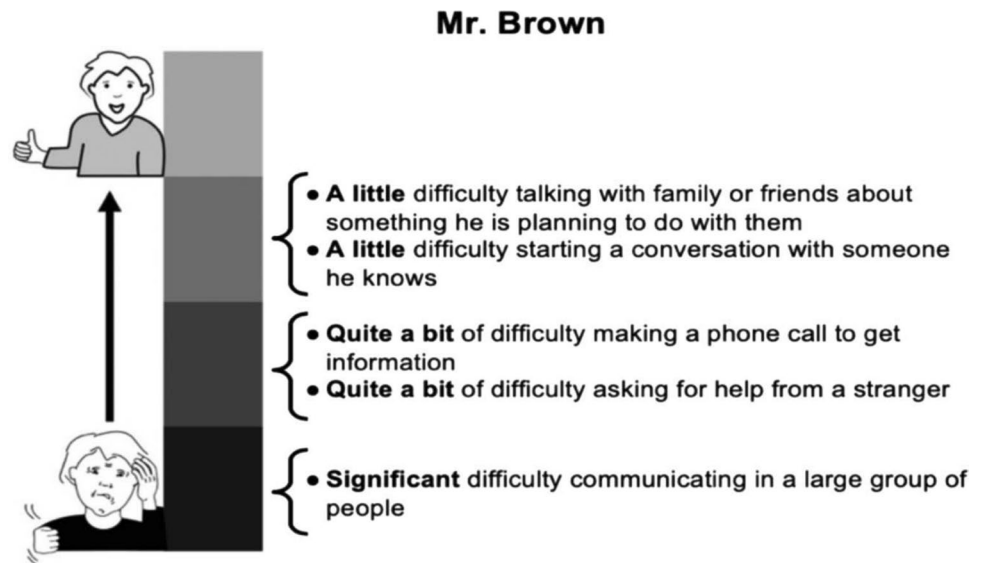
The NQ-SRA consists of 45 items that assess the “degree of involvement in one’s usual social roles, activities, and responsibilities; including work, family, friends, and leisure” [44]. Short forms are also available. As with the CPIB, it assesses aspects of participation but is not specific to communicative participation. Examples of NQ-SRA items include “In the past 7 days I have been able to do all of my regular activities with friends” and “In the past 7 days I have been able to keep up with my family responsibilities.”

Bookmarking materials and procedures

Bookmarking is a process by which stakeholders (clients, care partners, and clinicians) help determine how PROM score ranges should be clinically interpreted [23, 24, 27–29, 45]. A previous publication describes how we have made the bookmarking materials and procedures more accessible to individuals with cognitive/language conditions [39]. These materials and procedures are described briefly below and in more detail in the Supplementary Material.

The first step was to create vignettes about hypothetical clients with acquired cognitive/language conditions based on PROM items and responses. For example, two CPIB items and item responses could be “José’s condition interferes quite a bit talking with people that he does not know, but only a little bit talking with people that he does know.” As also described in our previous publication [39] the selection of items followed Victorson’s guidelines [46]. Consistent with previous PRO-Bookmarking studies [28, 29], we developed about ten vignettes per item bank and each vignette contained five items/responses. Vignettes were written at 0.5 SD intervals for the CPIB, ACOM, and NQ-Cog, and every 0.25 SD intervals for the NQ-SRA. The reason that the vignettes for the NQ-SRA were closer together is that there was less variability in most-likely responses in items at T-score levels < 35 and > 50, so we distributed the ten vignettes across a smaller range of scores to permit more precise classification. Each vignette was assigned an arbitrary surname (e.g., “Mr. Garcia”), and surnames represented the most common ancestries in the USA.

Fig. 1 Example of a vignette card from Cohen et al., (2021). This card depicts “Mr. Brown” who represents a CPIB T score of 40. As further described by Cohen et al. (2021), participants had 8–10 cards per PROM and discussed where bookmarks should be placed between adjacent cards that were in order and represented different scores. Reprinted from Cohen et al. (2021) with permission from the publisher



The second step was to put those vignettes on physical cards (8.5" × 5.5" cardstock) that could be viewed, sorted, and manipulated by participants. To be maximally accessible to individuals with communication disorders, the vignettes were displayed in large font with bullet points and with a graphic to indicate whether the items/responses described a high or low level of the trait being assessed [39]. Figure 1 shows an example of a CPIB vignette card. Representative examples of vignettes from every item bank are freely available as supplementary material to our previous publication [39].

The third step was to conduct bookmarking groups separately for each stakeholder and condition type: Adults with post-stroke aphasia, TBI, or PD (6 groups); care partners of people with aphasia, TBI, or PD (6 groups); and SLPs (5 groups) (Table 1). We adapted the bookmarking procedure to be accessible to people with cognitive/language conditions by breaking down tasks into small components, displaying activities graphically when possible, and using moderator(s) who were able to give impromptu communication support as needed because of their clinical training as a psychologist (authors M.L. and D.V.) or speech-language pathologist (authors A.L., J.B., and S.H.) [39]. After providing informed consent and completing introductions and practice examples, participants were introduced to a PROM and what the response options look like to respondents. They were then given the vignette cards associated with that PROM and asked to spend some time becoming familiar with them. Participants would then place “bookmarks” between adjacent vignettes that they felt represented a boundary between people whose challenges they would classify as *within normal limits*, or *mild*, *moderate*, or *severe* challenges. Group members sometimes disagreed in their initial placement of bookmarks, and it was the job of the

moderator(s) to facilitate a discussion that led to a consensus regarding why one vignette person’s challenges were *mild* but the adjacent person’s was *moderate*. Figure 2 depicts a visual communication aid used during this discussion. In the end, each group achieved consensus about the location of bookmarks for the item banks they were able to complete. As also discussed by Cohen, Harnish et al., [39] not every group bookmarked every item bank due to time, appropriateness to the condition, and other constraints. For example, TBI and PD groups did not bookmark the ACOM because it is specific to post-stroke aphasia. The PROMs were not bookmarked in a fully counterbalanced order. Because it was initially unclear how long the task would take with each stakeholder group, item banks were bookmarked in order of their importance to the aims of the grant that funded this investigation: first the CPIB, then the ACOM (for aphasia stakeholders only), then NQ-Cog or NQ-SRA followed by the other. Table 1 shows which groups bookmarked which item bank.

Placement of consensus cut points

The authors of this paper served as an adjudicating expert panel to interpret and synthesize the results of multiple bookmarking groups into a final set of cut points. When the groups largely agreed on the classification of adjacent vignettes, for example, T45 indicates *mild* challenges and T40 indicates moderate challenges, the consensus bookmark location was placed between them at T42.5. However, when the groups were split about how to classify a vignette, e.g., whether the vignette at T35 should be called *moderate* or *severe*, the consensus bookmark location was placed at that level such that any score above 35 is *moderate* and any score below 35 is *severe*.

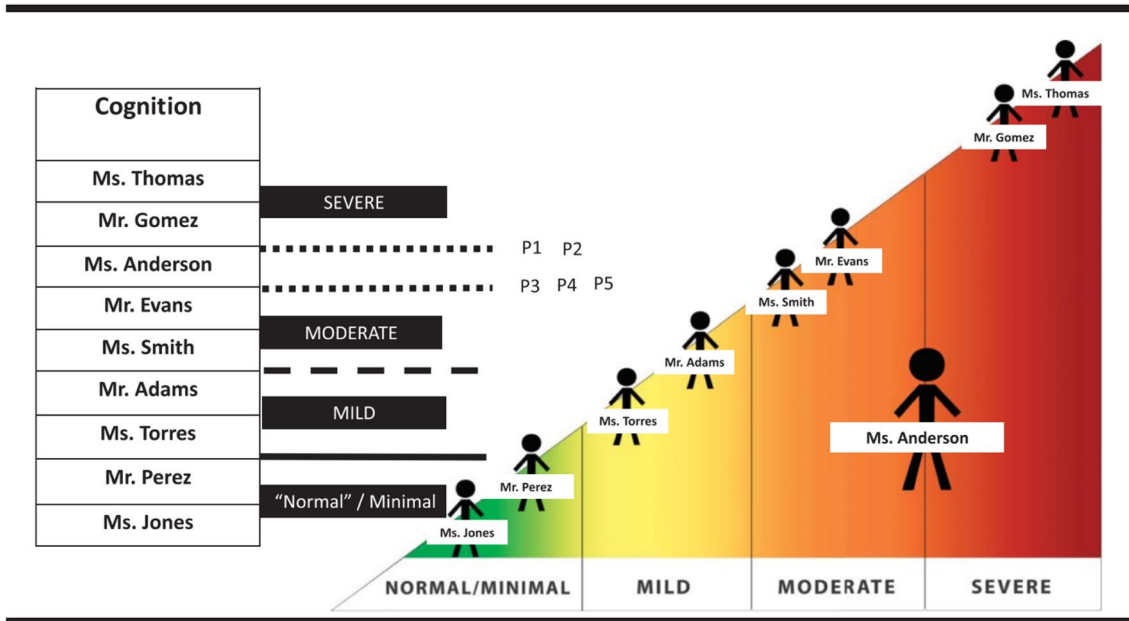


Fig. 2 Powerpoint display that was developed from feedback from stakeholders and used to graphically support the PRO-bookmarking task. As also described in Cohen et al. (2021), Fig. 2 depicts a scenario in which participants have agreed on all bookmark locations except the boundary between moderate and severe. It is being discussed whether “Ms. Anderson” should be classified as moderate or severe. Graphically, Ms. Anderson is the largest figure, indicating that she is the vignette under discussion. Previously classified vignettes are smaller, in psychometric order, and within the category bound-

ary to which participants have assigned them previously in the discussion. Participants 1 and 2 think Ms. Anderson should be classified as moderate (that is, the boundary for severe is between her and Mr. Gomez); Participants 3, 4, and 5 think she should be classified as severe (that, the boundary for severe is between her and Mr. Evans). As participants are persuaded and change their position, the moderator uses multimodal support to make the consensus-building process. Reprinted from Cohen et al., 2021 with permission from the publisher

Group ID	Description	T25	T30	T35	T40	T45	T50	T55	T60	T65
A	SLPs	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
B	SLPs	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
C	SLPs	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
D	SLPs	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
E	SLPs	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
F	Partners of PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
G	Partners of PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
H	Partners of PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
I	Partners of PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
J	Partners of PwPD	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
K	Partners of PwTBI	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
L	PwPD	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
M	PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
O	PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
P	PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
Q	PwTBI	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Blue	Blue
CPIB CONSENSUS CUTPOINTS		Severe			Moderate		Mild		Normal	
		T35			T45		T57.5			

Fig. 3 Cut points for the Communicative Participation Item Bank. CPIB = Communicative Participation Item Bank. Vignettes are shown by their associated T score ($M=50$, $SD=10$), where a T score equal to 50 matches the mean of that measure’s reference sample. For the CPIB, that reference sample is 701 individuals with multiple sclerosis, Parkinson’s disease, amyotrophic lateral sclerosis, and head and

neck cancer. Each colored cell represents how each group classified each vignette (e.g., the T25 vignette, the T30 vignette)—within normal limits (blue), mild (yellow), moderate (orange), or severe (red). The synthesized cut points, determined by the adjudicating expert panel, are at the bottom

Group ID	Description	T25	T30	T35	T40	T45	T50	T55	T60	T65	T70	T75
A	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
B	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
C	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
D	SLPs	Red	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
E	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
F	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
G	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
H	Partners of PWA	Red	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
I	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
M	PWA	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
O	PWA	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
P	PWA	Red	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
ACOM CONSENSUS CUTPOINTS		Severe			Moderate			Mild-Mod		Mild		
		T37.5			T52.5			T62.5				

Fig. 4 Cut points for the Aphasia Communication Outcome Measure. Vignettes are shown by their associated T score (M=50, SD=10), where a T score equal to 50 matches the mean of that measure’s reference sample. For the ACOM, that reference sample is 329 PWA. Because every participant in the reference sample had aphasia, there

is no “normal” range of functional communication. Each colored cell represents how each group classified each vignette (e.g., the T25 vignette, the T30 vignette)—within normal limits (blue), mild (yellow), moderate (orange), or severe (red). The synthesized cut points, determined by the adjudicating expert panel, are at the bottom

Group ID	Description	T20	T25	T30	T35	T40	T45	T50	T55	T60
A	SLPs	Red	Red	Yellow	Yellow	Yellow	Yellow	Blue	Blue	Blue
B	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
C	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
E	SLPs	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
F	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
G	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
H	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
I	Partners of PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
K	Partners of PwTBI	Red	Red	Yellow	Yellow	Yellow	Yellow	Blue	Blue	Blue
N	PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
O	PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
P	PWA	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
Q	PwTBI	Red	Red	Red	Yellow	Yellow	Yellow	Blue	Blue	Blue
NQ-COG Consensus Cut Points		Severe			Moderate		Mild	Normal		
		T30			T40		T47.5			

Fig. 5 Cut points for the Neuro-QoL Item Bank v2.0—Cognitive Function. NQ-Cog=Neuro-QoL Item Bank v2.0—Cognitive Function. Vignettes are shown by their associated T score (M=50, SD=10), where a T score equal to 50 matches the mean of that measure’s reference sample. For the NQ-Cog, that reference sample

matches a U.S. census-matched general population [53]. Each colored cell represents how each group classified each vignette (e.g., the T25 vignette, the T30 vignette)—within normal limits (blue), mild (yellow), moderate (orange), or severe (red). The synthesized cut points, determined by the adjudicating expert panel, are at the bottom

Results

Cut points

Figures 3, 4, 5, and 6 visually depict how each PRO-bookmarking group classified vignettes and the consensus cut points. The group IDs in Figs. 3, 4, 5, and 6 match the group IDs in Table 1 so the reader can see details about group characteristics. For the CPIB (Fig. 3), scores above T57.5 were found to represent “minimal problems” or a “normal” experience of communicative participation. Scores between

T57.5 and T45 represent *mild* challenges, scores between T45 and T35 represent “moderate” challenges, and scores below T35 represent “severe” challenges.

For the ACOM (Fig. 4), the consensus cut points were set as follows. Scores above T62.5 represent “mild” aphasia, scores between T62.5 and T52.5 represent “mild-to-moderate” aphasia, scores between T52.5 and T37.5 represent “moderate” aphasia, and scores below T37.5 represent “severe” aphasia. Reaching consensus on these cut points required more consideration than did the other PROMs. PWA extended the “mild” range of challenges further than other stakeholder groups. Because of this, the most contested

Group ID	Description	T30	T32.5	T35	T37.5	T40	T42.5	T45	T47.5	T50	T52.5
A	SLPs	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
B	SLPs	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
C	SLPs	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
D	SLPs	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
F	Partners of PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
G	Partners of PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
H	Partners of PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
I	Partners of PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
J	Partners of PwPD	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
K	Partners of PwTBI	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
L	PwPD	Red	Red	Orange	Orange	Orange	Orange	Orange	Blue	Blue	Blue
N	PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
O	PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
P	PWA	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
Q	PwTBI	Red	Red	Red	Orange	Orange	Orange	Orange	Blue	Blue	Blue
NQ-SRA Consensus Cut Points		Severe			Moderate		Mild		Normal		
					T37.5	T42.5	T47.5				

Fig. 6 Cut points for the Neuro-QoL Item Bank v1.0—Ability to Participate in Social Roles and Activities. NQ-SRA=Neuro-QoL Item Bank v1.0—Ability to Participate in Social Roles and Activities. Vignettes are shown by their associated T score ($M=50$, $SD=10$), where a T score equal to 50 matches the mean of that measure’s reference sample. For the NQ-SRA, that reference sample matches a U.S.

census-matched general population [53]. Each colored cell represents how each group classified each vignette (e.g., the T25 vignette, the T30 vignette)—within normal limits (blue), mild (yellow), moderate (orange), or severe (red). The synthesized cut points, determined by the adjudicating expert panel, are at the bottom

range (T-scores 62.5—52.5) was called “mild-to-moderate.” Because every participant in the reference sample had aphasia, there is no “normal” range of functional communication on the ACOM.

For the NQ-Cog, scores above T47.5 represent “minimal problems” or a “normal” experience of cognitive errors. Scores between T47.5 and T40 represent “mild” challenges, scores between T40 and T30 represent “moderate” challenges, and scores below T30 represent “severe” challenges. When Group H (partners of PWA) bookmarked the NQ-Cog they chose not to classify any vignette as “mild” (Fig. 5). This was the only example of a group deciding that a descriptor did not apply to any vignette. When we as the study team served as an adjudicating expert panel to set consensus bookmark locations, we decided that because all other groups perceived a score range as “mild,” the category was justified in the consensus decision.

For the NQ-SRA, scores above T47.5 represent “minimal problems” or a “normal” experience of participation in social roles and activities. Scores between T47.5 and T42.5 represent “mild” challenges,” scores between T42.5 and T37.5 represent “moderate” challenges, and scores below T37.5 represent “severe” challenges (Fig. 6). For all item banks, scores that fall exactly on the cut point could be described as “moderate-severe,” “mild-moderate,” or “borderline.”

Discussion

PROMs are making increasingly important contributions to evidence-based practice and person-centered care by measuring important outcomes that are otherwise difficult to quantify [7]. Several relatively new PROMs have been developed that are intended for or highly relevant for adults with acquired cognitive/language disorders, and both clients and clinicians have much to gain from PRO-informed clinical practice [7, 31, 47]. Previous studies have indicated that clinicians find it helpful to have descriptors such as mild, moderate, and severe to describe PROM score ranges [24], so the purpose of this study was to map those descriptors onto score ranges on these PROMs. In a previous publication [39], we reported how we adapted the bookmarking procedure to be accessible to individuals with acquired cognitive/ language impairments. Here, we report the results from bookmarking groups that set cut points for the CPIB, ACOM, NQ-Cog, and NQ-SRA.

The healthmeasures.net website, which hosts PROMIS, Neuro-QoL, and other measurement systems, has rule-of-thumb guidance for interpretation of the NQ-SRA and NQ-Cog based on the score distributions from calibration testing data [15, 48]. The guideline is that for measures that reference a general population sample, T-scores above T45 are considered “within normal limits,” scores between T45–40 are mild, scores between T40 and T30 are moderate, and scores below T30 are severe. To our knowledge, this is the first report of PRO-bookmarking-derived clinical cut scores

for the NQ-SRA and the second study to report bookmarking-derived scores for the NQ-Cog.

The first study that reported cut points for the NQ-Cog, Rothrock et al. [24], conducted a single bookmarking group with people with cancer and a single bookmarking group with oncologists. Overall, the NQ-Cog cut points set by our stakeholder groups are grossly congruent with (i.e., $\leq 0.5SD$ different from) the stakeholder groups reported by Rothrock et al. and with the guidelines of healthmeasures.net. Our groups' cut points for the NQ-SRA are grossly congruent with the healthmeasures.net guidelines for the mild and moderate thresholds, but different for the threshold of severe; whereas healthmeasures.net indicates that the severe range is for scores $< T30$, our consensus cut points for that measure is for scores $< T37.5$. This means that our group assigned the descriptor “severe” to a range of scores (T-scores from 30.0 up to 37.5) that healthmeasures.net would call moderate. The reason for this discrepancy is not clear, except that our cut point was based on bookmarking (i.e., the impression of stakeholders) rather than based on score distributions. Ultimately, however, clinically meaningful interpretation of PROM scores is most valid when interpretation guides are specific to the population completing the PROM [49]; what is considered a “mild” cognitive limitation for a person with cancer might be different than for a person with an acquired cognitive/language disorder.

This is the first study to report descriptors for CPIB score ranges. It is notable that the “mild” range begins at $T57.5$, a much higher T-score threshold than the NQ-Cog or NQ-SRA item banks. This was most-likely related to the composition of the reference sample. Whereas the reference samples for the NQ-Cog and NQ-SRA mirror the general population [15, 48], where $T50$ indicates a truly typical experience [50], the reference sample for the CPIB was a sample of people with communication disorders. This means that $T = 50$ indicates a typical experience of communicative participation for people with a communication disorder; thus, it is not surprising that our stakeholders determined that the “normal” range is for scores notably higher than $T50$. This highlights the importance of PRO-bookmarking. Without clear descriptions of score ranges, it could be difficult for a clinician to know how to interpret a CPIB score of, say, $T55$. That score is $0.5 SD$ above the mean, indicating better-than-average participation, but would still be classified by most of our stakeholder groups as indicating mild restrictions to communicative participation.

Previous bookmarking studies have reported mixed consistency between stakeholder groups, perhaps related to the types of stakeholders, the constructs being assessed, and the number of groups that were conducted. For the CPIB, NQ-Cog, and NQ-SRA, there was good consensus among our groups and so the consensus cut points were relatively faithful to the individual group cut points without needing

to reconcile major differences. There was less consistency between PWA groups and other stakeholders who bookmarked the ACOM, and the discrepancies were in the same direction as previous bookmarking reports; patient groups set higher thresholds (i.e., “tolerate” more symptoms) than clinician or care partner groups [24, 27, 29]. For example, Cook et al. [27] reported that people with multiple sclerosis agreed with clinicians on the cut points for PROMs of mobility and sleep disturbance, but set higher thresholds for PROMs that assessed fine motor function and fatigue. One possible explanation for this finding is that people with the condition acclimate to its symptoms (i.e., recalibrate what constitutes a “mild problem”) [27, 29]. Curiously, however, PWA (and other patient groups) did not place bookmarks in that way for any other PROM/construct. In fact, on the NQ-SRA they had a *lower* threshold (i.e., tolerated fewer symptoms) of restricted participation in social roles and activities than did care partners and half of the SLP groups, labeling $T47.5$ as “mild” whereas most others called it “normal.”

Despite these occasional discrepancies, there is generally good agreement among groups. One strength of this study is the number of groups that were conducted. Whereas previous PRO-bookmarking studies generally conducted a single group for each condition or stakeholder type [27, 29, 30, 36], we conducted multiple groups. This helps improve confidence in the consensus cut points. Still, it is difficult to know when saturation is reached—the point at which the addition of new data fails to change the consensus bookmark locations. The consensus cut points based on 12–16 focus groups are likely more stable than the cut points of individual groups, so interpretation of subgroup (e.g., PWA-only) cut points requires caution. Our SLP and aphasia stakeholder groups had the largest sample sizes, so those perspectives are likely the most stable (i.e., likely to be saturated). Although the cut points placed by other stakeholder types largely converged with those set by SLPs and aphasia stakeholders, it is possible that the addition of more data would reveal differences. It would be useful for future research to test these bookmark locations with more perspectives from TBI and PD stakeholders, as well as the perspectives of patients with other neurogenic communication disorders. We also cannot rule out the possibility that cut points were influenced by the items included in the vignettes, the order in which the PROMs were bookmarked, or the members that comprised the groups.

The results presented here are intended to help make these four PROMs more interpretable for clinicians and researchers who serve adults with cognitive/language disorders, improving their ability to quantify aspects of quality of life, especially those aspects that are difficult to quantify otherwise [7]. However, we would caution against over-reliance on these cut points alone for major clinical decision-making such as whether a disorder is present or not. It is important to

keep in mind that these cut points are not meant to diagnose a clinical condition, they do not necessarily apply to other causes of disorders that are not represented here, and our vignettes were limited to 0.25 or 0.5 SD intervals, limiting the precision of the cut points. Instead, the score ranges presented here can help clinicians determine roughly “how good or bad” a particular score is based on the average client with that level of symptoms/function. However, the descriptor may not apply to a particular client for whom participation, for example, is more of a priority than for the average client. Other important interpretation tools for PROMs include the minimal detectable change and minimally important difference values, which help determine “how much better or worse” a particular score is compared to a previous score [39, 49, 51, 52]. PROMs can serve as very useful data to include in a clinician’s delivery of evidence-based practice, but they are most valid and useful when interpreted alongside other assessment data (e.g., performance-based and clinician-rated assessments) [7].

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-022-03337-2>.

Acknowledgements The authors would like to thank Hope Morrison, Teresa Highberger, Mariah Graham, Autumn Brown, Kiera Gran, and Haley Robinson for their hard work and contributions to this study.

Funding This study was funded by the Gordon and Betty Moore Foundation (GBMF #5299 to MLC).

Data availability The datasets generated and/or analyzed during the current study are not publicly available because additional publications are planned but are available from the corresponding author upon reasonable request.

Declarations

Competing interests Financial interests: No author has financial interests to disclose. Non-financial interests: Dr. Hula is the lead developer of the Aphasia Communication Outcome Measure that is a topic of investigation in the paper.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the University of Delaware Institutional Review Board.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Patrick, D. L., Burke, L. B., Powers, J. H., Scott, J. A., Rock, E. P., Dawisha, S., O’Neill, R., & Kennedy, D. L. (2007). Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *10*(Suppl 2), S125-137. <https://doi.org/10.1111/j.1524-4733.2007.00275.x>
- Bartlett, S. J., & Ahmed, S. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 1: Introduction. *Journal of Clinical Epidemiology*, *89*, 114–118. <https://doi.org/10.1016/j.jclinepi.2017.04.012>
- Mayo, N. E., Figueiredo, S., Ahmed, S., & Bartlett, S. J. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 2: Terminology proposed to measure what matters in health. *Journal of Clinical Epidemiology*, *89*, 119–124. <https://doi.org/10.1016/j.jclinepi.2017.04.013>
- Bingham, C. O., Bartlett, S. J., Merkel, P. A., Mielenz, T. J., Pilkonis, P. A., Edmundson, L., & MooreSabharwal, L. R. K. (2016). Using patient-reported outcomes and PROMIS in research and clinical applications: Experiences from the PCORI pilot projects. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *25*(8), 2109–2116. <https://doi.org/10.1007/s11136-016-1246-1>
- Bingham, C. O., Noonan, V. K., Auger, C., Feldman, D. E., Ahmed, S., & Bartlett, S. J. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 4: Patient-reported outcomes can inform clinical decision making in chronic care. *Journal of Clinical Epidemiology*, *89*, 136–141. <https://doi.org/10.1016/j.jclinepi.2017.04.014>
- Cohen, M. L., Lanzi, A., & Boulton, A. J. (2021). Clinical use of PROMIS, Neuro-QoL, TBI-QoL, and other patient-reported outcome measures for individual clients with cognitive & language disorders. *Seminars in Speech and Language*, *42*(3), 192–210.
- Cohen, M. L., & Hula, W. D. (2020). Patient-reported outcomes and evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology*, *29*(1), 357–370. https://doi.org/10.1044/2019_AJSLP-19-00076
- Noonan, V. K., Lyddiatt, A., Ware, P., Jaglal, S. B., Riopelle, R. J., Bingham, C. O., Figueiredo, S., Sawatzky, R., Santana, M., Bartlett, S. J., & Ahmed, S. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 3: Patient-reported outcomes can facilitate shared decision-making and guide self-management. *Journal of Clinical Epidemiology*, *89*, 125–135. <https://doi.org/10.1016/j.jclinepi.2017.04.017>
- Wagner, L. I., Schink, J., Bass, M., Patel, S., Diaz, M. V., Rothrock, N., PearmanGershon, T. R., & PenedoRosenCella, F. J. S. D. (2015). Bringing PROMIS to practice: Brief and precise symptom screening in ambulatory cancer care. *Cancer*, *121*(6), 927–934. <https://doi.org/10.1002/ncr.29104>
- Baylor, C., & Darling-White, M. (2020). Achieving participation-focused intervention through shared decision making: Proposal of an age-and disorder-generic framework. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/2020_AJSLP-19-00043
- Torrence, J. M., Baylor, C. R., Yorkston, K. M., & Spencer, K. A. (2016). Addressing communicative participation in treatment planning for adults: A survey of U.S. speech-language

- pathologists. *American Journal of Speech-Language Pathology*, 25(3), 355–370. https://doi.org/10.1044/2015_AJSLP-15-0049
12. Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2016). Core outcomes in aphasia treatment research: An e-Delphi consensus study of international aphasia researchers. *American Journal of Speech-Language Pathology*, 25(4S), S729–S742. https://doi.org/10.1044/2016_AJSLP-15-0150
 13. Wallace, S. J., Worrall, L., Rose, T., & Dorze, G. L. (2017). Which treatment outcomes are most important to aphasia clinicians and managers? *An international e-Delphi consensus study. Aphasiology*, 31(6), 643–673. <https://doi.org/10.1080/02687038.2016.1186265>
 14. Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., Kong, A. P. H., Simmons-Mackie, N., Scarinci, N., & Gauvreau, C. A. (2017). Which outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation*, 39(14), 1364–1379. <https://doi.org/10.1080/09638288.2016.1194899>
 15. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., PROMIS Cooperative Group. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, 63(11), 1179–1194. <https://doi.org/10.1016/j.jclinepi.2010.04.011>
 16. Tulskey, D. S., Kisala, P. A., Victorson, D., Carlozzi, N., Bushnik, T., Sherer, M., Choi, S. W., Heinemann, A. W., Chiaravalloti, N., Sander, A. M., Englander, J., & Cella, D. (2016). TBI-QOL: Development and Calibration of Item Banks to Measure Patient Reported Outcomes Following Traumatic Brain Injury. *The Journal of Head Trauma Rehabilitation*, 31(1), 40–51. <https://doi.org/10.1097/HTR.0000000000000131>
 17. Tulskey, D. S., Kisala, P. A., Victorson, D., Tate, D. G., Heinemann, A. W., Charlifue, S., Kirshblum, S. C., Fyffe, D., Gershon, R., Spungen, A. M., & Cella, D. (2015). Overview of the spinal cord injury-quality of life (SCI-QOL) measurement system. *The Journal of Spinal Cord Medicine*, 38(3), 257–269. <https://doi.org/10.1179/2045772315Y.00000000023>
 18. National Institutes of Health. (2014). *A Decade of Discovery: The NIH Roadmap and Common Fund*. Retrieved from <https://commodfund.nih.gov/sites/default/files/ADecadeofDiscoveryNIHRoadmapCF.pdf>
 19. Ahmed, S., Ware, P., Gardner, W., Witter, J., Bingham, C. O., Kairy, D., & Bartlett, S. J. (2017). Montreal accord on patient-reported outcomes (PROs) use series—Paper 8: Patient-reported outcomes in electronic health records can inform clinical and policy decisions. *Journal of Clinical Epidemiology*, 89, 160–167. <https://doi.org/10.1016/j.jclinepi.2017.04.011>
 20. Chan, E. K. H., Edwards, T. C., Haywood, K., Mikles, S. P., & Newton, L. (2019). Implementing patient-reported outcome measures in clinical practice: A companion guide to the ISOQOL user's guide. *Quality of Life Research*, 28(3), 621–627. <https://doi.org/10.1007/s11136-018-2048-4>
 21. Hsiao, C.-J., Dymek, C., Kim, B., & Russell, B. (2019). Advancing the use of patient-reported outcomes in practice: Understanding challenges, opportunities, and the potential of health information technology. *Quality of Life Research*. <https://doi.org/10.1007/s11136-019-02112-0>
 22. Rose, M., & Bezjak, A. (2009). Logistics of collecting patient-reported outcomes (PROs) in clinical practice: An overview and practical examples. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 18(1), 125–136. <https://doi.org/10.1007/s11136-008-9436-0>
 23. Cella, D., Choi, S., Garcia, S., Cook, K. F., Rosenbloom, S., Lai, J.-S., Tatum, D. S., & Gershon, R. (2014). Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 23(10), 2651–2661. <https://doi.org/10.1007/s11136-014-0732-6>
 24. Rothrock, N. E., Cook, K. F., O'Connor, M., Cella, D., Smith, A. W., & Yount, S. E. (2019). Establishing clinically-relevant terms and severity thresholds for Patient-Reported Outcomes Measurement Information System® (PROMIS®) measures of physical function, cognitive function, and sleep disturbance in people with cancer using standard setting. *Quality of Life Research*. <https://doi.org/10.1007/s11136-019-02261-2>
 25. Basch, E., Deal, A. M., Kris, M. G., Scher, H. I., Hudis, C. A., Sabbatini, P., & Schrag, D. (2016). Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 34(6), 557–565. <https://doi.org/10.1200/JCO.2015.63.0830>
 26. Karantonis, A., & Sireci, S. G. (2006). The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, 25(1), 4–12. <https://doi.org/10.1111/j.1745-3992.2006.00047.x>
 27. Cook, K. F., Victorson, D. E., Cella, D., Schalet, B. D., & Miller, D. (2015). Creating meaningful cut-scores for Neuro-QOL measures of fatigue, physical functioning, and sleep disturbance using standard setting with patients and providers. *Quality of Life Research*, 24(3), 575–589. <https://doi.org/10.1007/s11136-014-0790-9>
 28. Cook, K. F., Cella, D., & Reeve, B. B. (2019). PRO-bookmarking to estimate clinical thresholds for patient-reported symptoms and function. *Medical Care*, 57, S13–S17. <https://doi.org/10.1097/MLR.0000000000001087>
 29. Kisala, P. A., Victorson, D., Nandakumar, R., Shermeyer, A., Fyffe, D., Heinemann, A. W., Dyson-Hudson, T. A., & Tulskey, D. S. (2020). Applying a bookmarking approach to setting clinically relevant interpretive standards for the spinal cord injury: Functional index/capacity (SCI-FI/C) basic mobility and self care item bank scores. *Archives of Physical Medicine and Rehabilitation*. <https://doi.org/10.1016/j.apmr.2020.08.026>
 30. Morgan, E. M., Mara, C. A., Huang, B., Barnett, K., Carle, A. C., Farrell, J. E., & Cook, K. F. (2017). Establishing clinical meaning and defining important differences for patient-reported outcomes measurement information system (PROMIS®) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. *Quality of Life Research*, 26(3), 565–586. <https://doi.org/10.1007/s11136-016-1468-2>
 31. Yorkston, K., & Baylor, C. (2019). Patient-reported outcomes measures: An introduction for clinicians. *Perspectives of the ASHA Special Interest Groups*. https://doi.org/10.1044/2018_PERS-ST-2018-0001
 32. Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The communicative participation item bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research: JSLHR*, 56(4), 1190–1208. [https://doi.org/10.1044/1092-4388\(2012\)12-0140](https://doi.org/10.1044/1092-4388(2012)12-0140)
 33. Cella, D., Lai, J.-S., Nowinski, C. J., Victorson, D., Peterman, A., Miller, D., Bethoux, F., Heinemann, A., & RubinCavazosReiderMoy, S. J. E. A. T. C. (2012). Neuro-QOL: Brief measures of health-related quality of life for clinical research in neurology. *Neurology*, 78(23), 1860–1867. <https://doi.org/10.1212/WNL.0b013e318258f744>
 34. Gershon, R. C., Lai, J. S., Bode, R., Choi, S., Moy, C., Bleck, T., Miller, D., Peterman, A., & Cella, D. (2012). Neuro-QOL:

- Quality of life item banks for adults with neurological disorders: Item development and calibrations based upon clinical and general population testing. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 21(3), 475–486. <https://doi.org/10.1007/s11136-011-9958-8>
35. Hula, W. D., Doyle, P. J., Stone, C. A., Austermann Hula, S. N., Kellough, S., Wambaugh, J. L., Rose, K. B., Schumacher, J. G., & St Jacques, A. (2015). The Aphasia Communication Outcome Measure (ACOM): Dimensionality, Item Bank Calibration, and Initial Validation. *Journal of speech, language, and hearing research: JSLHR*, 58(3), 906–919. https://doi.org/10.1044/2015_JSLHR-L-14-0235
 36. Cohen, M. L., Harnish, S. M., Lanzi, A. M., Brello, J., Hula, W. D., Victorson, D., ... Tulsky, D. S. (In Press). Establishing Severity Levels for Patient-Reported Measures of Functional Communication, Participation, and Perceived Cognitive Function for Adults with Acquired Cognitive and Language Disorders. *American Journal of Speech-Language Pathology*.
 37. Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multi-dimensional assessment of language function. *PLOS ONE*, 13(2), e0192773. <https://doi.org/10.1371/journal.pone.0192773>
 38. Swinburn, K., Porter, G., & Howard, D. (2004). Comprehensive aphasia test. *American Psychological Association*. <https://doi.org/10.1037/t13733-000>
 39. Cohen, M. L., Harnish, S., Lanzi, A., Brello, J., Victorson, D., Kisala, P. A., Nandakumar, R., & Tulsky, D. S. (2021). Adapting a PRO-bookmarking task to be accessible to adults with cognitive and language disorders. *Journal of Speech Language and Hearing Research*, 64, 4403–4412.
 40. Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
 41. Dalrymple-Alford, J. C., MacAskill, M., Nakas, C., Livingston, L., Graham, C., Crucian, G., Melzer, T. R., Kirwan, J., Keenan, R., Wells, S., & Anderson, T. (2010). The MoCA: Well-suited screen for cognitive impairment in Parkinson disease. *Neurology*, 75(19), 9. <https://doi.org/10.1212/WNL.0b013e3181fc29c9>
 42. Eadie, T. L., Yorkston, K. M., Klasner, E. R., Dudgeon, B. J., Deitz, J. C., Baylor, C. R., Miller, C. R., & Amtmann, D. (2006). Measuring communicative participation: A review of self-report instruments in speech-language pathology. *American Journal of Speech-Language Pathology*, 15(4), 307–320. [https://doi.org/10.1044/1058-0360\(2006/030\)](https://doi.org/10.1044/1058-0360(2006/030))
 43. Baylor, C., Oelke, M., Bamer, A., Hunsaker, E., Off, C., Wallace, S. E., Pennington, S., Kendall, D., & Yorkston, K. (2017). Validating the communicative participation item bank (CPIB) for use with people with aphasia: An analysis of differential item function (DIF). *Aphasiology*, 31(8), 861–878. <https://doi.org/10.1080/02687038.2016.1225274>
 44. HealthMeasures. (n.d.). List of Adult Neuro-QoL Measures. Retrieved from <https://www.healthmeasures.net/explore-measurement-systems/neuro-qol/intro-to-neuro-qol/list-of-adult-measures>
 45. Nagaraja, V., Mara, C., Khanna, P. P., Namas, R., Young, A., Fox, D. A., Laing, T., McCune, W. J., Dodge, C., Rizzo, D., & Khanna, D. (2018). Establishing clinical severity for PROMIS® measures in adult patients with rheumatic diseases. *Quality of Life Research*, 27(3), 755–764. <https://doi.org/10.1007/s11136-017-1709-z>
 46. Victorson, D. (2016). *Narrowing the Focus: Bookmarking Methodology in Rehabilitation*. Presented at the American Congress of Rehabilitation Medicine, Chicago, IL.
 47. de Riesthal, M., & Ross, K. B. (2015). Patient reported outcome measures in neurologic communication disorders: An update. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, 25(3), 114. <https://doi.org/10.1044/nnsld.25.3.114>
 48. Rothrock, N. E., Hays, R. D., Spritzer, K., Yount, S. E., Riley, W., & Cella, D. (2010). Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the patient-reported outcomes measurement information system (PROMIS). *Journal of Clinical Epidemiology*, 63(11), 1195–1204. <https://doi.org/10.1016/j.jclinepi.2010.04.012>
 49. Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102–109. <https://doi.org/10.1016/j.jclinepi.2007.03.012>
 50. Cohen, M. L., Weatherford, S., & Nandakumar, R. (2019). How normal are “normal” errors of language and cognition? *Journal of Speech, Language, and Hearing Research*, 62(5), 1468–1472. https://doi.org/10.1044/2018_JSLHR-L-18-0215
 51. Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., & Aaronson, N. K. (2006). Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*, 4(1), 70. <https://doi.org/10.1186/1477-7525-4-70>
 52. Sloan, J. A. (2005). Assessing the minimally clinically significant difference: Scientific considerations, challenges and solutions. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 2(1), 57–62. <https://doi.org/10.1081/COPD-200053374>
 53. HealthMeasures. (n.d.). Neuro-QoL reference populations. Retrieved from <https://www.healthmeasures.net/score-and-interpret/score-and-interpret-scores/neuro-qol/reference-populations>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.