

PROMIS[®] Instrument Maturity Model

Approved: April 11, 2012;

Revised 02/13, 04/13, 05/13

The **Instrument Maturity Model** describes the stages of instrument scientific development from conceptualization through evidence of psychometric properties in multiple diverse populations. The model is used in conjunction with the standards and guidance documents (<http://www.nihpromis.org/science/publications?AspxAutoDetectCookieSupport=1>) to assist developers in meeting the progressive scientific standard criteria from item pool or scale development to fully validated instruments ready for use in clinical research and practice.

Brief descriptions of each stage follows:

Stage 1: Developmental – Conceptualization & Item Pool Development

The latent trait or domain is conceptualized and defined according to the PROMIS domain framework. Literature reviews and qualitative methods (e.g., individual interviews and/or focus groups) have been used to conceptualize and define the domain. During this phase, attention to literacy, translatability, cultural and lifespan harmonization, and PROMIS guidelines for item construction is required. At the end of this phase, an item pool or scale will have been developed.

Stage 2: Developmental – Calibration Phase

The items have undergone calibration following psychometric analyses using “best practices” factor analysis and item response theory methods or methods appropriate for a different measurement model. In addition, limited information relating the item bank’s measurement properties to existing “legacy” instruments of the domain (concurrent validity) has been assessed. Some modifications to the item pool based on both the qualitative (e.g., cognitive testing or debriefing) and psychometric analyses have been completed. Information has been developed on measurement error across the domain. Instruments such as short forms or CATs have been assessed and defined. Differential item functioning (DIF) is assessed with respect to a minimal set of relevant demographic and language variables (e.g., age, gender, and race/ethnicity), and recommendations made concerning the potential impact of DIF on the use of the item bank and scores. Not all measures will be computer adaptive assessments based on item banks. At times, static forms are desirable or even more appropriate. For example, standardized, static health profile instruments can capture multi-dimensional health concepts across several item banks. Stage 2 instruments may be appropriate for use as outcome measures in selected research.

Stage 3: Public Release – Calibrated and Preliminary Validation Completed

The measurement properties, validity and reliability of the item bank and related instruments have been more fully assessed and meet the standards for release for public use. A Stage 3 bank meets the same criteria as a Stage 2A or 2B bank for the first eight rows of the Maturity Model. A Stage 3A bank has undergone additional prospective validity and reliability testing than that completed in prior levels. This work may be focused on comparison to an expanded set of legacy measures, which may include a specific clinical population or populations using cross-sectional studies to assess construct validity. The relevance of item content is also further supported in a Stage 3A bank. Stage 3B banks expand the evidence base as relevant to different audiences and applications. For example, they include longitudinal studies to assess responsiveness, mode studies, evaluation of translation into an alternative language and provide some interpretation guidelines in either a general or a clinical population, or both. Targeted data collection facilitates further evaluations for DIF with respect to other covariates beyond those assessed in Stage 2, which now may include education level, socioeconomic status, language translations etc. These item banks and related instruments may be appropriate as clinical research outcomes.

Stage 4: Maturing - Responsiveness and Expansion

These instruments benefit from continued expansion of the development and evaluation begun in Stage 3B. They have undergone continued reliability, validity and responsiveness testing in different clinical populations. DIF analyses have been expanded to include additional relevant known groups which may include socioeconomic status (SES), language translation(s), and literacy levels. These are considered more mature instruments. Based on their measurement characteristics (responsiveness, MIDs, etc.), use within clinical settings (e.g., to measure individual change) may be appropriate. In addition, the underlying item banks may be in the process of being iteratively improved.

Stage 5: Fully Mature User Support

These instruments have undergone very extensive reliability, validity and responsiveness testing across multiple clinical populations. Score interpretations (absolute level or change) have been developed and are used to understand the health of patients and to guide decision making and follow-up actions. These interpretations may emerge as the result of a history of widespread use of the instruments across populations and applications; or, they can be fostered by the developers of the measures who create a user-friendly administration, scoring and interpretation manual or course geared to different audiences for different uses of the measures. The highly mature measure has been widely adopted and used as evidenced by searches in data bases such as PubMed and ClinicalTrials.gov. These measures have received recognition or endorsement by a formal review process (e.g. COSMIN criteria; Medical Outcomes Trust criteria; FDA qualification, EMA labeling claim review, NQF endorsement, inclusion in DSM, etc.).

	Develop- mental Stage 1A	Develop- mental Stage 1B	Develop- mental Stage 2A	Develop- mental Stage 2B	Public Release in PROMIS/ Assessment Center 3A	Public Release in PROMIS/ Assessment Center 3B	Public Release in PROMIS/ Assessment Center 4	Public Release in PROMIS/ Assessment Center 5
Stage	Item Pool	Prelimi- nary Item Bank	Calibrated Item Bank	Item Bank, Profile or Global Health Measure - Preliminary Reliability/ Validity	Instruments - Validated	Instruments – longitudinal data to for prelim responsiveness – other research to expand use- fulness	Maturing Instruments &/or Item Bank Expansion	Instruments with Fully Mature User Support:
Descriptions	Conceptualized	Ready for Calibration	Dimension ality Assessed & Calibrated	Validity (Construct & Concurrent) – limited	Validity - concurrent & construct validity – cross sectional assessed	Prelim responsiveness	Extensive validity & responsiveness in general and pertinent population samples Item bank modifications - population specific or expansion/ refinement	How scores can be used to understand and respond to health care needs and differences in health is determined & documented
QUALITATIVE: Conceptual documentation and evidence supporting content validity	YES	YES	YES	YES	YES	YES	YES	YES
Dimensionality Specified	NO	YES	YES	YES	YES	YES	YES	YES
Domain Placement Specified (approved)	NO	YES	YES	YES	YES	YES	YES	YES
Item response theory (IRT): Item calibration; information and DIF analyses	NO	NO	YES	YES	YES	YES	YES	YES
Classical test theory (CTT): Evidence supporting dimensionality, reliability and validity (e.g. concurrent validity with legacy)	NO	NO	YES	YES	YES	YES	YES	YES
DIF Preliminary Assessed in Known Groups (e.g. age, race/ethnicity, and gender)	NO	NO	YES	YES	YES	YES	YES	YES

POPULATION: Sample variability reflects variability in construct	NO	NO	YES	YES	YES	YES	YES	YES
FORMAT: CAT and short form measures; Computer, paper forms	NO	NO	YES	YES	YES	YES	YES	YES
Scoring Algorithm Specified	NO	NO	NO	YES	YES	YES	YES	YES
Continued Documentation of Relevance of Item Content and Generalizability as needed	NO	NO	NO	NO	YES	YES	YES	YES
Validity: Concurrent and construct assessed with legacy measures	NO	NO	NO	NO	YES	YES	YES	YES
POPULATION: Expanded DIF analyses relevant population characteristics (e.g. educational status, socioeconomic status etc.)	NO	NO	NO	NO	YES	YES	YES	YES
CTT: Evidence supporting responsiveness and interpretation guidelines (MID, responder criteria)	NO	NO	NO	NO	NO	YES	YES	YES
POPULATION: Translation into one language that is spoken by large percentage of population (e.g. in US, Spanish languages.)	NO	NO	NO	NO	NO	YES	YES	YES
POPULATION: Evaluation in general population and multiple disease conditions including DIF analyses by health condition and language translations.	NO	NO	NO	NO	NO	YES	YES	YES
MODE: Evidence supporting multiple modes of administration (CAT, paper, IVRS, computer)	NO	NO	NO	NO	NO	NO	YES	YES
Continued expansion of DIF analyses across subpopulations as well as continued qualitative work on content validity and to generate items at the tails of the distribution.	NO	NO	NO	NO	NO	NO	YES	YES
POPULATION: Translation and psychometric evaluation into languages other than English	NO	NO	NO	NO	NO	NO	NO	YES
Measure is recognized/certified/endorsed /qualified by a recognized consensus review process conducted by NQF or FDA, for example.	NO	NO	NO	NO	NO	NO	NO	YES