Guidance on Modifications to HealthMeasures Items

HealthMeasures users often ask if they can make modifications to measures in English*. Sometimes this is fine. Other times we think it should be studied, and sometimes we discourage the modification. This document outlines our approach to modifications.

These changes to the respondent interface (e.g., on a computer, smart phone, paper version) are acceptable and do not require seeking permission. These interface changes are negligible. The item is still considered a valid HealthMeasure item and standard scoring can be applied.

- Removing the item IDs, response scores, and/or measure name
- Altering the measure name on a respondent-facing interface (e.g., changing "Depression" to "Mood").
- Removing the domain header (e.g., "Fatigue" from a PROMIS Profile)
- Altering the order of items in a fixed length short form or profile
- Including non-HealthMeasures items in an assessment that includes a HealthMeasure (e.g., administering Neuro-QoL Emotional and Behavioral Dyscontrol CAT plus two new items about impulsivity).
    - Note, the new items are NOT included in the HealthMeasure score.
- Presenting self-report items in a grid-like format with a scroll bar on a web-based interface (versus presenting items one at a time on a screen as done in a computer adaptive test and the NIH Toolbox and PROMIS iPad apps)
- Including the context in an instruction rather than alone (e.g., "Please respond to the questions below thinking about how you have been feeling in the past 7 days.") This example requires that the instruction is always visible when responding to the associated items. This would most likely be a grid-like format with multiple items on a screen.
- Underlining, italicizing, or bolding text (e.g., "In the past 7 days", "How would you rate your pain on **average**?") or removing existing underlining, italics, or bolded text. We don't encourage this, but if there is reasonable argument (e.g., administration platform does not support underlining in its user interface), we can accept it.

Significant Modifications

Modifications beyond what is described above are considered significant and should be considered carefully.

Caution

- A modified HealthMeasure item cannot be treated as a valid PROMIS/Neuro-QoL/ASCQ-Me/NIH Toolbox item unless rather extensive research demonstrates its psychometric equivalence to the original item.
- We do not recommend using the HealthMeasures Scoring Service or item parameters that are applied to the original (source) items unless and until data are provided to demonstrate the validity of doing so.
- Modified items require all necessary licenses and permission to translate.

Identify How the Modified Item will be Used

Outside of the modifications described above, the context of use influences our approach to change requests. In some instances, a researcher or clinician wants to make a modification to a HealthMeasure that they are using within their own study or clinic. There is no intention of disseminating the measure. In other instances, a researcher or clinician wants to make a modification to improve a HealthMeasure with the aim of publishing it as a HealthMeasure and/or having that modification adopted into HealthMeasures. Although our philosophy is the same in both contexts, we have stricter guidance for measures to be disseminated. Additionally, we have significant restrictions for distributors of measures such as electronic health record vendors and data capture applications. Contact Help@HealthMeasures.net for more information if you are a distributor.

Approach to Other Change Requests

Our philosophy is that measures can be improved and we welcome others to make improvements. We believe measurement problems or measure improvement are **based upon quantitative and/or qualitative data, not expert**

*If you are interested in modifying a HealthMeasure in a language other than English, please contact Help@HealthMeasures.net.

opinion alone. If a user feels that they can improve a HealthMeasure, they are welcome and encouraged **to test this modification**. Ideally, the problem is identified with data (e.g., respondent misinterpretation of an item, differential item functioning [DIF], insufficient content validity) and not expert opinion alone (e.g., "I think the recall period should be 3 days").

If you are modifying a measure for your own research or clinical practice, proceed with caution. From time to time, people believe they can improve upon an item as applied to their testing. Often, such attempts do not accomplish the intended improvement. Many modifications have already been tested and found to perform poorly. However, in most cases we do not have data to know how that modification will alter how the item performs. For this reason, we strongly encourage testing modifications before using widely. We do not prevent researchers or clinicians from making modifications within their own work. We do require modifications be described when presenting or publishing results (see Publication Checklist). If you would like our opinion on your proposed modification, we are happy to provide that.

If you are modifying a measure that you intend to publish as a HealthMeasure and/or have your measure adopted into HealthMeasures, you should request permission from help@HealthMeasures.net at the outset of your research. You will be required to provide data to support the modification. There are some modifications that HealthMeasures is unlikely to accept and some that would be welcome as they fill a measurement gap.

Change Requests that are Unlikely to be Adopted by HealthMeasures

Although we support evaluation of modifications, the HealthMeasures scientific teams have invested significant resources in measure development and have made decisions on some aspects of measurement that are not likely to be revised. If you are making a modification with the aim that it will be adopted by HealthMeasures, be aware that in some instances this is unlikely to be successful. Specifically:

- **T-score Metric**: PROMIS and Neuro-QoL utilize a T-score metric. Although supplemental metrics (e.g., percentiles) may be of interest, *alternative* metrics that would replace the T-score metric are discouraged.
- **Direction of Scores**: PROMIS and Neuro-QoL are scored so that high = more of the concept being measured. Sometimes this is desirable (e.g., physical function) and sometimes it is undesirable (e.g., fatigue). This decision was made through a deliberate consensus process. Requests to permanently alter the direction of scoring are likely to be declined.
    - However, users are welcome (and sometimes encouraged) to alter the display of scores to facilitate interpretability. For example, utilize two y-axes so that "up" can mean "good" for all measures. However measures are displayed, a viewer should be able to identify the final T-score for the measure as well as the direction of the T-score.
- **Item-Level Calibrations**: We strongly encourage the calculation of population or country-specific norms, rather than recalibrating measures for a specific population or language. Having a common metric across samples and languages is one of the unique benefits of using HealthMeasures instruments. This is made possible by use of a single set of calibrations, i.e., discrimination and location parameters. Given evidence of differential item functioning (DIF) by language or disease population, however, we support the adoption of country- or population-specific calibrations for those particular items showing non-trivial DIF (provided these local item calibrations are then rescaled to the PROMIS T-score mean and SD). For example of DIF analysis by language and parameter rescaling, please see Hays et al., 2017, *Quality of Life Research*. Calculation and reporting of population- or country-specific *norms* or *percentiles* is valuable and supported (e.g., the mean in the general population of a non-U.S. country may not be T=50). We further support research into measurement challenges (e.g., analyzing for DIF) and also support research modifications (e.g., removing an item) to improve a measure's performance.
- **Creating Disease or Condition-specific Measures:** PROMIS, Neuro-QoL, and NIH Toolbox developed measures to be appropriate across health conditions. This was due in part to the need to improve measurement in individuals with multiple chronic conditions. Sometimes a user requests to modify the respondent's instructions to answer thinking only of one condition (e.g., "In answering the items below, think about the impact of your **ARTHRITIC KNEE** on your physical function." This modification is really changing the nature of the question. In fact, it would be reasonable to consider if the original scoring can be used with such a modification or if

utilization of raw scores or recalibration would be more appropriate. In addition, research across several studies and multiple disease areas have cast doubt on whether respondents can make these differential attributions. Therefore, these kind of modifications will likely not be adopted. Adopting such modification would require evidence that it improves the measure. We recommend, for example, administering the item in its original and modified version separated by other questions or, conduct a randomized trial of one way versus the other. Then, analyze the results to see if there is a difference. It should also be studied how this modification impacts the item parameters.

HealthMeasures supports alternative approaches that retain item syntax and current calibrations. For example, researchers have had patients and clinicians review items from a given item bank to select items most relevant for a condition or to identify items that may not be appropriate for use in a given condition (e.g., Cook KF et al 2011 *QLR*). After this review, others have then also constructed new items to reflect domain content not included in the original measure (ideally co-calibrated with the existing measure, e.g., Schifferdecker KE et al 2018 *QLR*). Presentations and publications should include the item IDs and source item bank (e.g., PROMIS Depression Item Bank v1.0) for custom short forms to facilitate communication.

- **Recall period:** Some HealthMeasures utilize a recall period (e.g., "In the past 7 days…"). There are instances when a user would like to alter the recall period to be longer or shorter. We do not have data to evaluate if and how this alters measurement properties. This is a modification that should be tested if utilized in a research study or clinical practice. It is unlikely that an altered timeframe would replace an existing HealthMeasure. There may, however, be sufficient evidence for the addition of a new measure (e.g., PROMIS daily fatigue short form).

Modifications that are Likely to be Adopted

In general, modifications to items that are likely to be adopted by HealthMeasures 1) address a measurement gap or problem and 2) were tested using sound methodology. Researchers are encouraged to contact help@HealthMeasures.net at the outset of research if they aim to have modifications adopted by PROMIS, Neuro-QoL, ASCQ-Me, or NIH Toolbox.

Examples of Modifications Likely to be Adopted:

1. PROMIS does not include measures for adults' caregivers. A researcher wants to administer PROMIS Fatigue to caregivers to assess fatigue in people with dementia. The researcher found that many respondents asked questions about how to respond to the measure. This was also found during Neuro-QoL testing of measures with caregivers. Consequently, the researcher asked help@HealthMeasures.net if she could change the items from "How often did you…" to "How often did he or she…". Ideally, when proxy report measures are developed, one compares the proxy report to the self-report for the same sample. In this case, the researcher argued that individuals with dementia would not be able to provide reliable self-report data. The researcher proposed to evaluate the psychometric properties of the altered items. The researcher agreed to describe the modifications in any publications or presentations on data from the measure. These modifications are positioned to be adopted by PROMIS because the researcher a) successfully presented a measurement gap, b) articulated an approach to evaluate the modification, and c) agreed to describe that modification in publications.

2. A researcher wants to modify PROMIS Bowel Incontinence items for people with a stoma. Specifically, as people with a stoma have a pouch attached to their abdomen that collects stool, they do not experience soiling underwear in a bowel accident. The researcher proposes to modify the syntax of items that describe activities that are not done by people with a stoma. The researcher will conduct cognitive interviews on the revised items with patients with stoma and then translate those items into other languages. For example, a modification may be to change "soil my underwear" to "soil my clothing" as a stoma may soil clothing that is not underwear. These modifications may be able to be adopted by PROMIS because the researcher a) successfully presented a measurement gap, b) articulated an approach to evaluate the modification that included evaluation of the impact on item parameters, and c) agreed to describe that modification in publications. Note, however, that PROMIS has a philosophy of having measures that function well for people across chronic conditions. If the researcher proposed adding a modified item that is stoma-specific (e.g., "I am bothered by having accidents from my stoma"), PROMIS is not likely to adopt that modification in its measure or create a stoma-specific

version of its measure. Modifications that retain universality (e.g., "soil my underwear" modified to "soil my clothing") stand a better chance of adoption. Another acceptable approach is to develop universal items that represent content that is important to a specific population (e.g., patients with arthritis) and co-calibrate those items with the existing item bank. This modified bank would be an improvement for the specific population but able to be used across all populations.

Modifications for Translations

When translating measures developed in English into other languages, it is sometimes found that a modification to the English item would facilitate translations (e.g., inclusion of a metric equivalent). All modifications related to translations are to be directed to the HealthMeasures Director of Translations. The Director of Translations will decide on modifications in all cases. Contact translations@HealthMeasures.net for more information.

NIH Toolbox Language about Modifications in Terms of Use

1. Specific modification plans must be enumerated and submitted *in writing* to Provider (Help@HealthMeasures.net) for approval. This request must include the specific changes to be made and the rationale for the changes. This written request must be signed by the lead researcher.
2. User may not make modifications to any component of NIH Toolbox without written consent of Provider.
3. All modifications should be fully validated against the existing NIH Toolbox measure(s) on which they were based. Plans for a concurrent validation study should be included in the written request for modifications. If no validation study is planned, this must be noted in the written request, along with an explanation.
4. User must send validation study results to Provider for review prior to publication citing any results of said study or any results citing use of NIH Toolbox.
5. Provider will review validation study data and will inform User if modification to NIH Toolbox measure(s) constitute an "Approved Adaptation" or a "Non-Validated Adaptation" of NIH Toolbox. Provider will update HealthMeasures.net accordingly to inform other users and prospective users of these modifications, along with contact information for the lead researcher.
6. User must cite NIH Toolbox in any and all presentations, publications, or other third-party sharing of research data, indicating whether any measures used constitute Approved Adaptations or Non-Validated Adaptations.
7. **Provider maintains all rights to Approved Adaptations as well as Non-Validated Adaptations of NIH Toolbox tests and test items made by User. User may under no circumstances license, distribute, or share any components of NIH Toolbox with third parties, regardless of the extent of modifications, without official, written consent of PROVIDER. User is entitled to no ownership claim of intellectual property as a result of making any NIH Toolbox modifications, and is hereby enjoined from communicating or stating any such claims**.

Procedure for Review of Modifications

The HealthMeasures team at Northwestern University is the current body that reviews and approves modifications to HealthMeasures based on standards developed by a) the PROMIS Health Organization (PHO) Standards Committee and approved by the PHO board; b) the Neuro-QoL and ASCQ-Me scientific teams; and 3) the NIH Toolbox scientific team. Each entity will generate specific requirements for modifications such as:

- What are the requirements for testing adaptations?
- What evidence do researchers need to collect?
- When do adaptations need to be reviewed?
- When will a modification be accepted by PHO?
- Will a modified measure still be called a PROMIS measure?
- How is "non-trivial DIF" defined?