Clinical Use of PROMIS, Neuro-QoL, TBI-QoL, and Other Patient-Reported Outcome Measures for Individual Adult Clients with Cognitive and Language Disorders

Matthew L. Cohen[1,2], **Alyssa Lanzi**[1], **Aaron J. Boulton**[2]

[1.]Department of Communication Sciences & Disorders; University of Delaware; Newark, DE, USA

[2.]Center for Health Assessment Research and Translation; University of Delaware; Newark, DE, USA

Abstract

Patient-reported outcome measures (PROMs) assess health outcomes from the patient's perspective. The National Institutes of Health has invested in the creation of numerous PROMs that comprise the PROMIS, Neuro-QoL, and TBI-QoL measurement systems. Some of these PROMs are potentially useful as primary or secondary outcome measures, or as contextual variables for the treatment of adults with cognitive/language disorders. These PROMs were primarily created for clinical research and interpretation of group means. They also have potential for use with individual clients; however, at present there is only sparse evidence and direction for this application of PROMs. Previous research by Cohen & Hula (2020) described how PROMs could support evidence-based practices in speech-language pathology. This companion article extends upon that work to present clinicians with implementation information about obtaining, administering, scoring, and interpreting PROMs for individual clients with cognitive/language disorders. This includes considerations of the type and extent of communication support that is appropriate, implications of the relatively large measurement error that accompanies individual scores and pairs of scores, and recommendations for applying minimal detectable change values depending on the clinician's desired level of measurement precision. However, more research is needed to guide the interpretation of PROM scores for an individual client.

Keywords

Patient-Reported Outcome Measures; Speech-Language Pathology; Evidence-based practice; Communication disorders

Corresponding author: Matthew L. Cohen, Ph.D., Assistant Professor, mlcohen@udel.edu, ph: 302-831-7071, 100 Discovery Blvd., 6th Floor, Newark, DE 19716; Alyssa Lanzi, Ph.D., CCC-SLP, Research Assistant Professor, lanzia@udel.edu; Aaron J. Boulton, Ph.D., Research Assistant Professor, aboulton@udel.edu.

Introduction to the PROMIS, Neuro-QoL, and TBI-QoL Measurement Systems

Patient-reported outcome measures (PROMs) quantify aspects of health status from the patient's perspective, without interpretation by a healthcare provider or anyone else about how the patient functions or feels.[1] PROMs can assess many different facets of health status from multiple angles, such as symptom frequency, intensity, or burden; perceived difficulty completing a task; or restricted participation because of a condition. They are particularly useful in quantifying constructs for which there are no physical or behavioral markers[2] or for which the patient is the only valid rater, as in the case of subjective, unobservable constructs like stigma.[3] Describing these questionnaires as outcome measures suggests that they assess the target of a treatment and are psychometrically sensitive to change, although the same measures may be used for other applications as well.

PROMs are not novel but were given a place of prominence beginning around 2000 with the formalization of evidence-based practice and person-centered care.[3–5] PROMs were developed primarily for research purposes, however, there is a growing need and interest from clinicians to use these measures in clinical settings. Unfortunately, the research to guide speech-language pathologists (SLPs) in their use of PROMs with individual clients is sparse. Cohen & Hula[3] broadly described how SLPs could use PROMs to support their delivery of evidence based practice. This article aims to extend that discussion to discuss implementation information for SLPs about obtaining, administering, scoring, and interpreting PROMs for individual clients with cognitive and language disorders. Specifically, this paper describes implementation information for clinicians related to three large sets of PROMs: the Patient Reported Outcomes Measurement Information System (PROMIS),[6] the Quality of Life in Neurological Disorders (Neuro-QoL) measurement system,[7,8] and the Traumatic Brain Injury - Quality of Life (TBI-QoL) measurement system.[9] There are other, similar PROMs that are relevant to speech-language therapy like the Communication Participation Item Bank (CPIB)[2] and Aphasia Communication Outcome Measure,[10] but these are discussed elsewhere,[11] including other articles in this issue.

As described further in Cohen & Hula[3] and by other researchers,[6,12,13] the PROMIS initiative was first funded by the National Institutes of Health (NIH) in 2004. In order "to see and understand the full picture of how a disease, condition, or treatment affects patients, researchers needed an accurate, evidence-based tool for measuring symptoms…from the patient's perspective(p.36)."[14] Rather than researchers using different PROMs to assess the same construct, like anxiety, the intention of PROMIS was to create a single system with multiple PROMs (or "item banks") that would measure various health constructs experienced by the general public, as well as by those with chronic conditions. An item bank is a set of individual items that measure the same construct of interest and have shared meaning across individuals. Item banks differ from conventional multi-item measures such that item banks typically contain a larger number of items but are often administered using select subsets of items using special procedures, explained in more detail below. At the time of this writing, within PROMIS there are 253 item banks for adults, 62 child self-report item banks (ages 8-17), and 58 parent proxy-report item banks (ages 5-17). Previous PROMs of

similar constructs still exist of course, and group-level scores from many of these legacy measures can be psychometrically linked to the PROMIS metric for comparison and data aggregation.[15]

The Neuro-QoL measurement system was developed around the same time as PROMIS by many of the same investigators, with the intention to measure constructs that apply more specifically to people with neurological conditions rather than to the general public. Originally, Neuro-QoL had 12 adult item banks that were calibrated on adults with stroke, multiple sclerosis, Parkinson's disease, epilepsy, and amyotrophic lateral sclerosis (ALS). There are now additional item banks that were calibrated on individuals with Huntington's disease. Most relevant to adults with cognitive and language disorders receiving speech/ language services, Neuro-QoL includes item banks that assess perceived cognitive function, ability to participate in social roles and activities, satisfaction with social roles and activities, speech difficulties, and swallowing difficulties. Neuro-QoL also includes a 5-item scale that assesses functional communication, but this would not be useful to clinicians because the scale is uncalibrated.

The last measurement system presented here is the Traumatic Brain Injury – Quality of Life (TBI-QoL) measurement system.[9] The TBI-QoL system is composed of 22 item banks and scales specific to adults with TBI that can also be interpreted as composite scores.[16] Most relevant to adults with cognitive and language disorders receiving speech/language services, the TBI-QoL system includes item banks that assess functional communication,[17,18] cognitive function generally and executive function specifically,[19] emotional and behavioral dyscontrol, ability to participate in social roles and activities, and satisfaction with social roles and activities.[20]

PROMs that comprise these measurement systems (PROMIS, Neuro-QoL, and TBI-QoL) were developed using similar methods and standards and were calibrated using the same item-response theory (IRT) models, which confer a number of measurement advantages.[21–23] For more information about the applications of IRT-based measurement in speech-language pathology, the reader is referred to other articles in this issue as well as Baylor et al.[24] and Yorkston & Baylor.[11] One characteristic of IRT that is important to highlight for the purposes of this article is that IRT models allow a person's score to be estimated in several different ways. Within an item bank there are often several questions (or "items") and, although every item could be administered, good measurement precision can be achieved with relatively small subsets of items in different combinations. Items in the bank differ from one another in terms of how well they assess different score ranges. For example, in an item bank of physical functioning, the item "I can get out of bed without assistance" is better for assessing someone with a low level of physical functioning, while the item "I can run a marathon" is better for assessing a person with a high level of functioning.

One way that clinicians can administer items from an item bank is by using "short forms," which are predetermined subsets of items. PROMIS, Neuro-QoL, and TBI-QoL short forms generally include 10 items or fewer. Unless otherwise specified, the short forms that are widely distributed assess the full range of the health construct. However, short forms

can also be constructed that target a specific score range (e.g., high versus low physical functioning) or a subpopulation. For example, the CPIB has a condition-generic 10-item short form,[2] but one that is specific to hearing loss is also under development.[25] In general, short forms with more items (e.g., 10 items) have better measurement precision than those with fewer items (e.g., 4 items).

Alternatively, some IRT PROMs can be administered by computerized adaptive test (CAT). CATs automatically report the PROM score, whereas the user has to use a lookup table to find the T-score and standard error of measurement associated with the manually calculated raw score on a short form. A CAT is a computerized administration of items from an item bank in a flexible way that "zeroes in" on the respondent's score generally with the same or fewer number of items than a short form and with better precision.[21,26] As the name indicates, CATs require a computer, often with an internet connection, which may not be feasible in some clinical settings. CAT administrations can be customized by the administrator to include different stopping rules. For example, the administrator could set the CAT engine to administer items until a desired level of measurement precision is reached (typically Standard Error of Measurement < 3 on the T score metric) or until a maximum number of items, if respondent time/burden is a salient factor. For a more detailed description of CATs, we direct readers to Fergadiotis, Casilio, & Hula [27].

Obtaining Measures

PROMIS and Neuro-QoL item banks and short forms are free for use and are available on healthmeasures.net. The reader is referred to the "Search and View Measures" feature specifically. Search results will include links to a scoring guide that includes a lookup table to convert raw scores to T-scores. TBI-QoL measures and scoring tables can be obtained by emailing tbi-QoL@udel.edu. Users of the Research Electronic Data Capture (REDCap) software can download PROMIS, Neuro-QoL, and TBI-QoL Measures from the REDCap shared library or import these measures directly into their project, including auto-scored and CAT versions. At the time of this writing, health systems using the electronic health record systems by Cerner or Epic (2017 version or later) may have access to many PROMIS measures, including CAT versions.[28] To learn about other ways to administer and score PROMs, readers are referred to healthmeasures.net.

We recommend that users search for specific PROMs by searching the more condition-specific PROM systems before the more condition-generic systems. That is, if the user is assessing a client with TBI, we recommend that the user first see if the construct that they want to assess is available within the TBI-QoL measurement system. If not, we recommend that they then search Neuro-QoL, then PROMIS. If the user is assessing a client with another (non-TBI) neurological condition, we recommend that they search Neuro-QoL then PROMIS. Some constructs (e.g., anxiety) that are assessed by multiple PROM systems, have item banks that are in fact psychometrically linked. For example, the TBI-QoL versions of some measures produce a score on the PROMIS or Neuro-QoL metric. The reader is referred elsewhere for more information about this topic.[9,15]

Administering PROMs to Adults with Cognitive and Language Disorders

PROMIS measures were developed to be valid for people with chronic conditions broadly. Neuro-QoL adult item banks were developed to be valid for people with acquired neurological conditions, but to our knowledge did not include stroke survivors with aphasia and did not systematically assess the cognitive abilities of respondents. Although more evidence for validity with particular cognitive and language conditions could be useful, we think there is sufficient evidence for validity for these measures to be used with adults with acquired cognitive and language disorders. Based on our own experiences and previous research, we recommend the following procedures for administering PROMs to adults with acquired cognitive and language disorders.

In order to make the PROMs more accessible, clinicians may want to make superficial changes to the appearance of the short form, by changing, for example, the font size or spacing on the page or by excluding distracting information, like the item ID.[29–31] Additionally, the phrasing of the response option might be superficially modified to be grammatically correct. For example, some items ask, "how much difficulty…" with one response option being "none" (as in "none difficulty") instead of "no difficulty." Clinicians should be careful not to change the meaning of the item or response options when making such modifications.

Research conducted thus far has supported the use of an administrator to help support the respondent's understanding of and response to PROMs.[32] Although it is possible that the presence of an administrator could bias the respondent's responses, previous research indicates that the bias is small when assessing relatively non-sensitive constructs.[33,34] However, when feasible, a person other than the treating clinician might administer the PROMs in order to minimize any pressure on the client to report feeling better or worse than they are (e.g., to make the clinician feel good or to justify more treatment).

For clients with cognitive and language disorders, the administrator should be trained to provide multimodal communication support guided by the principles of supported conversation, such as using simple sentences, writing or circling key words and phrases, gesturing, summarizing, asking yes or no confirmation questions, observing non-verbal language, and allowing plenty of time to respond.[35–37] Tucker et al.[32] proposed a specific system of communication support for PROMs that has been used successfully with adults with post-stroke aphasia.[38,39] This hierarchy includes (a) repeating the question and choices, (b) simplifying and restating the question and reviewing the choice scale, (c) re-explaining the entire choice scale and repeating the question, (d) combining a yes-no question with the scale and response options, and (e) presenting the next question. An administrator can help the respondent understand relatively superficial aspects of the item, but should not define a core concept for which the respondent's subjective interpretation is the point of the assessment.[40] For example, one item in the Neuro-QoL Satisfaction with Social Roles and Activities item bank is "I am satisfied with my ability to do things for fun outside my home." The response options are *not at all*, *a little bit*, *somewhat*, *quite a bit*, *very much*. It would appropriate for the administrator to rephrase word "satisfied" (e.g., "feel good about"). It would also be ok for the administrator to repeat/emphasize

that the response options indicate different levels of agreement with the statement, for example, that "somewhat" communicates stronger agreement than "a little bit." But the administrator should not try to concretize the difference between the two because the point of the assessment is to capture the patient's subjective experience of the intensity/frequency of their symptoms. We acknowledge that this may be a different appraisal process for adults with cognitive and language disorders than for neurotypical adults[41], and more research is needed in this area.

To support the language skills of clients, one useful strategy is to allow nonverbal responses (e.g., by pointing to a response option). Alternatively, the respondent could hold up a certain number of fingers to indicate their response (refer to Figure 1). This is especially useful when administering PROMs by video call. Clinicians, however, should be attentive to the raw score values associated with each response because items with the same response option may be scored in different directions from one another.

Despite these accommodation efforts, some respondents may be unable to complete some or all PROMs depending on how cognitively/linguistically demanding the PROM is.[3] It is reasonable to think that clients who are unable to provide informed consent for treatment are also unable to provide thoughtful responses to PROM items. However, that is a relatively low bar for many clients, and it remains largely up to the clinician's judgement to decide whether the client is able to provide thoughtful responses. It is worth remembering, however, that a client's perspective does not need to cohere with the perspective of the clinician or a caregiver in order to be "valid." There are instances when the client's report may be inaccurate, for example, if a client with hemiplegia reports that they can easily run a marathon. But even in these instances, the client's report should be assumed to be a valid reflection of what they perceive to be true. Ultimately, PROMs are not intended to be the only source of clinical measurement; they are most useful when combined with other clinical information like performance-based and clinician-reported assessments.[3]

The clinician may want to solicit the perspective of a proxy respondent, although the use of one is controversial (see other articles within this issue). Specific to PROMIS and Neuro-QoL measures, the HealthMeasures website says that a caregiver proxy can be used with the following prompt: "The following questionnaires will ask about your care recipient's symptom and activity levels; about his/her ability to think, concentrate and remember things; questions specific to his/her condition; and questions related to his/her quality of life. Please answer the following questions *based on what you think your care recipient would say* (emphasis added)".[40] An alternate kind of proxy report (also called an informant report) is where the respondent provides their own perspective of the patient's condition.

Score Interpretation

PROMIS, Neuro-QoL, TBI-QoL, and similar PROMs produce norm-referenced standardized scores, meaning that a client's raw score is expressed in terms of how it relates to scores produced by a comparison or reference sample. Most normative samples are normally distributed, meaning that the scores form a symmetrical, bell-shaped curve. Figure 2 shows an example of a normal distribution where the x-axis represents the various PROM

scores that were produced by members of the normative sample, and the y-axis represents the number of people that produced that score. The mean, median, and mode are the same value, which is the score produced by the most people.

Normal distributions have predictable properties, such as the percentage of scores that fall within various distances from the mean. Typically, the distance from the mean is communicated in terms of standard deviation (SD) units, where nearly 100% of the distribution falls between −3.0 and +3.0 SD from the mean. A Z-score is the most basic standardized score that expresses how many SD units separate that client's score from the reference sample mean. For example, a Z-score of −1.0 means the client's score was 1 SD below the mean of the reference sample (see the person in blue in Figure 2). For normative samples that are normally distributed (i.e., form a bell-shaped curve), this means the Z-score of −1.0 is greater than or equal to 16% of scores produced by the reference sample (the 16th percentile).

The distance between a particular client's score and the mean of the normative sample can be reported in other ways as well. In fact, most PROMs describe a client's performance using a T-score, which is simply a linear transformation of a Z-score. Whereas interpretation of Z-scores involves negative numbers and decimal places, a T-score generally does not. To convert from a Z-score to a T-score, one simply multiplies the Z-score by 10 and adds 50 ($T = Z(10)+50$). Similarly, to convert in the other direction: $Z = (T-50)/10$. The reader is referred to Figure 2 for a comparison of standardized scores

How to Interpret a Single PROM Score

There are two major ways that PROM scores are interpreted for an individual client. One is by interpreting a single score, for example, from a single assessment session before treatment. The second is by comparing how much better or worse a score is than a previous score, for example, from before to after treatment. Ways to answer those questions are discussed below. Much of this information comes from healthmeasures.net – the online hub for PROMIS, Neuro-QoL, and related measurement systems – but with more elaboration and emphasis on content that is most relevant to SLPs treating adult clients with cognitive and language conditions.

Distance from the normative sample mean

Norm-referenced standardized scores, like T-scores, express the distance between a particular client's score and the mean of the normative sample. Because the normative sample typically forms a bell-shaped, normal curve, scores that are increasingly far from the mean are less common. Thus, distance from the mean communicates the typicalness of the score compared to the reference group, which is often interpreted as a marker of how remarkable the score is. For example, on a PROM of perceived cognitive function (where higher scores indicate better perceived functioning), a score of T=40 is −1.0 SD (16th percentile) below the reference sample, which indicates that the person perceives their cognition to be worse than 84% of how people in the reference sample perceive their cognition. Of course, the critically important consideration is the composition of the reference sample.

Many scores reference a normative sample whose demographics mirror the general population, for example, mirroring results from a recent U.S. Census. This is true for most PROMIS and Neuro-QoL scores.[6–8] For those general-population reference samples, distance from the mean has intuitive meaning. For example, if someone's symptoms of anxiety are greater than what is experienced by 98% of others in the U.S., it makes sense to interpret that level of anxiety as being unusually high. The website healthmeasures.net includes a number of guides for interpreting PROMIS and Neuro-QoL PROMs. For some PROMs, only scores in a single direction are interpreted as being clinically relevant. For example, only scores below T=45 on the PROMIS v2.0 Cognition Item Bank are interpreted (Figure 1, Interpretation A): mild (45>T>40), "moderate" (40>T>30), or "severe" (T>30). Anything above 45 is considered "within normal limits." On the other hand, PROMs of other constructs, like PROMIS Bank v2.0 - Satisfaction with Social Roles and Activities, are interpreted both above and below the mean (Figure 1, Interpretation C): very low (T<30), low (30<T<40), average (40<T<60), high (60<T<70), or very high (T>70).

Rather than referencing a general population sample, other PROMs reference a sample of people with a particular medical diagnosis or clinical condition (e.g., people with traumatic brain injury [TBI] or aphasia). These reference samples are most often used when the construct being measured is not widely experienced by the general population (e.g., bladder and bowel management problems) and especially when the primary application of the measure is to determine greater than/less than relationships between people or timepoints, rather than to determine whether the level of the construct is typical or not. For example, the ACOM references a sample of people with post-stroke aphasia.[10] It measures perceived functional communication, not primarily to determine *whether* the client has aphasia, but rather to measure relative differences between people with aphasia or between timepoints. For clinical reference samples like those, distance from the mean is still very useful but has less intuitive meaning. For example, an ACOM score of T=60 means the client's perceived functional communication is one SD above the mean of people with aphasia who were in the reference sample, which Hula and colleagues[10] reported was 329 people with aphasia, 73% of whom had ischemic stroke, 52% of whom had aphasia only (no motor speech disorder), etc.

Whereas for a general population sample, T=50 suggests normal/average/typical symptoms or function, this is not necessarily true for clinical reference samples, where T=50 might represent mild or moderate symptoms or impairment. Understanding the composition of the reference sample is important when interpreting distance from the mean. Further, if there are multiple reference samples to choose from, the demographic and clinical characteristics of the samples can help the clinician decide which one is most appropriate to interpret their client's score.

PRO-Bookmarking

In addition to distance from the mean of the reference sample, there are other ways of interpreting a PROM score. PRO-bookmarking is a qualitative method that allows stakeholders (e.g., patients, care partners, clinicians) to determine how score ranges should be interpreted. This can be especially helpful when the distance from the mean is less

intuitive, as with clinical reference samples. PRO-bookmarking is a relatively new method, reported in 7 papers since 2014.[42–48]

There are two main stages of this research. The first stage involves the construction of "vignettes" about hypothetical clients who would produce a particular PROM score, for example, at 0.25 or 0.5 SD intervals. Table 1 depicts the most likely responses for each item in the CPIB generic short form that would be selected by clients at each level of communication participation (T=25, 30, etc.). These are calculated based on the responses produced by the reference sample. Vignettes are then written about hypothetical people at each level by phrasing their most-likely responses as a symptom report. For example, we could say that "Ms. Williams'" (T=35) communication condition interferes *very much* with getting her turn in a fast-moving conversation, with giving someone detailed information, and with talking with someone she doesn't know. Her condition interferes quite a bit with communicating with someone she knows and with communicating in a small group of people.

The second stage of the PRO-bookmarking process asks stakeholders to review and discuss the vignettes and determine whether that client's impairment, limitation, or restriction should be described as *normal/minimal*, *mild*, *moderate*, or *severe*. For example, small panels of people with aphasia (or another condition) would review each vignette and, with assistance of a moderator, come to consensus about whether Ms. Williams' experience represents normal communication participation or mild, moderate, or severe levels of restricted participation. The end result would look like Interpretations A, B, or C (Figure 2) but would not necessarily map as neatly onto standard deviation units.

For the most part, PRO-bookmarking has not yet been reported for PROMs that are strongly relevant to speech-language pathology. One exception is a study by Rothrock et al.[45] that reported cut point PROM scores on the PROMIS Item Bank v2.0 – Cognitive Function for clients with cancers (or cancer treatments) that affect cognition. One group of clients and one group of clinicians produced cut points that slightly disagreed with one another but grossly resembled the cut points shown in Figure 1, Interpretation A.[45] Cook et al.[43] reported cut points for people with multiple sclerosis for Neuro-QoL measures of fatigue, physical functioning, and sleep disturbance. These may be relevant variables for clients and research participants with cognitive/language conditions but are not likely to be outcome measures. Interested readers are referred to several papers by Cohen et al. that are expected to be in press in 2021 and will report how the PRO-bookmarking materials and procedures were adapted for adults with neurogenic communication disorders, as well as the resulting cut points for the ACOM, CPIB, Neuro-QoL v2.0 Cognitive Functioning, Neuro-QoL v1.0 Ability to Participate in Social Roles and Activities, and the TBI-QoL Communication Item Bank.

Patient Acceptable Symptom State (PASS)

The PASS estimate is a score associated with most clients feeling satisfied with their health in that domain. It is established after asking a number of people with the condition to answer an anchor question after completing a given PROM. For example, after answering a PROM about pain, the respondent would be asked something like "taking into account all

the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?".[49] The PASS is defined as the PROM score that is associated with at least 75% of respondents indicating that the state of their health in that domain is satisfactory. To our knowledge, this method has not yet been applied to PROMs that are strongly relevant to adults with cognitive/language conditions.

How to Interpret PROM Score Change

In order to understand how to interpret PROM score change, for example, investigating improvement from baseline, it is important to first understand measurement error. The observed score that a client produces on any assessment is not necessarily a perfect reflection of that person's actual ability/function/symptoms/impairment (i.e., their "true score"). The observed score is subject to error from one or more sources: the measure itself, the assessment environment, the test administrator, the construct being measured, or another source. The degree to which a measure is free from error is known as its reliability or precision. IRT-based PROMs have levels of precision that vary within the construct being measured; they are generally more precise towards the mean of the normative distribution and less precise with more extreme scores.

All observed scores have a level of error that is associated with them. The error associated with a PROMIS or Neuro-QoL PROM score can generally be found in the same lookup table that the clinician uses to convert from raw scores to T-scores on a short form. This same error – called the standard error of measurement (SEM) – is what is used to compute a confidence interval around the observed score. The confidence interval expresses how near or far the true score is likely to be from the observed score. Highly reliable tests produce observed scores that are more accurate (i.e., more consistently close to the true score), as reflected by a smaller SEM and confidence interval. Less reliable tests produce scores that are less accurate (i.e., less consistently close to the true score), as reflected by a larger SEM and confidence interval. If one were to administer a PROM to the same person repeatedly (with no testing effects or change in the construct that is being measured), a highly reliable test will produce scores that are the same or highly similar to one another. A less reliable test will produce scores that are more different from one another, without any change in the person's true score.

Minimal detectable change

If a clinician were to determine whether a client's true score changed, as a result of an intervention, for example, it would be important to make sure that their score change was greater than the amount of score variability that could result from that measure's error alone. The main way to do this for PROM scores is by applying a formula for what is referred to as minimal detectable change (MDC)[50–53] or what some might call "statistically significant change." There are different MDC thresholds that could be computed depending on how statistically rigorous the user wants to be. The most commonly used MDC value is based on 95% confidence in both directions ("two-tailed;" meaning without a hypothesized direction of change). The formula for this is $MDC_{95} = [SEM_{score1} + SEM_{score2}]/2) * (1.96) * SQRT(2)$. The American Speech-Language-Hearing Association (ASHA) is currently

developing tools for interpreting MDC_{95} for the PROMs that are part of its National Outcomes Measurement System (T. Schooling, personal communication, October 27, 2020).

Figure 3 is reproduced with permission from Kozlowski et al. (2016) and shows a tabular way to determine MDC_{95} values for the Neuro-QoL v2.0 – Cognitive Function short form. The user finds the column associated with the first (e.g., baseline) score and then finds the row associated with the second (e.g., discharge) score. An increase in PROM score that is below the red diagonal band is 95% likely to reflect a "real" change," that is, an increase in the true score. A decrease in PROM score that is above the red diagonal band is 95% likely to reflect a decrease in the true score. PROM score changes that fall within the diagonal band are not sufficient to exceed the 95% threshold. It does not necessarily mean that the client did not experience real change, but it was not by enough to be statistically confident in the measurement of that change. Kozlowski et al.[52] present similar tables for other Neuro-QoL short forms, and supplementary material for that article includes an Excel-based calculator for users to enter score 1 and score 2 and determine the MDC.

However, there is reason to think that a MDC_{95} may be too strict for most clinical applications. For example, Figure 3 shows that a client who produces a baseline score of T=40 needs to demonstrate an 8.3-point (0.83 SD) improvement in order to exceed the MDC_{95} threshold, which is an unusually strong effect for cognitive/language interventions.[54,55] It may be more appropriate for the clinician to apply a less strict criterion of 2-tailed 90% confidence: $[SEM_{score1} + SEM_{score2}]/2) * (1.64) * SQRT(2)$[56,57,58(p2)]. If there is a hypothesized direction of score change, which generally there is in treatment settings (i.e., improvement), the clinician could apply a MDC with 1-tailed 90% confidence $[SEM_{score1} + SEM_{score2}]/2) * (1.28) * SQRT(2)$. Compared with the scenario above in which a score change of 8 is required for a $MDC_{95\ 2\text{-tailed}}$, a $MDC_{90\ 1\text{-tailed}}$ only requires a score change of 4.7 (.47 SD).

Minimally important difference

Whereas the MDC value indicates what score change is required for "statistically significant change," a complementary way to understand PROM score change is associated with a "clinically significant change."[59,60] This could be the minimum change value that is required for most clients to *perceive* an overall improvement. There is a lot of disagreement in the literature about what to call this value, much less the best way to estimate it. For this article, we will use the same term as healthmeasures.net, which refers to this as the minimally important difference (MID). However, it is also called minimal clinically important difference (MCID), subjectively significant difference, clinical significance, minimally detectable difference, and many other terms.[59] This value is estimated a number of different ways, although ultimately a single threshold will not represent meaningful change across different people, conditions, contexts, or levels of severity.[60]

The most common way to estimate the MID value is to compare PROM score change against an anchor that is external to the PROM. For example, if one were to administer a PROM of perceived cognitive function at multiple timepoints, the anchor could be a retrospective global rating of change such as, "Overall, since you began cognitive rehabilitation, how have your cognitive abilities changed?" (response options = a lot worse, a

little worse, no change, a little better, a lot better). With dozens or hundreds of respondents, it could be estimated what mean PROM score change was associated with respondents who also said "a little better" overall. This would be interpreted as the PROM score change associated with a small noticeable improvement.

MID values can be very useful when interpreting group data, such as from a clinical trial where even small effects could be statistically significant because of relatively small measurement error. MID estimates allow researchers to determine how many people in the trial experienced a score change that is believed to be perceptible and meaningful. At the time of this writing, healthmeasures.net reports that MID thresholds for PROMIS and Neuro-QoL measures are generally between 0.2 – 0.6 SD (2.0 – 6.0 units on the T-score metric).

MID thresholds may also be useful when matching clients to treatments. For example, an orthopedic surgery group reported that candidates for a particular surgery whose self-reported physical function was above T=42 on PROMIS Physical Function were only 6% likely to experience a clinically meaningful improvement (i.e., produce a T-score after surgery that exceeded the MID threshold).[61] Similarly, this same group reported that surgical candidates with a PROMIS Pain Interference T-score below 55 were only 5% likely to obtain meaningful pain relief. In this way, individual clients can be assessed at baseline and evaluated for their likely benefit based on contextual and clinical variables. To our knowledge, this has not been done in the field of speech-language pathology, but it seems like a worthwhile endeavor. For example, two clients with post-TBI pragmatic language challenges could be matched to one treatment over another based on relevant variables, e.g., as assessed by TBI-QoL scores of executive functioning, behavioral dyscontrol, or functional communication.

It is less clear whether and how the MID should be used retrospectively, once a client is enrolled in therapy. Because the error associated with two scores from an individual is relatively high, the MID is likely to be smaller than the MDC. This means that there is not enough measurement precision to detect a change that is perceptible to clients. The logic of the MID is also questionable when applying group-derived indices of meaningful change to an individual client. The MID is an attempt to humanize score change, to anchor quantitative ratings against what is meaningful to a representative sample of people with the condition. But for an individual client, the clinician simply could ask them whether they perceive a meaningful change. If the client denies experiencing meaningful change, the perspectives of other people with that condition are largely irrelevant. The MID may serve as an estimate for what PROM score would be required for *most* clients to notice, but ultimately the client in question can tell you whether they notice an improvement or not.

Cook et al.[62] proposed a clever way of estimating MID thresholds prospectively rather than retrospectively (which can be biased)[63,64] that they called the Idio-Scale Judgement method. In that study, 500 participants with multiple sclerosis (MS) rated their own fatigue on the Neuro-QoL Fatigue item bank. They were then shown vignettes that were 0.2 SD units away from their own score and asked to decide whether each symptom level would represent a

significant improvement or decline that would make a difference in their own daily life. The smallest change rated by that client as such was interpreted as their prospective MID.

Although Cook et al.[62] ended up pooling these values to create group mean estimates, we think a similar method could be very useful for clinicians wanting to estimate the MID of individual clients. As part of the initial evaluation, clinicians administering a short form or full item bank could ask their client to complete the measure as usual, indicating current function or symptoms. Using the same form (but maybe a different color pen) they could then ask the client to indicate the smallest change in one or more PROM items that they would accept as a meaningful improvement (e.g., to make a difference in their daily life). The score change associated with that change could be interpreted as *that client's* MID. This seems particularly in the spirit of PROMs, which are intended to assess health from the patient's perspective. This is similar to what Zeppieri & George[65] called "patient-determined success criteria." These authors studied 225 consecutive patients with musculoskeletal pain in an outpatient sports medicine physical therapy clinic and found that clients were quite consistent in their success criteria over time. Interestingly, the criterion score set by most clients (assessed prospectively) was higher than published MID values (assessed retrospectively). The reason for this is not entirely clear but it may be the case that clients are more satisfied with smaller gains than they would have predicted. Changing standards of comparison are discussed later in a section on "response shift."

The prospective setting of patient-centered success criteria is similar to goal attainment scaling (GAS), in which the client and clinician collaboratively develop a quantifiable, measurable rating scale of progress towards a personalized goal. [66] PROMs are supposed to have content that is representative of a health construct (e.g., communication participation broadly), whereby GAS goals are more granular and specific to the person (e.g., ordering at a restaurant). The two methods of measure serve as interest complements of one another, however, PROM score change may be better reflect generalized improvements from therapy. For example, improving a client's ability to order at a restaurant affected enough other behaviors so as to be detectable on a general PROM of communication participation.

Practical advice

PROMIS and Neuro-QoL were originally intended for clinical research purposes[14] which most often involve groups of clients and relatively small measurement error. Although there is great interest in using them for clinical practice purposes, with the exception of Kozlowski et al.,[52] there is very little concrete guidance, at the time of this writing, about how to use these measures in the service of an individual client and in the context of relatively large measurement error. Below is a summary of our recommendations for current practice with individual clients who have cognitive/language disorders, acknowledging that better data and more consensus in the field may yield better recommendations in the coming years. For example, the journal Quality of Life Research is planning a special issue on assessing meaningful change.

1. *Choose reliable measures*. If the clinician is going to measure something, they should do it with the highest level of measurement precision that is feasible for the client and setting. For assessing individual change over time, the test-

retest reliability is particularly important. Intraclass correlation coefficient (ICC) values between 0.75- 0.90 are considered "good," and those over 0.90 are considered "excellent".[67] More reliable assessments generally require more time to administer, but the resulting scores are associated with smaller confidence intervals and are better able to reflect real change when it happens. Concretely, this means using a CAT when feasible with a requirement for a reasonably small error (e.g., SEM < 3 or smaller on the T score metric), because a CAT can be more precise even if the same or fewer items are administered than exist on a comparable static short form [68]. Greater reliability can also be achieved by choosing a longer short form (e.g., 10-item instead of 4-item) from healthmeasures.net. .

2. *Consider using a one-tailed MDC threshold to assess reliable PROM change in one direction.* For most clinical contexts, a MDC_{95} is probably too strict for interpretation of individual clients' PROM change scores. For example, Lapin et al.[69] reported on 337 individuals 1 month and 6 months after stroke who completed four PROMIS measures alongside global ratings of change. In the domain of physical functioning, 66% of participants endorsed at least minimal improvement on the global rating of change, 53% produced a change score on PROMIS Physical Functioning that exceeded the MID for that measure, but only 22% produced a change score that exceeded the MDC_{95} threshold (6.5 T-score units). A $MDC_{90\ 1\text{-tailed}}$ may be more appropriate for interpretation of an individual client's PROM score change when there is a directional hypothesis (e.g., that the client's score will improve with rehabilitation services). A reanalysis of the Lapin et al.[69] data with a $MDC_{90\ 1\text{-tailed}}$ (3.0 T-score units) revealed that approximately 50% of participants would have exceeded that threshold, which is closer to those who reported minimal global improvement and very close the number of people whose change scores met criteria for the MID. There is a tradeoff, similar to the tradeoff between sensitivity and specificity of a cutoff score, between failing to detect real change when it happens because the threshold was too high (false negative), and overinterpreting score change due to error because the threshold was too low (false positive). There is no "right" answer about how strict versus lenient the criterion should be. Our recommendation for most clinical settings is the $MDC_{90\ 1\text{-tailed}}$.

    a. Given the MDC_{95} value, the $MDC_{90\ 1\text{-tailed}}$ value can be computed like this: $MDC_{90\ 1\text{-tailed}} = [MDC_{95} / 1.96] * 1.28$. The RehabMeasures Database is an excellent resource for finding MDC_{95} values, and ASHA is currently developing tools for interpreting MDC for the PROMs that are part of its NOMS (T. Schooling, Personal Communication, October 27, 2020).

    b. Given the two PROM scores and their associated SEM, the $MDC_{90\ 1\text{-tailed}}$ can be computed with this formula: $[SEM_{score1} + SEM_{score2}]/2) * (1.28) * SQRT(2)$.

    c. Given the test-retest reliability of the measure, the $MDC_{90\ 1\text{-tailed}}$ can be computed like this: [(10 * SQRT(1-ICC)) * 1.81]. Some have advocated for using estimates of internal consistency as the estimate of reliability, but we agree with those who advocated that the test-retest estimate is more appropriate.[70]

3. *Consider using the patient-centered Idio-Scale Judgement method* that we proposed above to estimate the MID for an individual client.

4. *Look for trends over time.* If a PROM is administered at several timepoints (e.g., every 5 sessions), a continuing positive trend likely reflects real change, whereas erratic fluctuations around the baseline score are more likely due to error. However, a positive trend could also be influenced by confounding variables such as social desirability bias and response shift (discussed below).

5. *Recognize that PROMs are intended to be interpreted alongside other data.* PROM data are intended to complement, not replace, other sources of information about a client. As with all clinical data, they require context and interpretation by a knowledgeable clinician, and are not necessarily valid or useful for all clients and situations.[3]

Response shift

PROMs assess a person's perception ("how often do you experience that symptom?") or evaluation (e.g., how difficult is it for you to [do that activity]?") of their health[71] The cognitive processes that underlie the appraisal of one's health are not entirely clear for neurotypical people, much less those with cognitive and language disorders, but it is believed that at least four components are involved: (a) a frame of reference, (b) a strategy by which individuals recall relevant details from memory, (c) a standard against which they appraise their health (e.g., other people with the condition or their health before the onset of the health condition), and (d) some algorithm that maps these components onto a response option[72–74] "Response shift" is what happens when a PROM score changes not because of a change in health status but because of a change in the respondent's appraisal process. For example, when asked about cognitive or communication challenges, an individual recovering from a stroke in an inpatient rehabilitation setting may compare their current health status to themself the week before, or to other people in the unit. However, when the individual returns home and to their usual routines and relationships, they may begin to see their health status as compared to their life before the stroke. A decline in PROM score when the individual returns home may not reflect a real decline in cognition or communication, but rather a change in their standard of comparison or the way the person defines concepts on the PROM like "difficult" or "better" or "a lot" or "health." Significantly more research is needed about response shift and whether some manifestations of response shift are predictable or consistent across patient groups or settings.

Concluding Thoughts

Although the PROMIS, Neuro-QoL, and TBI-QoL measurement systems were designed for clinical research, great strides have been made to apply them to clinical practice

settings.[75–81] ASHA has begun rolling out the inclusion of PROMs in its NOMS, which aims to provide useful group-level interval data to complement its ordinal functional communication measure data. However, there is limited research from any healthcare field about how PROM scores might be used with an individual client. This paper presented some ideas for how they might be applied to an adult client with a cognitive/language condition in speech-language therapy.

By assessing enough representative behaviors, impairments, and/or symptoms, a PROM produces a single score that describes a person's overall level of a health construct. For this reason, IRT-based PROMs are a very time-efficient way of measuring contextual variables that are relevant to rehabilitation but that may not change with therapy – like sleep, fatigue, depression, mobility, and social support. For use as primary outcome measures, however, their application is more nuanced. Often, a clinician sets one or more goals in therapy to address very specific or idiosyncratic behaviors, impairments, or symptoms that may not match PROM items verbatim. Even if the client's goal does closely match one or more PROM items, it is not clear if the client achieving their goal would "move the needle" enough to be reflected in a detectable change in overall PROM score. For example, in Table 1, Mr. Brown (who produces a CPIB T = 40) reports that his communication condition interferes "quite a bit" with his ability to communicate in a small group of people. If he addressed that challenge in therapy and his response to that item changed to "my condition interferes… a little" or "…not at all," the CPIB short form lookup table (not shown) indicates that his overall T-score would only improve +1 or +2 to T = 41 or T = 42, respectively. Thus, a change in that item alone is probably not sufficient to exceed published MDC or MID thresholds; improvement would also need to be reflected on other items. Consider the response differences across columns in Table 1, which illustrates most-likely responses at 0.5 SD intervals. In general, respondents need to change responses associated with 3 or 4 PROM items on the 10-item short form in order to move 0.5 SD. And because the categories are ordinal, it is not clear how much clinical change is required to jump a response category or if that amount of change is consistent across items.

For this reason, PROMs may be better able to detect changes associated with some therapy goals than with others: goals that match the content specificity of the PROM and are likely to result in smaller changes in several PROM items or bigger changes in fewer PROM items. For instances where they do not have the content specificity to be useful as a primary outcome measure, PROMs may be more useful as a secondary outcome measure that captures generalization of the therapy to multiple representative behaviors of the health construct and not just to the target behavior. For example, the client's progress toward their goal attainment in everyday life may be documented with something like goal attainment scaling (GAS)[82] while generalization may be assessed with performance-based measures and PROMs.

For example, semantic feature analysis is a therapy approach used to improve the naming of nouns, verbs, or both. The proximal goal of therapy may be to improve the naming of a closed set of words with the more distal goals that there would be improvements in naming untrained items and in the client's ability to participate in everyday communicative interactions. The naming of trained words, however successful (e.g., as measured by number

of correct words or response latency), may not generalize enough to affect the overall construct of communication participation as measured by the CPIB generic 10-item short form (Table 1). However, this does not necessarily mean that the therapy was unsuccessful.

One future direction to improve the clinical use of PROMs might be to capitalize on the flexibility of IRT-based measures to create custom short forms for each client. The client and clinician could jointly pick PROM items that most closely align with the client's challenges, needs, and therapy goals and use the patient-centered MID proposed herein to determine that client's success criterion. More research and development, however, are needed to guide the best use of PROMs for individual clients, and the interested reader is advised to keep searching the literature over the coming years for more development in this area.

Acknowledgments

References

1. Patrick DL, Burke LB, Powers JH, et al. Patient-Reported Outcomes to Support Medical Product Labeling Claims: FDA Perspective. Value in Health. 2007;10(s2):S125–S137. [PubMed: 17995471]

2. Baylor C, Yorkston K, Eadie T, Kim J, Chung H, Amtmann D. The Communicative Participation Item Bank (CPIB): item bank calibration and development of a disorder-generic short form. J Speech Lang Hear Res. 2013;56(4):1190–1208. doi:10.1044/1092-4388(2012/12-0140) [PubMed: 23816661]

3. Cohen ML, Hula WD. Patient-Reported Outcomes and Evidence-Based Practice in Speech-Language Pathology. Am J Speech Lang Pathol. 2020;29(1):357–370. doi:10.1044/2019_AJSLP-19-00076 [PubMed: 32011905]

4. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. National Academy Press; 2001.

5. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence Based Medicine: What It Is And What It Isn't: It's About Integrating Individual Clinical Expertise And The Best External Evidence. BMJ: British Medical Journal. 1996;312(7023):71–72. [PubMed: 8555924]

6. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010;63(11):1179–1194. doi:10.1016/j.jclinepi.2010.04.011 [PubMed: 20685078]

7. Cella D, Lai J-S, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology. 2012;78(23):1860–1867. doi:10.1212/WNL.0b013e318258f744 [PubMed: 22573626]

8. Gershon RC, Lai JS, Bode R, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. Qual Life Res. 2012;21(3):475–486. doi:10.1007/s11136-011-9958-8 [PubMed: 21874314]

9. Tulsky DS, Kisala PA, Victorson D, et al. TBI-QOL: Development and Calibration of Item Banks to Measure Patient Reported Outcomes Following Traumatic Brain Injury. J Head Trauma Rehabil. 2016;31(1):40–51. doi:10.1097/HTR.0000000000000131 [PubMed: 25931184]

10. Hula WD, Doyle PJ, Stone CA, et al. The Aphasia Communication Outcome Measure (ACOM): Dimensionality, Item Bank Calibration, and Initial Validation. J Speech Lang Hear Res. 2015;58(3):906–919. doi:10.1044/2015_JSLHR-L-14-0235 [PubMed: 25812178]

11. Yorkston K, Baylor C. Patient-Reported Outcomes Measures: An Introduction for Clinicians. Perspectives of the ASHA Special Interest Groups. Published online January 11, 2019:1–8. doi:10.1044/2018_PERS-ST-2018-0001

12. Amtmann D, Cook KF, Johnson KL, Cella D. The PROMIS initiative: involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. Arch Phys Med Rehabil. 2011;92(10 Suppl):S12–19. doi:10.1016/j.apmr.2011.04.025 [PubMed: 21958918]

13. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care. 2007;45(5 Suppl 1):S3–S11. doi:10.1097/01.mlr.0000258615.42478.55

14. National Institutes of Health. A Decade of Discovery: The NIH Roadmap and Common Fund; 2014. https://commonfund.nih.gov/sites/default/files/ADecadeofDiscoveryNIHRoadmapCF.pdf

15. PROsetta Stone: Linking Patient-Reported Outcome Measures. Accessed December 13, 2020. prosettastone.org

16. Tyner CE, Boulton AJ, Sherer M, Kisala PA, Glutting JJ, Tulsky DS. Development of Composite Scores for the TBI-QOL. Archives of Physical Medicine and Rehabilitation. 2020;101(1):43–53. doi:10.1016/j.apmr.2018.05.036 [PubMed: 31875840]

17. Cohen ML, Kisala PA, Boulton AJ, Carlozzi NE, Cook CV, Tulsky DS. Development and Psychometric Characteristics of the TBI-QOL Communication Item Bank. Journal of Head Trauma Rehabilitation. 2019;34(5):326–339. doi:10.1097/HTR.0000000000000528 [PubMed: 31498231]

18. Cohen ML, Tulsky DS, Boulton AJ, et al. Reliability and Construct Validity of the TBI-QOL Communication Short Form as a Parent-Report Instrument for Children with Traumatic Brain Injury. Journal of Speech Language and Hearing Research. Published online In Press.

19. Carlozzi NE, Tyner CE, Kisala PA, et al. Measuring Self-Reported Cognitive Function Following TBI: Development of the TBI-QOL Executive Function and Cognition-General Concerns Item Banks. Journal of Head Trauma Rehabilitation. 2019;34(5):308–325. doi:10.1097/HTR.0000000000000520 [PubMed: 31498230]

20. Heinemann AW, Kisala PA, Boulton AJ, et al. Development and Calibration of the TBI-QOL Ability to Participate in Social Roles and Activities and TBI-QOL Satisfaction With Social Roles and Activities Item Banks and Short Forms. Archives of Physical Medicine and Rehabilitation. 2020;101(1):20–32. doi:10.1016/j.apmr.2019.07.015 [PubMed: 31473208]

21. Kisala PA, Tulsky DS. Opportunities for CAT applications in medical rehabilitation: development of targeted item banks. J Appl Meas. 2010;11(3):315–330. [PubMed: 20847478]

22. PROMIS Health Organization. PROMIS Instrument Development and Psychometric Evaluation Scientific Standards, Version 2.0; 2013. http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf

23. Revicki D, Cella D. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. Quality of Life Research. 1997;6(6):595–600. [PubMed: 9330558]

24. Baylor C, Hula W, Donovan NJ, Doyle PJ, Kendall D, Yorkston K. An introduction to item response theory and Rasch models for speech-language pathologists. Am J Speech Lang Pathol. 2011;20(3):243–259. doi:10.1044/1058-0360(2011/10-0079) [PubMed: 21622595]

25. Miller CW, Baylor CR, Birch K, Yorkston KM. Exploring the Relevance of Items in the Communicative Participation Item Bank (CPIB) for Individuals With Hearing Loss. Am J Audiol. 2017;26(1):27–37. doi:10.1044/2016_AJA-16-0047 [PubMed: 28114665]

26. Cella D, Gershon R, Lai J-S, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res. 2007;16 Suppl 1:133–141. doi:10.1007/s11136-007-9204-6 [PubMed: 17401637]

27. Fergadiotis G, Casilio M, Hula WD. Fergadiotis G, Casilio M, & Hula DW (in press). Computer adaptive testing for the assessment of anomia severity. Seminars in Speech and Language.

28. HealthMeasures. Administration Platforms. Published October 10, 2020. https://www.healthmeasures.net/

29. Rose TA, Worrall LE, Hickson LM, Hoffmann TC. Guiding principles for printed education materials: design preferences of people with aphasia. Int J Speech Lang Pathol. 2012;14(1):11–23. doi:10.3109/17549507.2011.631583 [PubMed: 22136650]

30. Rose T, Worrall L, Hickson L, Hoffmann T. Do people with aphasia want written stroke and aphasia information? A verbal survey exploring preferences for when and how to provide stroke and aphasia information. Top Stroke Rehabil. 2010;17(2):79–98. doi:10.1310/tsr1702-79 [PubMed: 20542851]

31. Rose TA, Worrall LE, Hickson LM, Hoffmann TC. Aphasia friendly written health information: Content and design characteristics. International Journal of Speech-Language Pathology. 2011;13(4):335–347. doi:10.3109/17549507.2011.560396 [PubMed: 21682542]

32. Tucker FM, Edwards DF, Mathews LK, Baum CM, Connor LT. Modifying health outcome measures for people with aphasia. Am J Occup Ther. 2012;66(1):42–50. doi:10.5014/ajot.2012.001255 [PubMed: 22251830]

33. Kisala PA, Boulton AJ, Cohen ML, et al. Interviewer- versus self-administration of PROMIS® measures for adults with traumatic injury. Health Psychology. 2019;38(5):435–444. doi:10.1037/hea0000685 [PubMed: 31045427]

34. Rutherford C, Costa D, Mercieca-Bebber R, Rice H, Gabb L, King M. Mode of administration does not cause bias in patient-reported outcome results: a meta-analysis. Qual Life Res. 2016;25(3):559–574. doi:10.1007/s11136-015-1110-8 [PubMed: 26334842]

35. Brennan A, Worrall L, McKenna K. The relationship between specific features of aphasia-friendly written material and comprehension of written material for people with aphasia: An exploratory study. aphasiology Aphasiology. 2005;19(8):693–711.

36. Kagan A Supported conversation for adults with aphasia: methods and resources for training conversation partners. Aphasiology Aphasiology. 1998;12(9):816–830.

37. Purdy M Multimodal Communication Training in Aphasia: A Pilot Study. Published online 2014:14.

38. Baylor C, Oelke M, Bamer A, et al. Validating the Communicative Participation Item Bank (CPIB) for use with people with aphasia: an analysis of differential item function (DIF). Aphasiology. 2017;31(8):861–878. doi:10.1080/02687038.2016.1225274 [PubMed: 30918415]

39. Hunting Pompon R, Amtmann D, Bombardier C, Kendall D. Modifying and Validating a Measure of Chronic Stress for People With Aphasia. Journal of Speech Language and Hearing Research. 2018;61(12):2934. doi:10.1044/2018_JSLHR-L-18-0173

40. HealthMeasures. Obtain and Administer Measures. https://www.healthmeasures.net/explore-measurement-systems/promis/obtain-administer-measures

41. Cohen ML, Boulton AJ, Lanzi A, Sutherland E, Hunting Pompon R. Psycholinguistic Features, Design Attributes, and Respondent-Reported Cognition Predict Response Time to Patient-Reported Outcome Measure Items. Quality of Life Research. Published online 2021.

42. Cella D, Choi S, Garcia S, et al. Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. Qual Life Res. 2014;23(10):2651–2661. doi:10.1007/s11136-014-0732-6 [PubMed: 24938431]

43. Cook KF, Victorson DE, Cella D, Schalet BD, Miller D. Creating meaningful cut-scores for Neuro-QOL measures of fatigue, physical functioning, and sleep disturbance using standard setting with patients and providers. Qual Life Res. 2015;24(3):575–589. doi:10.1007/s11136-014-0790-9 [PubMed: 25148759]

44. Cook KF, Cella D, Reeve BB. PRO-Bookmarking to Estimate Clinical Thresholds for Patient-reported Symptoms and Function: Medical Care. 2019;57:S13–S17. doi:10.1097/MLR.0000000000001087 [PubMed: 30985591]

45. Rothrock NE, Cook KF, O'Connor M, Cella D, Smith AW, Yount SE. 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. Qual Life Res. Published online August 13, 2019. doi:10.1007/s11136-019-02261-2

46. Mann CM. Identifying clinically meaningful severity categories for PROMIS pediatric measures of anxiety, mobility, fatigue, and depressive symptoms in juvenile idiopathic arthritis and childhood-

onset systemic lupus erythematosus. Quality of Life Research. Published online 2020:12. doi:10.1007/s11136-020-02513-6

47. Nagaraja V, Mara C, Khanna PP, et al. Establishing clinical severity for PROMIS® measures in adult patients with rheumatic diseases. Qual Life Res. 2018;27(3):755–764. doi:10.1007/s11136-017-1709-z [PubMed: 28983738]

48. Kisala PA, Victorson D, Nandakumar R, et al. 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. Published online November 2020:S000399932031251X. doi:10.1016/j.apmr.2020.08.026

49. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Annals of the Rheumatic Diseases. 2007;66(Supplement 3):iii40–iii41. doi:10.1136/ard.2007.079798 [PubMed: 17934093]

50. Lehman LA, Velozo CA. Ability to Detect Change in Patient Function: Responsiveness Designs and Methods of Calculation. Journal of Hand Therapy. 2010;23(4):361–371. doi:10.1016/j.jht.2010.05.003 [PubMed: 20638823]

51. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Third. F.A. Davis Company; 2015.

52. Kozlowski AJ, Cella D, Nitsch KP, Heinemann AW. Evaluating Individual Change With the Quality of Life in Neurological Disorders (Neuro-QoL) Short Forms. Arch Phys Med Rehabil. 2016;97(4):650–654.e8. doi:10.1016/j.apmr.2015.12.010 [PubMed: 26740062]

53. Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui K-K. Evaluating the Statistical Significance of Health-Related Quality-Of-Life Change in Individual Patients. Eval Health Prof 2005;28(2):160–171. doi:10.1177/0163278705275339 [PubMed: 15851771]

54. Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. Cochrane Database Syst Rev. 2016;(6):CD000425. doi:10.1002/14651858.CD000425.pub4 [PubMed: 27245310]

55. das Nair R, Cogger H, Worthington E, Lincoln NB. Cognitive rehabilitation for memory deficits after stroke. Cochrane Stroke Group, ed. Cochrane Database of Systematic Reviews. Published online September 1, 2016. doi:10.1002/14651858.CD002293.pub3

56. Lin K, Hsieh Y, Wu C, Chen C, Jang Y, Liu J. Minimal Detectable Change and Clinically Important Difference of the Wolf Motor Function Test in Stroke Patients. Neurorehabil Neural Repair. 2009;23(5):429–434. doi:10.1177/1545968308331144 [PubMed: 19289487]

57. Wu C, Chuang L, Lin K, Lee S, Hong W. Responsiveness, Minimal Detectable Change, and Minimal Clinically Important Difference of the Nottingham Extended Activities of Daily Living Scale in Patients With Improved Performance After Stroke Rehabilitation. Archives of Physical Medicine and Rehabilitation. 2011;92(8):1281–1287. doi:10.1016/j.apmr.2011.03.008 [PubMed: 21807147]

58. Chen C, Shen I, Chen C, Wu C, Liu W-Y, Chung C. Validity, responsiveness, minimal detectable change, and minimal clinically important change of Pediatric Balance Scale in children with cerebral palsy. Research in Developmental Disabilities. 2013;34(3):916–922. doi:10.1016/j.ridd.2012.11.006 [PubMed: 23291508]

59. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Review of Pharmacoeconomics & Outcomes Research. 2011;11(2):171–184. doi:10.1586/erp.11.9 [PubMed: 21476819]

60. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102–109. doi:10.1016/j.jclinepi.2007.03.012 [PubMed: 18177782]

61. Baumhauer JF. Patient-Reported Outcomes — Are They Living Up to Their Potential? N Engl J Med. 2017;377(1):6–9. doi:10.1056/NEJMp1702978 [PubMed: 28679102]

62. Cook KF, Kallen MA, Coon CD, Victorson D, Miller DM. Idio Scale Judgment: evaluation of a new method for estimating responder thresholds. Quality of Life Research. 2017;26(11):2961–2971. doi:10.1007/s11136-017-1625-2 [PubMed: 28624901]

63. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. Journal of Clinical Epidemiology. 2010;63(7):760–766.e1. doi:10.1016/j.jclinepi.2009.09.009 [PubMed: 20056385]

64. Schwartz CE, Powell VE, Rapkin BD. When global rating of change contradicts observed change: Examining appraisal processes underlying paradoxical responses over time. Qual Life Res. 2017;26(4):847–857. doi:10.1007/s11136-016-1414-3 [PubMed: 27722830]

65. Zeppieri G, George SZ. Patient-defined desired outcome, success criteria, and expectation in outpatient physical therapy: a longitudinal assessment. Health Qual Life Outcomes. 2017;15(1):29. doi:10.1186/s12955-017-0604-1 [PubMed: 28143546]

66. Krasny-Pacini A, Evans J, Sohlberg MM, Chevignard M. Proposed Criteria for Appraising Goal Attainment Scales Used as Outcome Measures in Rehabilitation Research. Archives of Physical Medicine and Rehabilitation. 2016;97(1):157–170. doi:10.1016/j.apmr.2015.08.424 [PubMed: 26343173]

67. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012 [PubMed: 27330520]

68. Embretson S, Reise S. Item Response Theory for Psychologists. Lawrence Earlbaum Associates; 2000.

69. Lapin B, Thompson NR, Schuster A, Katzan IL. Clinical Utility of Patient-Reported Outcome Measurement Information System Domain Scales: Thresholds for Determining Important Change After Stroke. Circ Cardiovasc Qual Outcomes. 2019;12(1). doi:10.1161/CIRCOUTCOMES.118.004753

70. Haley SM, Fragala-Pinkham MA. Interpreting Change Scores of Tests and Measures Used in Physical Therapy. Phys Ther. 2006;86(5):735–743. doi:10.1093/ptj/86.5.735 [PubMed: 16649896]

71. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health Qual Life Outcomes. 2004;2:16. doi:10.1186/1477-7525-2-16 [PubMed: 15038830]

72. Jobe JB. Cognitive psychology and self-reports: models and methods. Qual Life Res. 2003;12(3):219–227. [PubMed: 12769134]

73. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. Health Qual Life Outcomes. 2004;2:14. doi:10.1186/1477-7525-2-14 [PubMed: 15023229]

74. Tourangeau R, Rips LJ, Rasinski KA. The Psychology of Survey Response. Cambridge University Press; 2000.

75. Aaronson NK, Elliot TR, Greenhalgh J, et al. User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice, Version 2. International Society for Quality of Life Research; 2015. http://www.isoqol.org/UserFiles/file/UsersGuide.pdf

76. Ayers DC, Zheng H, Franklin PD. Integrating patient-reported outcomes into orthopaedic clinical practice: proof of concept from FORCE-TJR. Clin Orthop Relat Res. 2013;471(11):3419–3425. doi:10.1007/s11999-013-3143-z [PubMed: 23925525]

77. Bingham CO, Bartlett SJ, Merkel PA, et al. Using patient-reported outcomes and PROMIS in research and clinical applications: experiences from the PCORI pilot projects. Qual Life Res. 2016;25(8):2109–2116. doi:10.1007/s11136-016-1246-1 [PubMed: 26914103]

78. Chan EKH, Edwards TC, Haywood K, Mikles SP, Newton L. Implementing patient-reported outcome measures in clinical practice: a companion guide to the ISOQOL user's guide. Quality of Life Research. 2019;28(3):621–627. doi:10.1007/s11136-018-2048-4 [PubMed: 30448911]

79. Hsiao C-J, Dymek C, Kim B, Russell B. Advancing the use of patient-reported outcomes in practice: understanding challenges, opportunities, and the potential of health information technology. Quality of Life Research. Published online January 25, 2019. doi:10.1007/s11136-019-02112-0

80. Santana MJ, Haverman L, Absolom K, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. Qual Life Res. 2015;24(7):1707–1718. doi:10.1007/s11136-014-0903-5 [PubMed: 25589231]

81. Wagner LI, Schink J, Bass M, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer. 2015;121(6):927–934. doi:10.1002/cncr.29104 [PubMed: 25376427]

82. Malec JF. Goal Attainment Scaling in Rehabilitation. Neuropsychological Rehabilitation. 1999;9(3-4):253–275. doi:10.1080/096020199389365

Cognitive Skills

Please respond to each question or statement by holding up the number of fingers that corresponds to your answer.

How much DIFFICULTY do you currently have…

managing your appointments?				
None	A little	Somewhat	A lot	Cannot do

Figure 1.
Example of a Non-verbal Response System
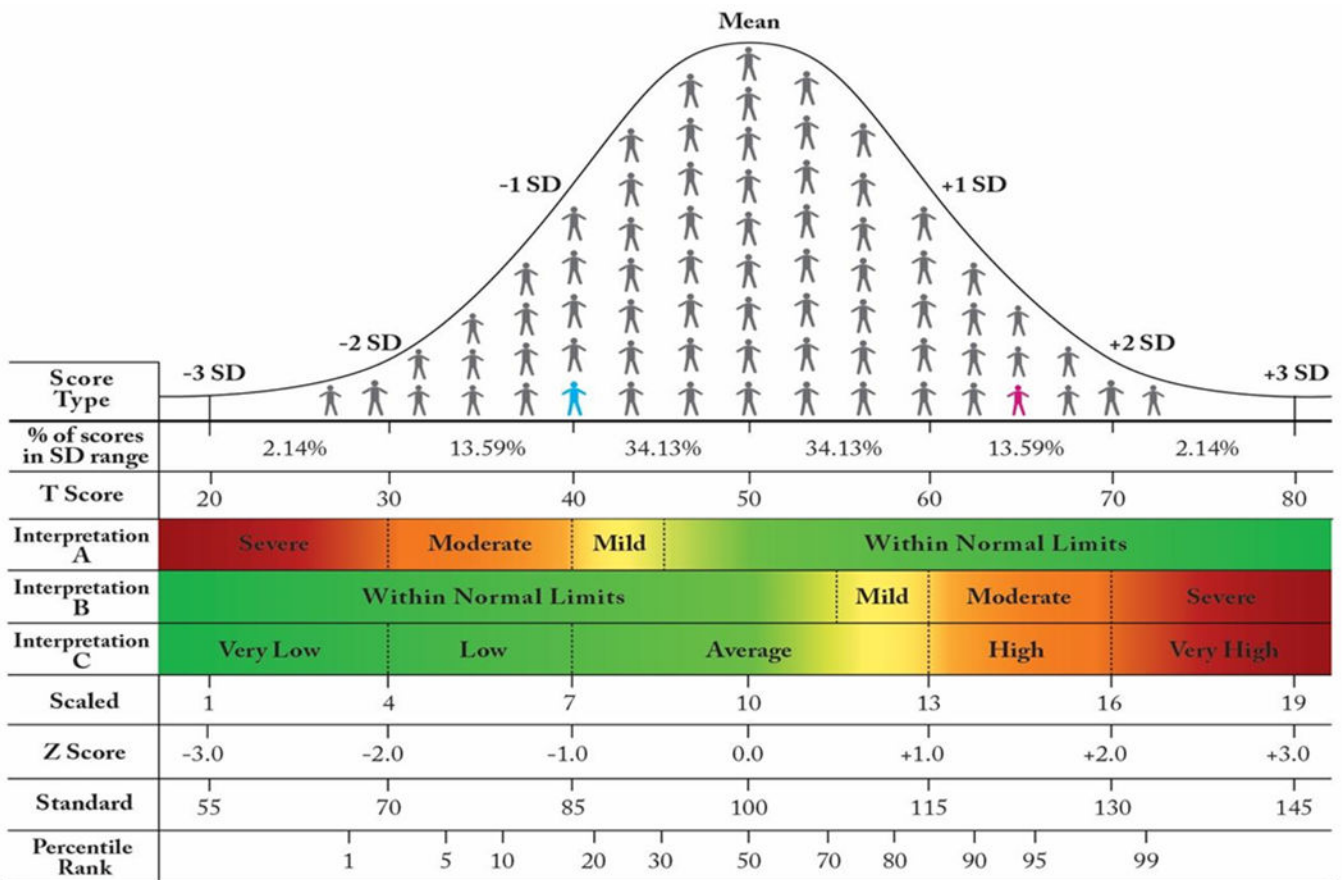
Figure 2.
A Normal Distribution, Standardized Scores, and SD-Based PROM Score Interpretation
The person in blue is −1.0 SD below the mean. The person in pink is 1.5 SD above the
mean. PROM Interpretations A and B are mirror images of one another.
Abbreviation: SD, standard deviation.

Page 25

Figure 3.

MDC$_{95}$ Thresholds for Neuro-QoL v2.0 – Cognitive Function

Reprinted from Kozlowski et al.[52] with permission from the copyright holder. Other tables and MDC calculators available as supplementary materials to that paper.

Table 1.

Most Likely Responses to Items[a] on the CPIB Generic Short Form

CPIB Short Form Item[b]	T-score								
	25	30	35	40	45	50	55	60	65
…Talking with people you know?	1	2	2	3	3	3	4	4	4
…Communicating when you are out in your community (e.g., errands; appointments)?	1	1	2	2	2	3	3	4	4
…Asking questions in a conversation?	1	1	2	2	3	3	3	4	4
…Communicating in a small group of people?	1	1	2	2	3	3	3	4	4
…Trying to persuade a friend or family member to see a different point of view?	1	1	1	2	2	3	3	4	4
…Talking with people you do NOT know?	1	1	1	2	2	3	3	4	4
…Getting your turn in a fast-moving conversation?	1	1	1	1	2	2	3	3	4
…Giving someone DETAILED information?	1	1	1	2	2	3	3	3	4
…Communicating when you need to say something quickly?	1	1	1	1	2	2	3	3	4
…Having a long conversation with someone you know about a book, movie, show or sports event?	1	1	1	2	2	3	3	4	4
Vignettes Names[c]	Ms. Rodriguez	Mr. Bell	Ms. Williams	Mr. Brown	Ms. Garcia	Mr. Davis	Ms. Ward	Mr. Nelson	Ms. Miller

[a] Response options: 1 = very much; 2 = quite a bit; 3 = a little; 4 = not at all.

[b] The item stem was "does your condition interfere with…"

[c] Vignette names at the bottom of the table indicate that vignettes were written about hypothetical people at those T-score levels. However, the vignettes included items and responses from all 46 items of the CPIB, not exclusively items that are on the generic short form.

Abbreviation: CPIB, Communication Participation Item Bank.