



Development and calibration data for the Healthcare Access Item Bank: a new computer adaptive test for persons with type 2 diabetes mellitus

Noelle E. Carlozzi^{1,6} · Michael A. Kallen² · Jonathan P. Troost³ · Alexa Bragg⁴ · Jessica Martin-Howard⁴ · Barbara De La Cruz⁵ · Jennifer A. Miner⁶ · Ioana Moldovan⁵ · Brian W. Jack^{4,5} · Suzanne Mitchell^{4,5}

Accepted: 8 October 2022 / Published online: 31 October 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Purpose The purpose of this study was to develop a new measure to evaluate the ability to receive medical services when needed among persons with type 2 diabetes mellitus.

Methods The Healthcare Access measure was developed using data from 225 persons with type 2 diabetes mellitus who completed an item pool comprised of 54 questions pertaining to their experience accessing healthcare services.

Results Exploratory and confirmatory factor analyses supported the retention of 45 items. In addition, a constrained graded response model (GRM), as well as analyses that examined item misfit and differential item functioning (investigated for age, sex, education, race, and socioeconomic status), supported the retention of 44 items in the final item bank. Expert review and GRM item calibration products were used to inform the selection of a 6-item static short form and to program the Healthcare Access computer adaptive test (CAT). Preliminary data supported the reliability (i.e., internal consistency and test–retest reliability) and validity (i.e., convergent, discriminant, and known-groups) of the new measure.

Conclusions The new Healthcare Access item bank can be used to examine the experiences that persons with type 2 diabetes mellitus have with healthcare access, to better target treatment improvements and mitigate disparities; it will be available as a part of the Neuro-QoL measurement system through healthmeasures.net and the PROMIS Application Programmable Interface (API) in early 2023.

Keywords Diabetes, mellitus, type 2 · Health services · Delivery of healthcare · Patient-reported outcome measures · Outcomes assessment, healthcare

✉ Noelle E. Carlozzi
carlozzi@med.umich.edu
https://coda.med.umich.edu/

- ¹ Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, MI, USA
- ² Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
- ³ Michigan Institute for Clinical and Health Research, University of Michigan, Ann Arbor, MI, USA
- ⁴ Department of Family Medicine, Boston University School of Medicine, Boston, MA, USA
- ⁵ Department of Family Medicine, Boston Medical Center, Boston, MA, USA
- ⁶ Department of Physical Medicine & Rehabilitation, Center for Clinical Outcomes Development and Application (CODA), University of Michigan, North Campus Research Complex, 2800 Plymouth Road, Building NCRC B14, Room G216, Ann Arbor, MI 48109-2800, USA

Introduction

Improving equitable access to healthcare is a priority for many institutions across the United States (US), yet there are many well-documented disparities in treatment and outcomes in the US population. More than 37 million Americans have been diagnosed with diabetes, accounting for over \$327 billion in annual healthcare costs [1, 2]. In addition, patients with type 2 diabetes mellitus (T2DM) experience over 7.7 million hospital admissions per year, and approximately 1 in 5 hospitalizations involve patients with T2DM [3], due to complications driven by unmet health-related social needs, among them an inadequate built environment, economic instability, lack of community/social supports [4], and limited access to medical care. In a recent systematic review, financial constraints and limited access to health services and management were highlighted as two

key healthcare barriers to T2DM care and management [5]. Furthermore, in the US, a number of studies that examine disparities in healthcare access identified disparities in healthcare access for different subgroups of individuals, including those of Hispanic/Latino ethnicity [6–10], Blacks/African Americans [9–11], and women [12]. In many cases, social determinants of health, including education and socioeconomic status, have been associated with these disparities [4, 13, 14].

Healthcare access has been defined by the Institute of Medicine as “the timely use of personal health services to achieve the best health outcomes” [15] and has been broadly thought to encompass healthcare coverage (i.e., insurance), healthcare services (screening, treatment, prevention, and usual care), and timeliness for securing services (the ability to get services once a service need is identified) [16]. Healthcare access refers to the physical availability, affordability, and acceptability of healthcare services [17]. There is currently no consensus approach to evaluating healthcare access, and, as such, current approaches are varied. Many national survey studies employ a series of self-report proxy-type questions to determine health insurance status, whether or not there has been a time in the recent past when the participant was unable to get healthcare that was believed to be needed, and/or emergency room use in the recent past; yet these self-report questions are not standardized and vary across studies (yielding different results) [18]. There are also concerns about whether these types of proxy questions adequately reflect the healthcare access concerns of the patients themselves [19]. Another approach to evaluating healthcare access is to use statistical analyses of healthcare utilization or medication data (using hospital records or medical claims data) alongside demographic variables (e.g., race/ethnicity data, age, insurance data) to determine if rates of utilization differ between different groups, or if outcomes (i.e., morbidity/mortality) differ across groups [20]. This type of approach can yield robust quantitative information, but it still does not directly address patient-specific reports about healthcare access. This problem is also inherent in other analytical approaches to evaluating healthcare access, such as the Geographic Information System Mapping (GIS) and related spatial analytic techniques that create, manage, analyze, and map health services [21] but, again, are devoid of any patient-reported perceptions about healthcare access.

In addition to these commonly used but widely varying approaches, there are also a handful of standardized and validated self-report surveys that examine perceived access to healthcare. For example, the Perceptions of Access to Health Care Services Questionnaire is a 31-item measure that assesses accessibility, affordability, accommodation, acceptability, and awareness. While a total score from the Access to Health Care Services Questionnaire meets established recommendations for internal consistency reliability,

three of the six subscales fall below what is considered acceptable for internal consistency reliability (i.e., Cronbach’s $\alpha < 0.70$) [22]. Other limitations of this measure include the following: 1) its development was informed by experts (i.e., professional providers) not by key stakeholders (i.e., patients); and 2) performance validity, convergent/discriminant validity, known-groups validity, and predictive validity have not been established [22]. Other measures exist but are limited in terms of scope; either they focus on a specific medical condition (e.g., healthcare services for someone with tuberculosis [23], or cancer [24]) or on a single aspect of access (e.g., access to mental health services [25–27], access to physical and built environment for people with disabilities [28], geographic accessibility [29]) or healthcare access for a specific type of hospital system (e.g., the military healthcare system [30], the healthcare system of a particular country) [31, 32]. Regardless of the current approach, there is general consensus that we need to do a better job evaluating healthcare access [18, 20].

To this end, we have developed a new measurement system that is designed to evaluate important social determinants of health that contribute to healthcare disparities. This measurement system, the Re-Engineered Discharge for Diabetes Computer Adaptive Test (REDD-CAT), included the development of a new patient-reported outcome (PRO) item bank to capture the experiences that persons with T2DM have with access to healthcare services, including healthcare coverage, actual provision of healthcare services, and timeliness of receiving services. Specifically, this new PRO was developed as an extension to the Patient-Reported Outcomes Measurement Information System (PROMIS) [33], a measurement system designed to assess concepts that are characteristic to health-related quality of life and appropriate for use in individuals with chronic diseases and conditions. PROMIS and its complementary measures offer several advantages over other systems, given that measurement development has involved both classical test theory and item response theory-based analytical approaches [34, 35]. The combination of these analytical approaches allows for the development of item banks that can be administered as a long form (an administration format where all items in the bank are administered), a static short form (SF; an administration format that includes the administration of a fixed set of items typically selected to capture items of varying levels of endorsement difficulty), or as a computerized adaptive test (CAT; a smart test that employs a standard first item followed by item administration based on the response to a previously administered item [36–42]). These types of measures allow for sensitive assessment of a broad range of symptomatology using a minimum number of items (typically 4–6), without sacrificing precision [43]. In essence, CATs have the advantage of brevity (i.e., minimal administration burden) and tend to be equivalent to or outperform

the psychometric reliability and validity of more traditional static measures, even when the number of items administered across these formats (CATs versus traditional static measures) are identical [44]. Furthermore, PRO measures that were developed using item response theory yield “estimated scores” even if only a single item is administered; threats to score validity because of missing data (that are inherent to classical test theory-based scores requiring the successful completion of all test items to generate an “estimated score”) are considerably less relevant for these types of measures.” [45] We describe the development of this new item bank, the REDD-CAT Healthcare Access Item Bank, below.

Methods

Study participants

This analysis included a total of 225 individuals with T2DM. Data collection efforts were part of a broader effort to develop the REDD-CAT measurement system, a new patient-reported outcomes measure suite designed to capture important social determinants of health, which included the development of a new measure of Healthcare Access (described herein), as well as two additional measures that are also included in this issue of Quality of Life Research (Illness Burden [46] and Medication Adherence [47]). We used three strategies to identify and screen potentially eligible participants at a safety-net urban hospital. Weekly lists of diabetes outpatients with upcoming appointments from Boston Medical Center’s Clinical Data Warehouse were generated. In addition, inpatient census reports were produced via the electronic health record to identify eligible inpatients with T2DM. Participants who were previously enrolled in a T2DM research study and had agreed to be contacted for future research opportunities were also contacted. Study inclusion criteria included the following: aged 18+, diagnosis of T2DM, ability to communicate in English, willingness to participate, and capacity to consent. Given that study participation required, at minimum, a 5th-grade reading level (which is the recommendation for low health literacy precautions)[48], reading comprehension was assessed using the Wide Range Achievement Test 4th Edition (WRAT4) Reading Subtest [49]. Participants who were able to read the first 10 words correctly on the WRAT4 completed the study assessments independently; otherwise, research assistants helped administer the questionnaire and recorded participants’ responses. All study activities were conducted in accordance with the Boston Medical Center/Boston University Medical Campus Institutional Review Board.

Measures

All measures that were used in this study were either publicly available (i.e., Neuro-QoL and PROMIS) or used with permission from the authors (i.e., HEAL measures and Econ-QOL).

Healthcare access

The Healthcare Access item bank was developed according to established PRO development methodology [50]. This new measure was designed to assess participants’ concerns and experiences accessing healthcare. An iterative process was used to refine the Healthcare Access item pool (i.e., an uncalibrated set of items); this included feedback from expert review, item reading-level assessment, translatability review, cognitive interviews, and a final consensus meeting attended by study team investigators. This was followed by a second iterative process, which is described below (under Item Bank Development). Response options included two 5-point Likert scales (1 = never, 2 = rarely, 3 = sometimes, 4 = usually, 5 = always; 1 = not at all, 2 = a little bit, 3 = somewhat, 4 = quite a bit, 5 = very much). The final item bank (i.e., calibrated set of items) is scored on a T-score metric (mean = 50; *SD* = 10) with higher scores indicating more ease in accessing healthcare services. Note that T-scores are normalized relative to the calibrations sample; thus, T-scores for Healthcare Access are normalized relative to a calibration sample of people with T2DM. For reliability and validity analyses, we examined T-scores derived from the full item bank, from computer adaptive testing (CAT; scores were simulated using Firestar Version 1.3.2) [51], and from a 6-item short form.

Mental health measures

Participants completed two measures from the Neuro-QoL measurement system: Neuro-QoL Depression (perceptions of sadness) and Neuro-QoL Anxiety (perceptions of worry, fear, and hyperarousal) [52–54]. Both measures were administered as computer adaptive tests; response options are on a Likert scale (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always). Resulting scores are based on a T-score metric (mean = 50; *SD* = 10) normalized relative to a calibration sample of people with neurological conditions, with higher scores indicating more depression and anxiety, for Depression and Anxiety, respectively. These measures were used to examine the discriminant validity of the new Healthcare Access item bank. Studies in the general population and other diverse clinical samples (i.e., adults with epilepsy, multiple sclerosis, and Huntington disease) support the reliability of the Neuro-QoL Depression and Anxiety item banks [54–60]. Across these studies, reliability is supported by

excellent internal consistency (all Cronbach's alphas ≥ 0.91 regardless of the sample being examined), as well as adequate to excellent test–retest reliability (ICCs range from 0.68 to 0.82 for 7-day test–retest for the epilepsy, Parkinson's disease, and multiple sclerosis cohorts, and 0.95 for 3-day test–retest correlations in the Huntington disease cohort). Convergent validity was supported across all cohorts with moderate to large correlations of these measures with other measures of mental health and or emotional well-being (all correlations ≥ 0.60 across the different measures and different cohorts), and discriminant validity was supported by correlations that were lower in magnitude between Anxiety or Depression and other measures of social health or physical functioning. These measures were also able to differentiate among people with epilepsy with differing levels of disease severity, and people with multiple sclerosis reporting better versus worse global health. There are also data to support responsiveness in Huntington disease and epilepsy.

Substance use/abuse measures

Participants also completed two measures from the Patient-Reported Outcomes Measurement Information System (PROMIS), including PROMIS Alcohol Use [61] (alcohol use; consequences and expectancies of drinking) and PROMIS Severity of Substance Use [62] (past 30 days). Resulting scores are based on a T-score metric (mean = 50; $SD = 10$) normalized relative to a calibration sample of people from the general population, with higher scores indicating more alcohol use and more severe substance abuse, for Alcohol Use and Severity of Substance Use, respectively. These measures were used to examine the discriminant validity of the new Healthcare Access measure. The reliability and validity of these measures have been supported by research in both the general population and in participants of community addiction treatment programs [62–64]. Internal consistency reliability was excellent (all Cronbach's alpha > 0.93) and convergent validity has been supported by a moderate correlation with other alcohol use screening measures (all $r > 0.55$), and responsiveness is supported by a modest improvement in scores.

Healing encounters and attitudes lists (HEAL) measures [65]

Two measures from the HEAL Measurement System were also administered to participants, including the 7-item Patient–Provider Connection SF (trust and satisfaction with one's healthcare provider) and the 6-item Health Care Environment SF (perceptions about staff being respectful and the healthcare environment being comfortable). Responses for these measures are on a 5-point Likert scale (1 = not at all, 2 = a little bit, 3 = somewhat, 4 = quite a bit, 5 = very much). Resulting scores are based on a T-score metric (mean = 50;

$SD = 10$) normalized relative to a calibration sample of people from the general population, with higher scores indicating more positive healing encounters. These measures were used to examine the convergent validity of the new Healthcare Access item bank. Initial development work supported the reliability and validity of the HEAL measures in the general population and for people with chronic conditions [65]. Specifically, internal consistency reliability was excellent for the Patient–Provider Connection and Healthcare Environment measures (Cronbach's alpha = 0.96 and 0.92, respectively) and these measures also demonstrated moderate relationships with a measure of outpatient clinical care (r 's 0.38 and 0.39, respectively), providing preliminary support for validity.

Economic quality of life (Econ-QOL) short form [66]

The 8-item Econ-QOL short form was administered to evaluate perceptions about economic and financial HRQOL. Responses for this measure are on a 5-point Likert scale (1 = never, 2 = rarely, 3 = sometimes, 4 = usually, 5 = always). This measure is scored on a T-score metric ($M = 50$; $SD = 10$), with higher scores reflecting worse economic quality of life. We examined those patients with “better” economic quality of life, i.e., scores one standard deviation below the mean (≤ 40), versus those with “worse” economic quality of life, i.e., scores one standard deviation above the mean (≥ 60). The reliability and validity of the Econ-QOL measure are supported by the measurement development papers in individuals with disabilities (e.g., traumatic brain injury, spinal cord injury, and stroke) [67–69], as well as in caregivers of persons with traumatic brain injury [70]; internal consistency reliability is excellent (all Cronbach's alpha > 0.91), there are moderate correlations with self-reported income (r 's > 0.46), and scores can discriminate between those above versus below (or possibly below) the poverty line.

Hospital readmissions

Medical record data were used to identify the inpatient admissions during the previous 6-month period for each participant. Those individuals with two or more inpatient readmissions in the previous 6 months were considered “high risk” for readmission; individuals with no or one inpatient admission in the previous 6-month period were considered “low risk” for readmission.

Data collection

All self-report data were collected using REDCap [71, 72]. Assessments were completed on either a personal or study-owned mobile device or desktop/laptop computer or

by phone interview with the study staff. Specifically, 200 participants completed the survey using a study device, one completed the study via telephone assisted by a study research assistant and three completed the study on a home device by connecting to a study survey link sent by email. A total of 21 participants used more than one method to complete the survey. Fifteen participants partially completed the survey in person at Boston Medical Center on a study device and finished completing survey questions by phone with staff administering questions verbally (this includes, but is not limited to, participants who received reading assistance from staff due to failing the WRAT4) and recorded participant responses directly in REDCap. Three participants completed the REDD-CAT survey partially in person on a study device and partially at home using study link sent by email. Three additional participants began the survey responses in person on a study device, continued the survey by phone with staff, and finished the survey using the emailed study link.

Statistical analyses

Item bank development: qualitative analyses

The item bank development process was conducted according to the published measurement development standards [73] and is detailed in the Supplementary File 2. This item pool was informed by both the concerns raised in the previous qualitative studies, as well as by a literature review that identified an existing measure that was specific to the military healthcare system: the TBI-CareQOL Military Frustration measure [30]. The final Healthcare Access item pool included items adapted from the TBI-CareQOL Military Frustration measure [30], as well as newly drafted items that reflected the specific concerns that were raised in the above-reported qualitative studies. As noted previously, the item pool was revised using an iterative process that included expert review, cognitive interviews, reading-level assessment, and translatability review (to ensure future acceptability for adaptations into Spanish and other languages).

Item bank development: quantitative analyses

Following the literature review and qualitative analysis of interview data, classical test theory (CTT) and item response theory (IRT) analytic approaches were used to develop the calibrated item bank. A detailed summary of our analytical approaches can be found in the Supplementary File 1. In brief, our quantitative analyses were as follows. We identified a unidimensional set of items using full-sample exploratory and confirmatory factor analyses (EFA, CFA), in conjunction with clinical input [81–83]; these analyses were conducted using Mplus (version 7.4) [84]. For EFA, we considered the item set to have unidimensional characteristics

if the ratio of eigenvalue 1 to eigenvalue 2 was ≥ 4 and the proportion of variance accounted for by eigenvalue 1 was ≥ 0.40 . We excluded items with sparse cells (response categories with $n < 5$ respondents), items with low item-adjusted total score correlations (< 0.40), and items that were non-monotonic (monotonicity was examined using non-parametric IRT models of item-rest plots and expected score by latent trait plots; Testgraf Software [85]). For CFA, we considered an item set to be unidimensional if: the comparative fit index (CFI) was ≥ 0.90 , the Tucker-Lewis index (TLI) was ≥ 0.90 , and the root mean square error of approximation (RMSEA) was < 0.10 [44, 82, 86–90]. For comparative fit purposes, we also obtained the chi-square value for model fit and its associated p value. We deleted items with low factor loadings ($\lambda < 0.50$) and items that were locally dependent (i.e., residual correlation > 0.20 ; correlated error modification index ≥ 100) [81–83; 91–95]. When CFA overall model fit criteria were not fully met, we conducted confirmatory bi-factor analyses (CBFA) [83, 96] to obtain comparators to traditional fit analyses. CBFA can be used to assess whether the data are “unidimensional enough” to fit with a unidimensional measurement model [97]. CBFA provides a set of indices to assess factor strength, including omega, omega-Hierarchical (omega-H), and explained common variance (ECV). For our interests, omega-H provides a dimensionality index: A threshold general factor omega-H value > 0.80 has been recommended for establishing a measure’s essential unidimensionality [98].

Next, a constrained graded response model (GRM), i.e., a common-slope IRT model that is appropriate when sample sizes are less than $N = 500$ [99], was used to estimate item parameters. We excluded items with significant misfit ($S-X^2 / df$ effect size > 3) [100–103]. We also excluded items with impactful differential item functioning (DIF): (1) a statistically significant ($p < 0.01$) group-specific item parameter difference, with a weighted area beneath the curve [wABC] effect size > 0.30 [104], for any DIF candidate item tested; plus (2) $> 2\%$ of DIF-corrected vs. uncorrected score differences exceeding individual case uncorrected score standard errors. DIF analyses were conducted for factors theorized to be potentially biasing, given that $n \approx 100$ participants per DIF factor subgroup were available [105]. This included an examination of DIF for age (< 60 vs. ≥ 60 years), sex (male vs. female), education (\leq high school vs. $>$ high school), and socioeconomic status (pay rent/mortgage: *never/rarely/sometimes* vs. *usually/always*; pay bills on time: *never/rarely/sometimes* vs. *usually/always*). DIF analyses were conducted in IRTPRO (version 3.1.2) [106] using iterative Wald-2 testing, a process that establishes a set of DIF-free items against which candidate DIF items can be examined to determine if they exhibit DIF [107, 108]. In Wald-2 testing Step 1, we identified a DIF-free set of anchor items, while in Step 2, we tested any identified candidate items for DIF.

Subgroup-specific parameters were estimated for each candidate DIF item, using the constrained GRM. The parameters were then compared across subgroups—total DIF, slope-related DIF, and threshold-related DIF—to identify the statistically significant parameter differences with non-trivial effect sizes. The suite of IRT-based analyses was followed by a final CFA analysis designed to confirm that the final item set was essentially unidimensional (using the same item-level and overall model fit criteria outlined above).

Calibration parameters (i.e., slope and threshold estimates) from our GRM analyses were used to program computer adaptive test (CAT) administration of the final item bank. For a more rigorous and realistic CAT performance assessment, we simulated the item responses from $N=2000$ cases drawn from a clinical population (i.e., having a mean one SD in the direction of worse health status). CAT administration parameters (e.g., number of items to administer, targeted score reliability level) were optimized to balance response burden and score precision. In addition, a 6-item short form (SF) was constructed, using clinician input and item-level statistics, including item score-level information values. SF items were purposefully selected to represent the full range of concept coverage while simultaneously referencing item calibration and calibration-related statistics (e.g., item slope, thresholds, average item difficulty, and item information). CAT scores were simulated using Firestar software [109].

Preliminary reliability and validity analyses

Healthcare Access score data were normally distributed and appropriate for parametric analyses. Internal consistency reliability was examined using Cronbach's alpha for full bank and SF scores and an IRT-based estimate [110] for the simulated CAT scores (a priori criterion for an acceptable reliability level specified as ≥ 0.70 [111]). The percentages of participants who had the highest possible and lowest possible scores for the full bank and the newly developed SF were obtained to establish potential ceiling and floor effects, respectively. We divided the raw CAT item response score by the number of items administered in order to examine floor and ceiling effects for the CAT (i.e., a quotient score of "1" was considered a "floor effect" and a quotient score of "5" a "ceiling effect"). A priori criteria for acceptable floor and ceiling effects were specified as $\leq 20\%$ [112, 113].

Convergent and discriminant validity of the Healthcare Access item bank were examined using Pearson correlations. Convergent validity would be supported by moderate to strong correlations ("moderate" = $r's \geq 0.36$ – 0.67 and "high" = $r's$ between 0.68 and 0.89) between Healthcare Access and community factors related to healthcare (i.e., HEAL Health Care Environment and HEAL Patient–Provider Connection) [114]. Discriminant validity would be

supported by weak correlations ("low" = $r's \leq 0.35$) between Healthcare Access and mental health/substance use/abuse (i.e., Neuro-QoL Depression, Neuro-QoL Anxiety, PROMIS Alcohol Abuse, and PROMIS Severity of Substance Abuse) [114].

Known-groups validity was examined using independent sample t -tests to compare 1) those at high risk for readmission (i.e., ≥ 2 inpatient admissions in the past six months) versus those at low risk for readmission (no or one inpatient admission in the past six months) and 2) those with "worse" economic quality of life (Econ-QOL scores ≥ 60) versus those with "better" economic quality of life (Econ-QOL scores ≤ 40). Known-groups validity would be supported by: (1) those at high-risk reporting worse healthcare access than those at low risk for readmission and (2) those with "worse" economic quality of life reporting worse healthcare access than those with "better" economic quality of life. Finally, we expected greater than 16% of participants at high risk for readmission or with "worse" economic quality of life to have Healthcare Access scores ≥ 1 SD below the mean [115].

Sample size requirements

Sample size specifications were informed by needing to ensure stable parameter calibration for the constrained GRM modeling and for the Wald-2 DIF analyses used in the item bank development process. Published guidelines indicate that a constrained GRM is appropriate for sample sizes less than $N=500$ [99], and, further, that sample sizes of at least $N=200$ are recommended to conduct stable constrained GRM parameter estimation [99, 116]. Published guidelines also indicate that DIF analyses, when using the iterative Wald-2 method, are appropriate for subgroup sizes $\sim \geq 100$ participants [117].

Results

Study participants

Two hundred and twenty-five persons with T2DM participated in this study; data come from a study focused on the development of new patient-reported outcomes that capture important social determinants of health, including the development of the Illness Burden item bank [46] and the Medication Adherence item bank [47], which are also published in this issue of *Quality of Life Research*. Descriptive information is provided in Table 1. Briefly, our sample was, on average, 57.7 years of age ($SD=11$), 52% female, predominantly Black/African American (75%), and 83% non-Hispanic/Latino. Just under a quarter (24%) of the sample reported needing help reading materials from the hospital/

Table 1 Descriptive data for study participants

Variables	Study participants (<i>N</i> = 225)
Age M (SD)	57.7 (11.0)
Sex <i>n</i> (%) Male	107 (48)
Race <i>n</i> (%)	
Caucasian	40 (18)
African American	169 (75)
Other	16 (7)
Ethnicity <i>n</i> (%) Hispanic or Latino	16 (7)
Marital Status <i>n</i> (%)	
Single, Never married	123 (55)
Married/cohabitating	34 (15)
Separated/divorced	48 (21)
Widowed	19 (8)
Missing	1 (<1)
Education <i>n</i> (%)	
Less than High School	42 (19)
High School Graduate or Equivalent	73 (32)
Some College	78 (35)
College Degree	19 (8)
Master Degree or more	13 (6)
Self-reported Annual Income <i>n</i> (%)	
Less than \$5,000	39 (17)
\$5,000–\$14,999	75 (33)
\$15,000–\$29,999	45 (20)
\$30,000–\$49,999	27 (12)
Greater than \$49,999	18 (8)
Missing	21 (9)
How confident are you in filling out medical forms by yourself? <i>n</i> (%)	
Extremely confident	105 (47)
Quite a bit confident	46 (20)
Somewhat confident	45 (20)
A little confident	18 (8)
Not at all confident	10 (4)
Missing	1 (<1)
Do you usually ask someone to help you read materials you receive from the hospital or doctor?, <i>n</i> (%)	
Yes	54 (24)
No	169 (75)
Missing	2 (1)
HbA1c M (SD)	8.1 (2.2)

doctor, and roughly half (47%) of the sample reported an annual income of less than \$15 K per year.

Item bank development

Table 2 outlines the primary findings from the item bank development process. Briefly, EFA analyses supported the unidimensionality of the item pool: The ratio of eigenvalue 1 to eigenvalue 2 was 9.48; eigenvalue 1 accounted for 56.65% of modeled variance, while eigenvalue 2 accounted for only

5.97%. Of the 54 items in the Healthcare Access item pool, one item was eliminated due to a low item-adjusted total score correlation (inclusion criterion $r \geq 0.40$), and nine items were eliminated due to high residual correlations (inclusion criterion $r \leq 0.20$). Subsequent IRT modeling of the 44 remaining items revealed no item misfit (see Supplemental File 3 for item fit chi-square values, degrees of freedom, p values, and chi-square /degrees of freedom quotients). No items had impactful DIF for any of the investigated DIF factors. A final CFA model of these 44 items

Table 2 Unidimensional modeling and analyses for Healthcare Access Item Pool

Domain	Unidimensional modeling										Initial item performance			IRT modeling	
	Item pool	EFA E1/E2 ratio (crit- rion > 4)	Percent of variance for EI (crit- rion > 40)	1-factor CFA loading (crit- rion < .50)	1-factor CFA residual correlation (crit- rion > .20)	1-factor CFA modification index (crit- rion > 100)	Item-adjusted total score correlations (Crite- rion < .40)	Sparse cells (crit- rion < 10)	Problems with mono- tonicity	IRT item misfit	DIF	Interim/Final item bank			
Healthcare Access	54 items	9.48	56.65	0 items	9 items	0 items	1 item	0 items	0 items	0 items	0 items	0 items	44 items		

CFA Confirmatory Factor Analysis; *EFA* Exploratory Factor Analysis; *IRT* Item Response Theory; *DIF* Differential Item Functioning

indicated good item-level and overall model fit (Table 3). Because our CFA model fit criteria were fully met, it was not necessary to conduct CBFA.

The final item calibration estimates are presented in Supplemental File 4. The common slope was 2.08, and thresholds ranged from − 3.53 to +0.07 for the full item set. Test information was excellent (i.e., ≥ 10, with corresponding reliabilities ≥ 0.90) for scaled scores between approximately 10 and 60 (i.e., from −4 SDs to +1 SD; see Fig. 1 for the test information function and standard errors, plotted by theta); marginal reliability was 0.93. With minimum number of items = 4, maximum number of items = 12, and targeted score-level reliability = 0.85, CAT administration tended to use the fewest items (i.e., 4) from the item bank from approximately theta = − 2.0 to theta = − 0.2; the maximum number of items (i.e., 12) was administered by the CAT at theta scores ≥ +0.5 (See Fig. 2, which displays the number of items administered by examinee plotted as a function of theta).

A 6-item SF was constructed using items from the final item bank, employing calibration and calibration-based statistics (e.g., slope, item characteristic curves, item information, and average item difficulty), in conjunction with item content-related clinical coverage considerations. A look-up table to convert raw (summed) scores to T-scores is available in Supplemental File 5. The reliability of the SF was examined on a measurement continuum from approximately theta = − 3.0 (T-score = 20) to + 1.0 (T-score = 60). Score-level reliabilities were very good (i.e., ≥ 0.80) for T-scores between 20 and 55, and good or very good (i.e., ≥ 0.70) for T-scores between 20 and 59.

Preliminary reliability and validity analyses

Internal consistency reliability was excellent for both the CAT and SF administrations (Table 4). The different administration formats were generally free of floor and ceiling effects, although there was evidence for a slight ceiling effect for the SF administration (Table 4).

Correlations supported convergent and discriminant validity (Table 5). Correlations between Healthcare Access and community factors related to healthcare were moderate, supporting convergent validity. In addition, correlations between Healthcare Access and mental health and substance use/abuse measures were generally low, supporting discriminant validity.

Known-groups validity was also generally supported (Table 6). Findings for those at high risk for readmission were in the expected direction, but this difference did not meet the conventional levels of significance (p value = 0.06). As expected, those with “worse” economic quality of life indicated significantly worse healthcare access than did those with “better” economic quality of life. Those

Table 3 Final model fit and reliability characteristics for the Healthcare Access Item Bank

Domain	Item bank	Chi-square (degrees of freedom); p value	CFI (criterion > .90)	TLI (criterion > .90)	CFA-based RMSEA (criterion < .10)	Alpha Reliability (criterion > .80)	IRT SE-based Reliability (criterion > .80)	Response Pattern/Marginal Reliability (criterion > .80)
Healthcare Access	44 items	1497.3 (902); < .001	.96	.96	.054	.97	.97	.93

Chi-square chi-square test of model fit, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index, *RMSEA* Root Mean Square Error of Approximation

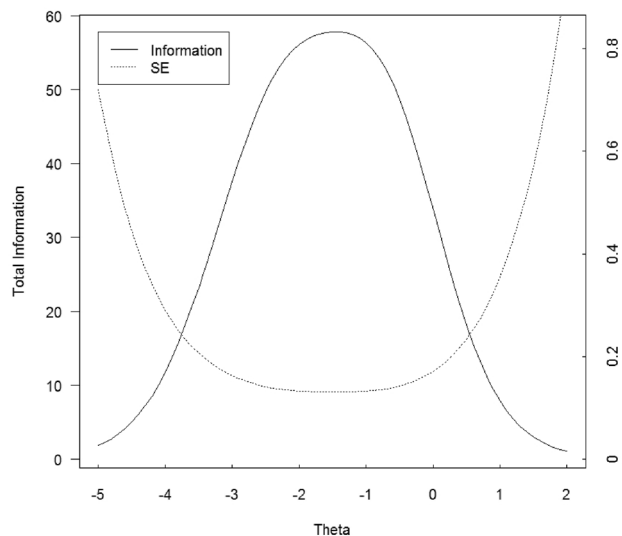


Fig. 1 Healthcare Access Test Information Plot. In general, we would like total test information per score level to be ≥ 10.0 and the resultant standard error to be ≤ 0.32 (which would provide a score-level reliability of ≥ 0.90). This figure shows excellent total test information (left y axis) and standard errors (right y axis) for Healthcare Access scaled T-scores between approximately 10 and 60 (i.e., x axis: theta - 4 to +1)

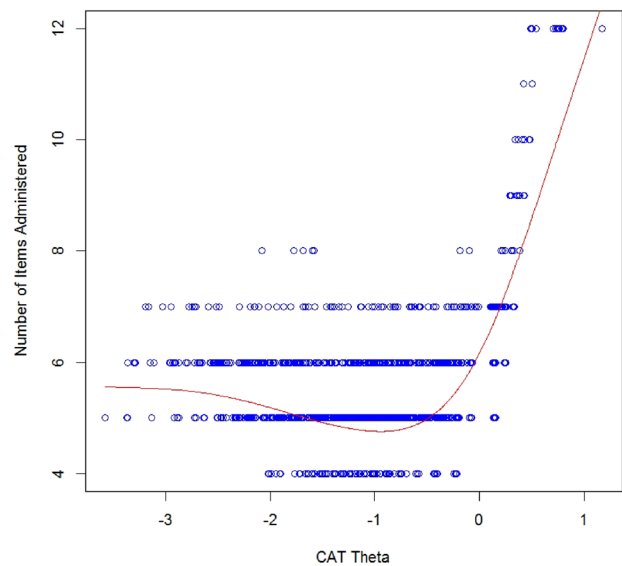


Fig. 2 Simulation data for the Healthcare Access Number of CAT Items by CAT Theta. In this figure, the number of items administered by examinee (the individual blue circles) is plotted as a function (the red curvilinear line) of theta. The figure shows the number of CAT items used for different score levels in standard deviation units: From approximately - 2.0 SD units to - 0.2 SD units, the CAT tended to use the minimum of four items from the item bank; at approximately $\geq +0.5$ SD units ,the maximum of 12 items from the item bank was used by the CAT. (Color figure online)

Table 4 Descriptive data for the different Healthcare Access administration formats

	N	Internal consistency reliability	Mean (SD)	% at floor	% at ceiling
Healthcare access—CAT	225	0.97	49.6 (9.24)	0.00	9.78
Healthcare access—Short form	225	0.82	49.8 (8.65)	0.00	21.78

CAT Computer adaptive test

individuals at high risk for readmission, as well as those with “worse” economic quality of life, were at greater risk of having problems with Healthcare Access (i.e., greater than 16% of individuals in those groups had scores ≥ 1 SD below the mean, relative to the general population of individuals with T2DM).

Discussion

This manuscript describes the development and preliminary psychometric examination of a new patient-reported outcome measure, the REDD-CAT Healthcare Access item bank. This new measure, designed to capture important patient-reported concerns related to access to healthcare services, can be administered as either a computer adaptive

test (CAT) or a 6-item short form. Factor analyses supported the development of this unidimensional measure, and differential item functioning (DIF) analyses indicated that items were devoid of bias for age, sex, education, and socioeconomic status. This new measure will be publicly available through healthmeasures.net, as well as the PROMIS Application Programmable Interface (API), in early 2023.

Preliminary examination of the psychometric properties of this measure indicated that, for both administration formats (i.e., CAT and SF administrations), internal consistency reliability was excellent. In addition, the measures were free of excessive floor or ceiling effects, although there was a slight ceiling effect for the SF administration (but not the CAT administration) of the measure. While this was not consistent with our a priori specifications (we expected those effects to be $\leq 20\%$ and found a ceiling effect for the SF of 21.78%, which slightly exceeded this expectation), this minimal ceiling effect will not have a negative impact on its clinical utility, given that we expect this measure to be used as a screening measure to help identify those individuals with healthcare access problems (which would be impacted by the presence of a floor effect, not a ceiling effect).

With regard to validity, we found evidence for convergent, discriminant, and known-groups validity. Specifically, the overall pattern of correlations was as expected; we had moderate correlations between the Healthcare Access scores and HEAL Health Care Environment and Patient–Provider Connection scores, supporting convergent validity. These findings are consistent with recent research examining barriers to Healthcare Access among a diverse sample of

Table 5 Convergent and discriminant validity for the Healthcare Access Item Bank

Healthcare access score	Convergent validity		Discriminant validity			
	HEAL Healthcare environment	HEAL Patient–provider connection	Neuro-QoL depression	Neuro-QoL anxiety	PROMIS alcohol use	PROMIS Severity of substance use (past 30 days)
Healthcare access full bank score	0.42	0.37	– 0.37	– 0.32	– 0.30	0.11

CAT Computer adaptive test; SF Short form

Table 6 Known-groups validity

Healthcare access score	Low risk for readmission in the past 6 months			High Risk for Readmission in the past 6 months			<i>t</i>	<i>p</i>	<i>d</i>	
	<i>n</i>	Mean (SD)	%	<i>n</i>	Mean (SD)	%				
										Impaired ^a
Full bank score	175	50.7 (9.6)	13.7	50	47.7 (9.6)	26.0	1.92	0.06	0.31	
		“Better” Economic Quality of Life (scores ≥ 60)			“Worse” Economic Quality of Life (scores ≤ 40)					
Full bank score	16	52.2 (11.0)	18.8	63	46.9 (8.4)	22.2	– 2.14	0.04	0.59	

^aHealthcare access score ≤ 40

publicly insured adults that identified health insurance coverage, logistical barriers, and patient–provider trust among the top-endorsed barriers to access [118]. We also had low correlations between the Healthcare Access full bank scores and Depression scores as well as alcohol/substance use scores, which provided support for discriminant validity. The relationship between the Healthcare Access full bank scores and Depression scores did not meet the a priori cutoff (i.e., <0.36) for discriminant validity; the magnitude of the relationship was 0.366, which just exceeded this specification. Given that this was the only relationship that did not meet the pre-specified criterion for convergent and discriminant validity, and more than 75% of our validity analyses were consistent with proposed expectations, Healthcare Access scores still meet the established standards for construct validity [119].

In addition to convergent and discriminant validity, known-groups validity was also supported. For example, individuals with worse economic quality of life reported significantly worse healthcare access and were at higher risk for elevated scores on the new REDD-CAT Healthcare Access measure, relative to those with better economic quality of life. This finding is consistent with a considerable body of literature that finds a robust relationship socioeconomic status and healthcare access [4, 13, 14, 118, 120, 121]. In addition, there was a trend for those at high risk for readmission to have worse reported Healthcare Access compared to those with low readmission risk. This is consistent with findings showing that individuals who do not have healthcare access to post-acute care services in the community are more likely to end up at the hospital [122]. Those individuals with worse economic quality of life, as well as those with high readmission risk, also had elevated rates of healthcare access problems, as hypothesized.

This is the first comprehensive measurement system, to our knowledge, to include a patient-reported measure about healthcare access, which is an important social determinant of health that is related to readmission risk in persons with T2DM. This new measure, REDD-CAT Healthcare Access, can be administered as a computer adaptive test, a format where each subsequent item administered is selected based on a participant's previous response, in essence maximizing brevity without losing the precision that is afforded by administering more items [44]. This new measure is scored on a T-score metric ($M = 50$, $SD = 10$), with higher scores indicating better perceived healthcare access. This type of standardized score increases the clinical utility of the tool, given that obtained scores can be directly compared to the reference group (in this case, other individuals with T2DM). For example, individuals with scores of 40 or less (i.e., ≤ 1 SD below the mean) report more significant barriers to Healthcare Access than 83.9% of the broader T2DM population. Furthermore, individuals with scores 30 or less (i.e., ≤ 2

SD s below the mean) are reporting barriers at a rate that exceeds 97.9% of their peers. When deciding among the different administration modalities (long form, SF, or CAT), we would encourage people to focus on using either the SF or the CAT administration, as the overall participant burden (of administering all of the items in the bank) is not outweighed by the small gains in precision of the long form administration. When deciding between SF and CAT administration, there are practical reasons to consider. For example, CAT administration requires an electronic delivery platform with a live internet connection to run the CAT administration programs that are included as part of the PROMIS API. In addition, these electronic data capture platforms (such as REDCap or healthmeasures.net) may incur additional costs for the researcher/clinician. Regardless, given the slightly better psychometric performance of the REDD-CAT Healthcare Access CAT (i.e., the absence of the ceiling effect), we would recommend CAT administration when not precluded by practical considerations that would warrant SF administration.

Bearing in mind the advances offered by this new measure to capture and identify individuals with T2DM who have significant Healthcare Access concerns, we acknowledge several study limitations. First, while the sample size met minimal requirements for the analyses (i.e., EFA, CFA, constrained GRM, and DIF) that informed the development of this new measure, and while this approach has been successfully applied to PRO measurement development in other settings [30, 123], larger samples tend to exhibit more stable calibration parameter estimates and provide a more robust estimate of DIF; as such, future work that confirms EFA, CFA, calibration parameters, and DIF analyses in independent samples is needed. Similarly, the preliminary reliability and validity data that are reported herein should be replicated in an independent sample. In addition, the CAT data presented were based on simulations and therefore need to be replicated in a clinical sample using an actual CAT engine. In addition, the study sample setting was an urban safety-net health system, which is likely to treat patient that lack adequate healthcare coverage and have patients with a high number of unmet social needs relative to other types of healthcare systems [2]. As such, results may not be generalizable to other types of hospital systems. The study population also included mainly non-white T2DM patients, which may limit generalizability to Caucasians.

In sum, the REDD-CAT Healthcare Access measure provides a brief, reliable, and valid assessment of T2DM patients' healthcare access concerns. This new measure can be used to aid discharge planning for those individuals with T2DM who have recently been hospitalized, in order to screen for individuals that may be experiencing difficulties within the healthcare system. In addition, although this measure was developed specifically for use in persons with

T2DM, it may also have clinical utility in other medical populations.

Conclusions

The REDD-CAT measurement system is the first comprehensive system designed to assess social determinants of health among persons with T2DM. The new REDD-CAT Healthcare Access measure captures an important social determinant of behavior, namely, patient experiences with access to healthcare services, including healthcare coverage, actual provision of healthcare services, and timeliness of receiving services. This measure will be available for public use as a part of the healthmeasures.net platform. Additional research is needed to elucidate the mediation and moderation effects that perceived healthcare access has on outcomes in these individuals. Understanding these relationships will help inform targeted interventions designed to minimize readmission risk and improve patient HRQOL.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-022-03278-w>.

Acknowledgements We thank the investigators and research associates who worked on the study, the study participants, and organizations who supported recruitment efforts. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author contributions NEC: Principal Investigator; data coordination and analysis site; analysis design; initial draft of introduction, methods, results, and discussion; incorporation of revisions. MAK: Study Co-Investigator and Statistician; assistance with analysis design; primary statistician for measurement development portions of manuscript (e.g., factor analyses and IRT analyses); drafted statistical analysis section and a template for the results section. JPT: Statistician and Data Analyst; primary statistician for reliability and validity analyses; assistance writing methods and results sections; review and feedback on manuscript drafts. AB: Study Research Coordinator; responsible for data collection; review and feedback on manuscript drafts (critical review of the methods). JMH: Study Grants Manager and Project Coordinator; review and feedback on manuscript drafts; assistance with study regulatory documents. BDLC: Project Manager; review and feedback on manuscript drafts; assistance with synthesis of revisions. JAM: Study Data Manager; critical review of the methods; review and feedback on manuscript drafts. IM: Study Research Coordinator; responsible for data collection; review and feedback on manuscript drafts (critical review of the methods). BWJ: Principal Investigator of the PRET study; review and feedback on manuscript drafts (critical review of the summary of the qualitative work that informed this study). SM: Principal Investigator; data collection site; study design; critical review of the introduction and discussion; review and feedback on manuscript drafts.

Funding Work on this manuscript was supported by grant number R21DK121092 (PIs Carlozzi; Mitchell [admin]) from the National Institutes of Health (NIH), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and by the National Center for Advancing Translational Sciences (NCATS) for the Michigan Institute for Clinical and Health Research (UL1TR002240).

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. Dr. Mitchell has received compensation as a lecturer on Relationship Centered Care by Merck & co and holds equity in See Yourself Health Inc, a digital health company.

Ethical approval Approval was obtained from the ethics committee of Boston Medical Center. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to participate Informed consent was obtained from all participants prior to their participation in this study.

References

1. Agency for Healthcare Research and Quality. (2016). *Statistical Brief #72 Healthcare Cost and Utilization Project (HCUP)*. AHRQ.
2. U.S Department of Health and Human Services, & Prevention, C. f. D. C. a. (2020). National diabetes statistics report 2020: Estimates of diabetes and its burden in the United States. *cdc.gov*.
3. Elixhauser, A., & Steiner, C. (2006). *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Agency for Health Care Policy and Research.
4. Clark, M. L., & Utz, S. W. (2014). Social determinants of type 2 diabetes and health in the United States. *World Journal of Diabetes*, 5(3), 296–304.
5. Ong, S. E., Koh, J. J. K., Toh, S. E. S., Chia, K. S., Balabanova, D., McKee, M., Perel, P., & Legido-Quigley, H. (2018). Assessing the influence of health systems on type 2 diabetes mellitus awareness, treatment, adherence, and control: A systematic review. *PLoS ONE*, 13(3), e0195086.
6. Mansyur, C. L., Rustveld, L. O., Nash, S. G., & Jibaja-Weiss, M. L. (2015). Social factors and barriers to self-care adherence in Hispanic men and women with diabetes. *Patient Education and Counseling*, 98(6), 805–810.
7. Titus, S. K., & Kataoka-Yahiro, M. (2021). Barriers to access to care in Hispanics With type 2 diabetes: A systematic review. *Hisp Health Care Int*, 19(2), 118–130.
8. Titus, S. K., & Kataoka-Yahiro, M. (2019). A Systematic review of barriers to access-to-care in Hispanics with type 2 diabetes. *Journal of Transcultural Nursing*, 30(3), 280–290.
9. Harris, M. I. (2001). Racial and ethnic differences in health care access and health outcomes for adults with type 2 diabetes. *Diabetes Care*, 24(3), 454–459.
10. Canedo, J. R., Miller, S. T., Schlundt, D., Fadden, M. K., & Sanderson, M. (2018). Racial/Ethnic disparities in diabetes quality of care: The role of healthcare access and socioeconomic status. *Journal of Racial and Ethnic Health Disparities*, 5(1), 7–14.
11. Campbell, J. A., & Egede, L. E. (2020). Individual-, community-, and health system-level barriers to optimal type 2 diabetes care for inner-city African Americans: An integrative review and model development. *The Diabetes Educator*, 46(1), 11–27.
12. Suresh, N., & Thankappan, K. R. (2019). Gender differences and barriers women face in relation to accessing type 2 diabetes care: A systematic review. *Indian Journal of Public Health*, 63(1), 65–72.
13. Tan, X., Lee, L. K., Huynh, S., Pawaskar, M., & Rajpathak, S. (2020). Sociodemographic disparities in the management of type

- 2 diabetes in the United States. *Current Medical Research and Opinion*, 36(6), 967–976.
14. Bijlsma-Rutte, A., Rutters, F., Elders, P. J. M., Bot, S. D. M., & Nijpels, G. (2018). Socio-economic status and HbA1c in type 2 diabetes: A systematic review and meta-analysis. *Diabetes/Metabolism Research and Reviews*, 34(6), e3008.
 15. Institute of Medicine Board of Health Care Services. (2010). *Future directions for the National Healthcare Quality and Disparities Reports*. National Academic Press.
 16. Healthy People. (2020). Access to Health Services. Retrieved October 16, 2022, from <http://www.healthypeople.gov/2020/topics-objectives/topic/Access-to-Health-Services>
 17. Gulliford, M., Figueroa-Munoz, J., Morgan, M., Hughes, D., Gibson, B., Beech, R., & Hudson, M. (2002). What does 'access to health care' mean? *Journal of Health Services Research & Policy*, 7(3), 186–188.
 18. Berk, M. L., & Schur, C. L. (1998). Measuring access to care: Improving information for policymakers. *Health Affairs (Millwood)*, 17(1), 180–186.
 19. Quinn, M., Robinson, C., Forman, J., Krein, S. L., & Rosland, A. M. (2017). Survey instruments to assess patient experiences with access and coordination across health care settings: available and needed measures. *Medical Care*, 55(1), S84–S91.
 20. Centers for Medicare & Medicaid. (2018). Measures management system. Retrieved October 16, 2022, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Access-Measures.pdf>
 21. McLafferty, S. L. (2003). GIS and health care. *Annual Review of Public Health*, 24, 25–42.
 22. Hoseini-Esfidarjani, S. S., Negarandeh, R., Delavar, F., & Janani, L. (2021). Psychometric evaluation of the perceived access to health care questionnaire. *BMC Health Services Research*, 21(1), 638.
 23. Scatena, L. M., Wysocki, A. D., Beraldo, A. A., Magnabosco, G. T., Brunello, M. E., Netto Ruffino, A., Nogueira Jde, A., Silva Sobrinho, R. A., Brito, E. W., Alexandre, P. B., Monroe, A. A., & Villa, T. C. (2015). Validity and reliability of a health care service evaluation instrument for tuberculosis. *Revista de Saude Publica*, 49, 7.
 24. Evensen, C. T., Yost, K. J., Keller, S., Arora, N. K., Frentzel, E., Cowans, T., & Garfinkel, S. A. (2019). Development and testing of the CAHPS cancer care survey. *Journal of Oncology Practice/American Society of Clinical Oncology*, 15(11), e969–e978.
 25. Fortney, J. C., Pyne, J. M., Hawrilenko, M., Bechtel, J. M., Moore, D., Nolan, J. P., Pfeiffer, P., Shushan, S., Shore, J. H., & Bowen, D. (2021). Psychometric properties of the assessment of perceived access to care (APAC) instrument. *Journal of Ambulatory Care Management*, 44(1), 31–45.
 26. Clement, S., Brohan, E., Jeffery, D., Henderson, C., Hatch, S. L., & Thornicroft, G. (2012). Development and psychometric properties the barriers to access to care evaluation scale (BACE) related to people with mental ill health. *BMC Psychiatry*, 12, 36.
 27. Pyne, J. M., Kelly, P. A., Fischer, E. P., Miller, C. J., Wright, P., Zamora, K., Koenig, C. J., Stanley, R., Seal, K., Burgess, J. F., & Fortney, J. C. (2020). Development of the perceived access inventory: A patient-centered measure of access to mental health care. *Psychological Services*, 17(1), 13–24.
 28. Drum, C. E., Horner-Johnson, W., & Walsh, E. S. (2012). Construction and validation of the outpatient health care usability profile (OHCU). *Disability and Health Journal*, 5(4), 292–297.
 29. Haggerty, J. L., & Levesque, J. F. (2017). Validation of a new measure of availability and accommodation of health care that is valid for rural and urban contexts. *Health Expectations*, 20(2), 321–334.
 30. Carlozzi, N., Lange, R., French, L., Kallen, M., Boileau, N., Hanks, R., Nakase-Richardson, R., Massengale, J., Sander, A., Hahn, E., Miner, J., & Brickell, T. (2020). TBI-CareQOL military health care frustration in caregivers of service members/veterans with traumatic brain injury. *Rehabilitation Psychology*, 65(4), 360–376.
 31. Kaitelidou, D., Economou, C., Galanis, P., Konstantakopoulou, O., Siskou, O., Domete, S., de Boer, D., Boerma, W. G., & Groenewegen, P. P. (2019). Development and validation of measurement tools for user experience evaluation surveys in the public primary healthcare facilities in Greece: A mixed methods study. *BMC Family Practice*, 20(1), 49.
 32. Desborough, J., Banfield, M., & Parker, R. (2014). A tool to evaluate patients' experiences of nursing care in Australian general practice: Development of the patient enablement and satisfaction survey. *Australian Journal of Primary Health*, 20(2), 209–215.
 33. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., DeVellis, R., DeWalt, D., Fries, J., Gershon, R., Hahn, E., Lai, J.-S., Pilkonis, P., Revicki, D., ... Hays, R. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested in its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, 63, 1179–1194.
 34. Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. L. Erlbaum Associates.
 35. Wu, M., Jen, T.-H., & Tam, H. P. (2016). *Educational Measurement for Applied Researchers: Theory into Practice* (1st ed., pp. 1 online resource (XIV, 306 pages 114 illustrations, 374 illustrations in color). Imprint: Springer, Singapore
 36. Bjorner, J. B., Chang, C. H., Thissen, D., & Reeve, B. B. (2007). Developing tailored instruments: Item banking and computerized adaptive assessment. *Quality of Life Research*, 16(Suppl 1), 95–108.
 37. Cella, D., Gershon, R., Lai, J. S., & Choi, S. (2007). The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*, 16(Suppl 1), 133–141.
 38. van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized adaptive testing : theory and practice*. Kluwer Academic.
 39. Rose, M., Bjorner, J. B., Fischer, F., Anatchkova, M., Gandek, B., Klapp, B. F., & Ware, J. E. (2012). Computerized adaptive testing—ready for ambulatory monitoring? *Psychosomatic Medicine*, 74(4), 338–348.
 40. Smits, N., Paap, M. C. S., & Bohnke, J. R. (2018). Some recommendations for developing multidimensional computerized adaptive tests for patient-reported outcomes. *Quality of Life Research*, 27(4), 1055–1063.
 41. Thissen, D., Reeve, B. B., Bjorner, J. B., & Chang, C. H. (2007). Methodological issues for building item banks and computerized adaptive scales. *Quality of Life Research*, 16(Suppl 1), 109–119.
 42. Wainer, H., & Dorans, N. J. (2000). *Computerized adaptive testing : A primer* (2nd ed.). Lawrence Erlbaum Associates.
 43. Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS((R)) adult health profiles: efficient short-form measures of seven health domains. *Value Health*, 22(5), 537–544.
 44. Lai, J. S., Cella, D., Choi, S., Junghaenel, D. U., Christoudolou, C., Gershon, R., & Stone, A. (2011). How item banks and its applications can influence measurement practice in rehabilitation medicine: A PROMIS fatigue item bank example. *Archives of Physical Medicine & Rehabilitation*, 92(Suppl 1), S20–S27.
 45. Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, 36(5), 648–662.

46. Mitchell, S., Kallen, M. A., Troost, J. P., Bragg, A., Marin-Howard, J., Moldovan, I., Miner, J. A., Jack, B. W., & Carlozzi, N. E. (Under Review). Development and calibration data for the Illness Burden Item Bank: A new computer adaptive test for persons with type 2 diabetes mellitus. *Quality of Life Research*.
47. Carlozzi, N. E., Kallen, M., Troost, J., Bragg, A., Marin-Howard, J., Moldovan, I., Miner, J. A., Jack, B. W., & Mitchell, S. (In Press). Development and calibration data for the Medication Adherence Item Bank: A new computer adaptive test for persons with type 2 diabetes mellitus. *Quality of Life Research*.
48. Hahn, E. A., & Cella, D. (2003). Health outcomes assessment in vulnerable populations: Measurement challenges and recommendations. *Archives of Physical Medicine and Rehabilitation*, 84(4 Suppl 2), S35–42.
49. Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT 4: Wide range achievement test professional manual*. Psychological Assessment Resources Inc.
50. PROMIS® Instrument Development and Psychometric Evaluation Scientific Standards, http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf. (Vol. 2019)
51. Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33(8), 644–645.
52. Cella, D., Nowinski, C., Peterman, A., Victorson, D., Miller, D., Lai, J.-S., & Moy, C. (2011). The neurology quality of life measurement (Neuro-QOL) Initiative. *Archives of Physical Medicine and Rehabilitation, Supplement*, 92(Suppl 1), S28–S36.
53. Cella, D., Lai, J. S., Nowinski, C., Victorson, D., Peterman, A., Miller, D., Bethoux, F., Heinemann, A., Rubin, S., Cavazos, J., Reeder, A., Sufit, R., Simuni, T., Holmes, G., Siderowf, A., Wojna, V., Bode, R., McKinney, N., Podrabsky, T., ... Moy, C. (2012). Neuro-QOL: Brief measures of health-related quality of life for clinical research in neurology. *Neurology*, 78, 1860–1867.
54. Gershon, R., Lai, J.-S., Bode, R., Choi, S., Moy, C., Bleck, T., Miller, D., Peterman, A., & Cella, D. (2012). Neuro-QOL: Quality of life item banks for adults with neurological disorders: Item development and calibrations based upon clinical and general population testing. *Quality of Life Research*, 21(3), 475–486.
55. Carlozzi, N. E., Boileau, N. R., Roche, M. W., Ready, R. E., Perlmutter, J. S., Chou, K. L., Barton, S. K., McCormack, M. K., Stout, J. C., Cella, D., Miner, J. A., & Paulsen, J. S. (2020). Responsiveness to change over time and test-retest reliability of the PROMIS and Neuro-QoL mental health measures in persons with Huntington disease (HD). *Quality of Life Research*, 29(12), 3419–3439.
56. Carlozzi, N. E., Goodnight, S., Kratz, A. L., Stout, J. C., McCormack, M. K., Paulsen, J. S., Boileau, N. R., Cella, D., & Ready, R. E. (2019). Validation of neuro-QoL and PROMIS mental health patient reported outcome measures in persons with huntington disease. *J Huntingtons Dis*, 8(4), 467–482.
57. Miller, D. M., Bethoux, F., Victorson, D., Nowinski, C. J., Buono, S., Lai, J. S., Wortman, K., Burns, J. L., Moy, C., & Cella, D. (2015). Validating neuro-QoL short forms and targeted scales with people who have multiple sclerosis. *Multiple Sclerosis*. <https://doi.org/10.1177/1352458515599450>
58. Nowinski, C. J., Siderowf, A., Simuni, T., Wortman, C., Moy, C., & Cella, D. (2016). Neuro-QoL health-related quality of life measurement system: Validation in parkinson's disease. *Movement Disorders*, 31(5), 725–733.
59. Victorson, D., Cavazos, J. E., Holmes, G. L., Reeder, A. T., Wojna, V., Nowinski, C., Miller, D., Buono, S., Mueller, A., Moy, C., & Cella, D. (2014). Validity of the neurology quality-of-life (Neuro-QoL) measurement system in adult epilepsy. *Epilepsy Behavior*, 31, 77–84.
60. Healy, B. C., Zurawski, J., Gonzalez, C. T., Chitnis, T., Weiner, H. L., & Glanz, B. I. (2019). Assessment of computer adaptive testing version of the Neuro-QOL for people with multiple sclerosis. *Multiple Sclerosis*, 25(13), 1791–1799.
61. Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Lawrence, S. M., & Daley, D. C. (2016). Validation of the alcohol use item banks from the patient-reported outcomes measurement information system (PROMIS((R))). *Drug and Alcohol Dependence*, 161, 316–322.
62. Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Lawrence, S. M., Hilton, T. F., Daley, D. C., Patkar, A. A., & McCarty, D. (2015). Item banks for substance use from the patient-reported outcomes measurement information system (PROMIS((R))): Severity of use and positive appeal of use. *Drug and Alcohol Dependence*, 156, 184–192.
63. Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Lawrence, S. M., & Daley, D. C. (2016). Validation of the alcohol use item banks from the patient-reported outcomes measurement information system (PROMIS). *Drug and Alcohol Dependence*, 161, 316–322.
64. Pilkonis, P. A., Yu, L., Colditz, J., Dodds, N., Johnston, K. L., Maihoefer, C., Stover, A. M., Daley, D. C., & McCarty, D. (2013). Item banks for alcohol use from the patient-reported outcomes measurement information system (PROMIS): Use, consequences, and expectancies. *Drug and Alcohol Dependence*, 130(1–3), 167–177.
65. Greco, C. M., Yu, L., Johnston, K. L., Dodds, N. E., Morone, N. E., Glick, R. M., Schneider, M. J., Klem, M. L., McFarland, C. E., Lawrence, S., Colditz, J., Maihoefer, C. C., Jonas, W. B., Ryan, N. D., & Pilkonis, P. A. (2015). Measuring nonspecific factors in treatment: item banks that assess the healthcare experience and attitudes from the patient's perspective. *Quality of Life Research*. <https://doi.org/10.1007/s11136-015-1178-1>
66. Tulsy, D. S., Kisala, P. A., Lai, J. S., Carlozzi, N., Hammel, J., & Heinemann, A. W. (2015). Developing an item bank to measure economic quality of life for individuals with disabilities. *Archives of Physical Medicine and Rehabilitation*, 96(4), 604–613.
67. Heinemann, A. W., Magasi, S., Hammel, J., Carlozzi, N. E., Garcia, S. F., Hahn, E. A., Lai, J. S., Tulsy, D., Gray, D. B., Hollingsworth, H., & Jerousek, S. (2015). Environmental factors item development for persons with stroke, traumatic brain injury, and spinal cord injury. *Archives of Physical Medicine and Rehabilitation*, 96, 589–595.
68. Tulsy, D. S., Kisala, P. A., Lai, J. S., Carlozzi, N., Hammel, J., & Heinemann, A. W. (2015). Developing an item bank to measure economic quality of life for individuals with disabilities. *Archives of Physical Medicine and Rehabilitation*, 96, 604–613.
69. Sherer, M., Nick, T. G., Sander, A. M., Melguizo, M., Hanks, R., Novack, T. A., Tulsy, D., Kisala, P., Luo, C., & Tang, X. (2017). Groupings of persons with traumatic brain injury: A new approach to classifying traumatic brain injury in the post-acute period. *The Journal of Head Trauma Rehabilitation*, 32(2), 125–133.
70. Carlozzi, N. E., Graves, C., Troost, J., Kallen, M. A., Tulsy, D. S., Hanks, R. A., Ianni, P. A., Miner, J. A., & Sander, A. M. (In Press). Initial evidence for reliable and valid use of scores on the 8-item Econ-QOL short form to measure economic quality of life in caregivers of persons with traumatic brain injury. *Archives of Physical Medicine & Rehabilitation*.
71. Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381.
72. Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., Consortium R. E. (2019). The REDCap consortium:

- Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208.
73. PROMIS® Instrument Development and Psychometric Evaluation Scientific Standards. https://www.mcgill.ca/can-pro-network/files/can-pro-network/promisstandards_vers2.0_final.pdf.
 74. Strunin, L., Stone, M., & Jack, B. (2007). Understanding rehospitalization risk: Can hospital discharge be modified to reduce recurrent hospitalization? *Journal of Hospital Medicine*, 2(5), 297–304.
 75. Jack, B., Greenwald, J., Forsythe, S., O'Donnell, J., Johnson, A., Schipelliti, L., Goodwin, M., Burniske, G. M., Hesko, C., Paasche-Orlow, M., Manasseh, C., Anthony, D., Martin, S., Hollister, L., Jack, M., Jhaveri, V., Casey, K., & Chetty, V. K. (2008). Developing the tools to administer a comprehensive hospital discharge program: The reengineered discharge (RED) program. In B. J. Henriksen, M. A. Keyes, & M. L. Grady (Eds.), *Advances in patient safety: New directions and alternative approaches* (Vol. 3). Agency for Healthcare Research and Quality.
 76. Anthony, D., Chetty, V. K., Kartha, A., McKenna, K., DePaoli, M. R., & Jack, B. (2005). Re-engineering the hospital discharge: An example of a multifaceted process evaluation. In K. Henriksen, J. B. Battles, E. S. Marks, & D. I. Lewin (Eds.), *Advances in patient safety: From research to implementation (Volume 2: Concepts and Methodology)*. Agency for Healthcare Research and Quality.
 77. Mitchell, S. E., Laurens, V., Weigel, G. M., Hirschman, K. B., Scott, A. M., Nguyen, H. Q., Howard, J. M., Laird, L., Levine, C., Davis, T. C., Gass, B., Shaid, E., Li, J., Williams, M. V., & Jack, B. W. (2018). Care transitions from patient and caregiver perspectives. *Annals of Family Medicine*, 16(3), 225–231.
 78. Li, J., Brock, J., Jack, B., Mittman, B., Naylor, M., Sorra, J., Mays, G., Williams, M. V., Project A. T. (2016). Project ACHIEVE—using implementation research to guide the evaluation of transitional care effectiveness. *BMC Health Services Research*, 16, 70.
 79. Naylor, M. D., Shaid, E. C., Carpenter, D., Gass, B., Levine, C., Li, J., Malley, A., McCauley, K., Nguyen, H. Q., Watson, H., Brock, J., Mittman, B., Jack, B., Mitchell, S., Callicoate, B., Schall, J., & Williams, M. V. (2017). Components of comprehensive and effective transitional care. *Journal of the American Geriatrics Society*, 65(6), 1119–1125.
 80. Cancino, R., Esang, M., Martin, J., Morton, S., Kazis, L., & Jack, B. (2014). *Development of a patient-centered instrument to evaluate 30-day hospital readmission*. Academy Health.
 81. McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates Inc.
 82. Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research*, 18(4), 447–460.
 83. Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(Suppl 1), 19–31.
 84. Muthén, L. K., & Muthén, B. O. (2011). *Mplus user's guide*. Muthén & Muthén.
 85. Ramsay, J. (2000). *TestGraf a program for the graphical analysis of multiple choice test and questionnaire data*. McGill University.
 86. Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). Guilford Press.
 87. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
 88. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling—a Multidisciplinary Journal*, 6(1), 1–55.
 89. Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. SAS Institute Inc.
 90. Lai, J.-S., Zelko, F., Krull, K., Cella, D., Nowinski, C., Manley, P., & Goldman, S. (2014). Parent-reported cognition of children with cancer and its potential clinical usefulness. *Quality of Life Research*, 23(4), 1049–1058.
 91. Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education*, 80(1), 26–44.
 92. Kaplan, D. (1989). Model modification in covariance structure analysis: application of the expected parameter change statistic. *Multivariate Behav Res*, 24(3), 285–305.
 93. Lujben, T. C., & Boomsma, A. (1988). Statistical guidance for model modification in covariance structure analysis. *Compstat*, 1988, 335–340.
 94. Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models for detection of misspecifications. *Structural Equation Modeling*, 16, 561–582.
 95. Saris, W. E., Satorra, A., & Sorbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 105–129). Jossey-Bass.
 96. Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114(3), 552–566.
 97. Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73, 5–26.
 98. Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223–237.
 99. Samejima, F. (1969). *Estimation of Latent Ability Using a Response Pattern of Graded Scores (Psychometric Monograph No. 17)*. Psychometric Society.
 100. Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91(1), 25–39.
 101. Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143–165.
 102. Crisan, D. R., Tendeiro, J. N., & Meijer, R. R. (2017). Investigating the practical consequences of model misfit in unidimensional IRT models. *Applied Psychological Measurement*, 41, 439–455.
 103. Zhao, Y. (2017). Impact of IRT item misfit on score estimates and severity classifications: An examination of PROMIS depression and pain interference item banks. *Quality of Life Research*, 26(3), 555–564.
 104. Edelen, M. O., Stucky, B. D., & Chandra, A. (2015). Quantifying 'problematic' DIF within an IRT framework: Application to a cancer stigma index. *Quality of Life Research*, 24(1), 95–103.
 105. Clauser, B. E., & Hambleton, R. K. (1994). Review of differential item functioning. *Journal of Educational Measurement*, 31(1), 88–92.
 106. Cai, L., Thissen, D., & du Toit, S. H. C. (2015). *IRTpro for Windows [Computer software]*. Scientific Software International.
 107. Woods, C. M., Cai, L., & Wang, M. (2013). The langer-improved wald test for DIF testing with multiple groups: evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*, 73(3), 532–547.

108. Wang, M., & Woods, C. M. (2017). Anchor selection using the wald test anchor-all-test-all procedure. *Applied Psychological Measurement, 41*(1), 17–29.
109. Choi, S. W., Podrabsky, T., & McKinney, N. (2012). Firestar-D: Computerized adaptive testing simulation program for dichotomous item response theory models. *Applied Psychological Measurement, 36*(1), 67–68.
110. Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Maihoefer, C. C., & Lawrence, S. M. (2014). Validation of the depression item bank from the patient-reported outcomes measurement information system (PROMIS) in a three-month observational study. *Journal of Psychiatric Research, 56*, 112–119.
111. Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105.
112. Andresen, E. M. (2000). Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine & Rehabilitation, 81*(12 Suppl 2), S15–20.
113. Cramer, D., & Howitt, D. L. (2004). *The Sage dictionary of statistics*. Sage.
114. Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic and Medical Sonography, 1*, 35–39.
115. Heaton, R. K., Miller, S. W., Taylor, J. T., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Psychological Assessment Resources Inc.
116. Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14*, 59–71.
117. Clauser, B. E., & Hambleton, R. K. (1994). Review of differential item functioning. *Journal of Educational Measurement, 31*(1), 88–92.
118. Call, K. T., McAlpine, D. D., Garcia, C. M., Shippee, N., Beebe, T., Adeniyi, T. C., & Shippee, T. (2014). Barriers to care in an ethnically diverse publicly insured population: Is health care reform enough? *Medical Care, 52*(8), 720–727.
119. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42.
120. McMaughan, D. J., Oloruntoba, O., & Smith, M. L. (2020). Socioeconomic status and access to healthcare: Interrelated drivers for healthy aging. *Frontiers in Public Health, 8*, 231.
121. Craig, L., Cruz, S. M., & Alper, H. (2019). In J. Alper, S. M. Cruz & L. Graig (Eds.), *Improving Access to and Equity of Care for People with Serious Illness: Proceedings of a Workshop*. Washington (DC)
122. Lewis, J. M., DiGiacomo, M., Currow, D. C., & Davidson, P. M. (2011). Dying in the margins: Understanding palliative care and socioeconomic deprivation in the developed world. *Journal of Pain and Symptom Management, 42*(1), 105–118.
123. Ruo, B., Choi, S. W., Baker, D. W., Grady, K. L., & Cella, D. (2010). Development and validation of a computer adaptive test for measuring dyspnea in heart failure. *Journal of Cardiac Failure, 16*(8), 659–668.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.