Development of the Cochlear Implant Quality of Life Item Bank

Theodore R. McRackan,[1] Brittany N. Hand,[2] Cochlear Implant Quality of Life Development Consortium,[3] Craig A. Velozo,[2] and Judy R. Dubno[1]

Objectives: Functional outcomes following cochlear implantation have traditionally been focused on word and sentence recognition, which, although important, do not capture the varied communication and other experiences of adult cochlear implant (CI) users. Although the inadequacies of speech recognition to quantify CI user benefits are widely acknowledged, rarely have adult CI user outcomes been comprehensively assessed beyond these conventional measures. An important limitation in addressing this knowledge gap is that patient-reported outcome measures have not been developed and validated in adult CI patients using rigorous scientific methods. The purpose of the present study is to build on our previous work and create an item bank that can be used to develop new patient-reported outcome measures that assess CI quality of life (QOL) in the adult CI population.

Design: An online questionnaire was made available to 500 adult CI users who represented the adult CI population and were recruited through a consortium of 20 CI centers in the United States. The questionnaire included the 101 question CIQOL item pool and additional questions related to demographics, hearing and CI history, and speech recognition scores. In accordance with the Patient-Reported Outcomes Measurement Information System, responses were psychometrically analyzed using confirmatory factor analysis and item response theory.

Results: Of the 500 questionnaires sent, 371 (74.2%) subjects completed the questionnaire. Subjects represented the full range of age, durations of CI use, speech recognition abilities, and listening modalities of the adult CI population; subjects were implanted with each of the three CI manufacturers' devices. The initial item pool consisted of the following domain constructs: communication, emotional, entertainment, environment, independence, listening effort, and social. Through psychometric analysis, after removing locally dependent and misfitting items, all of the domains were found to have sound psychometric properties, with the exception of the independence domain. This resulted in a final CIQOL item bank of 81 items in 6 domains with good psychometric properties.

Conclusions: Our findings reveal that hypothesis-driven quantitative analyses result in a psychometrically sound CIQOL item bank, organized into unique domains comprised of independent items which measure the full ability range of the adult CI population. The final item bank will now be used to develop new instruments that evaluate and differentiate adult CIQOL across the patient ability spectrum.

Key words: Cochlear implantation, Cochlear implants, Patient reported outcomes measurement information system, Patient-reported outcome measures, Quality of life.

[1]Department of Otolaryngology-Head and Neck Surgery, [2]Division of Occupational Therapy, and [3]Cochlear Implant Quality of Life Development Consortium, Medical University of South Carolina, Charleston, South Carolina, USA.

INTRODUCTION

Substantial evidence exists that cochlear implantation has a dramatic impact on a patient's life (Vermeire et al. 2005; Olze et al. 2011; McRackan et al. 2018a, 2018b). However, the standard outcome measures used to assess this impact have primarily focused on word and sentence recognition under conditions that are not representative of typical communication environments and, therefore, correlate poorly with patients' self-reported communication abilities (Capretta et al. 2016; Moberly et al. 2018; McRackan et al. 2018a, 2018b). How cochlear implant (CI) users listen, communicate, and interact with their environment is far more complex than revealed by commonly used speech recognition tasks, even tasks that include background noise (McRackan et al. 2018a, 2018b). For example, most CI users rely on both auditory and visual cues for communication (Stevenson et al. 2017), and converse with both ears in the sound field with multiple communication partners in complex listening environments, factors that are not captured with our current test measures. In addition, current outcome measures do not assess the impact of CIs on the social, emotional, and functional aspects of CI users' lives. These gaps provide support for the development and use of a CI-specific quality of life (QOL) instrument that can comprehensively assess patients' experiences beyond speech recognition abilities.

Results of several studies and meta-analyses have revealed the significant impact of hearing loss on QOL (Chia et al. 2007) and the improvement that results after cochlear implantation (McRackan et al. 2018a, 2018b). QOL is typically composed of several constructs or domains that can differ across QOL instruments. For example, following implantation, adult CI recipients may have improved social function (Hinderink et al. 2000; Vermeire et al. 2005; Olze et al. 2011; Chung et al. 2012), better emotional well-being or mental health (Vermeire et al. 2005; Looi et al. 2011; Olze et al. 2011; Kobosko et al. 2015), and decreased listening effort, in addition to improved communication (Hughes et al. 2018). However, the degree of domain-specific QOL improvement varies greatly depending on the instrument used to measure these outcomes (McRackan et al. 2018a, 2018b) with some showing no improvement in social (Damen et al. 2007; Klop et al. 2008; Arnoldner et al. 2014) or emotional function (Damen et al. 2007; Klop et al. 2008; Arnoldner et al. 2014).

Consistent with these results, evaluation of health-related QOL through patient-reported outcomes measures (PROMs) has been increasingly emphasized by the Food and Drug Administration (Patrick et al. 2007), Centers for Medicare and Medicaid Services, and specifically in cochlear implantation by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health. However, no QOL instruments have been developed for the adult CI population using rigorous, psychometrically sound methodologies. Rather,

<zdoi; 10.1097/AUD.0000000000000684>

the most common instruments used with CI recipients are hearing-specific instruments that have not been validated for use by these patients. These include the Hearing Handicap Inventory for Adults/Elderly (HHIA/E) (Ventry & Weinstein 1982; Newman et al. 1990), Speech, Spatial and Quality of Hearing Scale (SSQ) (Gatehouse and Noble 2004), and the Abbreviated Profile of Hearing Aid Benefit (Cox and Alexander 1995), which were developed and validated primarily for individuals who have mild to moderate hearing loss and those who use hearing aids. As such, researchers and clinicians cannot be confident that these instruments accurately and reliably capture the constructs they purport to measure for adults with CI. Moreover, adults with CI may face unique barriers to QOL that may not be captured by instruments developed for and validated with other individuals.

In addition to hearing-specific PROMs, CI-specific PROMs have been used, especially the Nijmegen Cochlear Implant Questionnaire (NCIQ) (Hinderink et al. 2000). The NCIQ includes six QOL domains (basic sound perception, advanced sound perception, speech production, self-esteem, activity, and social interactions), each consisting of 10 items that were developed based on expert opinion. Initial validation and testing of the NCIQ was completed using a sample of 91 participants (including 46 controls) from a single clinical site (Hinderink et al. 2000). Although expert consensus was reached on the domains and items included in the NCIQ, the domains do not include certain QOL domains that CI users perceive to be important (McRackan et al. 2017; Hughes et al. 2018). This speaks to the value of directly soliciting input from CI users to ensure that a QOL instrument adequately captures concepts that are important to members of the target population. Moreover, since the NCIQ was established, more rigorous methods of psychometric testing during PROM development have become standard (Cella et al. 2007; Hays et al. 2007; Pilkonis et al. 2011). As such, there is a clear need to develop and validate CI-specific PROMs using modern psychometric techniques and recruitment of a large, representative sample from multiple clinical sites.

The National Institutes of Health established the Patient-Reported Outcomes Measurement Information System (PROMIS) in 2004 to develop, evaluate, and disseminate PROMs that focus on patient-centered outcomes. Since that time, PROMIS has established rigorous guidelines for how PROMs should be developed and validated (Fig. 1). This process includes five steps: (1) comprehensive literature search of existing measures (Klem et al. 2009; McRackan et al. 2018a, 2018b), (2) focus groups with thematic analysis of discussed topics to create a question or item pool (DeWalt et al. 2007; McRackan et al. 2017), (3) cognitive interviewing for feedback on language and item clarity (DeWalt et al. 2007), (4) confirmation of domain factor structure and item analysis using item response theory (IRT) to develop the item bank (Hays et al. 2007), and (5) validity testing of final QOL instruments (Pilkonis et al. 2014).

IRT is the modern standard for evaluation of items for inclusion in a PROM and offers several advantages over classical test theory, which was previously the standard for PROM development and used in the legacy hearing- and CI-specific QOL instruments (discussed later). Perhaps most importantly, instruments developed with IRT are considered to have psychometric properties that are sample and test independent (Prieto et al. 2003). Classical test theory is predicated on observed and true scores, which are sample dependent as subjects will have higher true scores on easy tests and lower true scores on more difficult tests. In contrast, IRT focuses on the measurement of an underlying latent trait, commonly referred to as person ability or person measure, which remains independent of test difficulty.
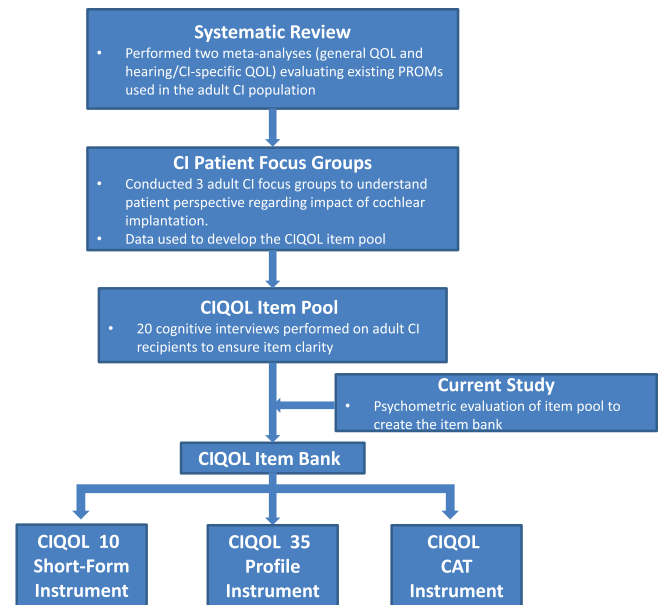

Figure 1. Flow diagram of steps involved in the development of a QOL instrument using the PROMIS guidelines and our progress to date. The current study establishes the CIQOL item bank, which will be used to develop subsequent instruments. CAT indicates computerized adaptive testing; CI, cochlear implant; PROMIS, Patient-Reported Outcomes Measurement Information System; QOL, quality of life

A second advantage of IRT is the focus on item-level, rather than test-level, psychometrics. IRT analyses provide data on each individual item to determine its characteristics and utility for inclusion in subsequent instruments. Through IRT, researchers can evaluate each item in the pool for ceiling and floor effects, identify fit to the model, match individual item difficulty level to person ability level, and ensure that the items cover the ability range of the population of interest. This analysis leads to the development of an item bank that measures and differentiates individuals across the range of ability levels. This item bank then serves as the source for items to be used for subsequent PROMs including short form, profile, and computerized adaptive testing (CAT) instruments. With the psychometric properties of each item established, selection of items for each instrument is based on highest discrimination across the ability range and best match between item difficulty and subject ability. The result is an optimized instrument with increased capacity to differentiate individuals across a greater range of the latent trait—termed precision (Rose et al. 2008).

Three core assumptions about the item bank must be met (Reeve et al. 2007): (1) items only contribute to one domain of QOL (unidimensionality), (2) responses to each item are unrelated to responses to other items (local independence), and (3) items fit the IRT measurement model (item fit). Confirmatory factor analysis (CFA) is used to confirm unidimensionality and local independence. Items are eliminated from the item pool if they do not substantially contribute to the unidimensional QOL domain captured by the other items, or if responses to the item are dependent upon responses to other items in the pool. In addition, indicators of item fit to the IRT model, such as infit

and outfit, are examined to ensure that the included items adequately measure the construct of interest for individuals at ability levels close to and far from the item difficulty.

Building on past work, the current manuscript represents the transition of our research from qualitative methodologies to quantitative analysis to evaluate our item pool using IRT (McRackan et al. 2017). The goal is to create an item bank that will be the source for items to be used in instruments that provide comprehensive and patient-centered evaluations of QOL in the adult CI population.

MATERIALS AND METHODS

Subjects

In order to enroll a large and diverse sample with respect to age at implantation, sex, CI listening modalities, and communication abilities, we established a Cochlear Implant Quality of Life Development Consortium, which includes 20 CI centers that represent all regions of the United States. Recruitment flyers were distributed to CI recipients electronically and on paper through these centers. Interested patients then emailed our research team to be enrolled. To be enrolled, patients must have been 18–89 years of age, a CI user for at least 1 year, and not have received a CI for single-sided deafness. This upper age limit was selected as individuals over the age of 89 are considered a special population by our institutional review board, requiring in-person consent, which was not possible given the online nature of the study's data collection. Data collection was performed using REDCap (Research Electronic Data Capture)—a secure, web-based data collection platform.

Sample size needs were determined based on CFA, which was the most sample size-dependent portion of the analysis. Sample sizes of 300 are considered conservative for CFA under a variety of sample conditions based on Monte Carlo simulations (MacCallum et al. 1999). We assumed a 60% response rate and therefore sent questionnaires to the first 500 subjects who contacted the research team.

Of the 500 patients emailed, 408 (81.6%) returned questionnaires. Subjects with missing data in item pool responses (n = 37, 7.4%) were excluded from analyses resulting in 371 (74.2%) subjects with complete data. The 371 subjects who completed the questionnaire included more females than males (Table 1). Most were married but did not have children <18 years of age living in the household. The majority lived in suburban environments, and essentially equal numbers lived in urban and rural locales. The vast majority had some education beyond a high school diploma and either worked full time or were retired. Subjects were fairly evenly split among the household income categories except in the lowest ($0–$20,000). All regions of the United States were represented with the South Atlantic region having the highest number of subjects (25.3%). Individuals from our institution represented only 2.9% of those who completed the questionnaire. Subjects represented the full range of age at implantation, duration of CI use, speech recognition abilities, and listening modalities of the adult CI population and used all three CI manufacturers' devices (Tables 2 and 3).

Data Collection

The questionnaire consisted of three sections: (1) subject demographics, (2) hearing and CI history (including speech recognition scores), and (3) the CIQOL item pool. Subject demographics are displayed in Table 1. US Census Bureau definitions

TABLE 1. Subject Demographics

	N (%)
Sex	
Male	149 (40.2)
Female	222 (59.8)
Marital status	
Single, never married	54 (14.6)
Married/domestic partnership	251 (67.7)
Widowed	24 (6.5)
Separated/divorced	42 (11.3)
Have children <18 in the home	
Yes	56 (15.1)
No	315 (84.9)
Environment where subject lives	
Urban	81 (21.8)
Suburban	214 (57.7)
Rural	76 (20.5)
Race	
Asian	3 (0.8)
Black or African American	3 (0.8)
White	351 (94.6)
More than one race	4 (1.1)
Not reported	10 (2.7)
Ethnicity	
Hispanic or Latino	13 (3.5)
Not Hispanic or Latino	300 (80.9)
Not reported	58 (15.6)
Combined household income	
$0–$20,000	26 (7.0)
$20,001–$50,000	63 (16.9)
$50,001–$80,000	87 (23.4)
$80,001–$110,000	66 (17.7)
>$110,000	93 (25.0)
Unknown/not reported	36 (9.7)
Highest level of education	
No schooling completed	0 (0)
Nursery school to 8th grade	1 (0.2)
Some high school, no diploma	2 (0.5)
High school graduate or equivalent	27 (7.2)
Some college	55 (14.8)
Trade/Tech/Vocational training	17 (4.5)
Associate degree	37 (9.9)
Bachelor's degree	112 (30.1)
Master's degree	75 (20.2)
Professional degree	18 (4.9)
Doctorate degree	27 (7.2)
Employment status	
Employed, working ≥40 hr per week	120 (32.3)
Employed, working <40 hr per week	40 (10.7)
Not employed, looking for work	8 (2.1)
Not employed, not looking for work	17 (4.5)
Retired	166 (44.7)
Disabled, not able to work	20 (5.3)
Region	
Northeast New England	17 (4.5)
Northeast Mid-Atlantic	30 (8.0)
Midwest East North Central	57 (15.3)
Midwest West North Central	33 (8.8)
South Atlantic	94 (25.3)
South East South Central	18 (4.8)
South West South Central	30 (8.0)
West Mountain	37 (9.9)
West Pacific	52 (14.0)
Not reported	3 (0.8)

TABLE 2. Subject Hearing and CI History

	Mean (SD)	Range
Age	59.5 (14.9)	19–88
Duration of hearing loss before CI (yr)	27.1 (18.4)	0–80
Duration of CI use (yr)	7.6 (6.5)	1.0–33.0
CNC Word scores (%; n = 173)	69.6 (24.4)	0–100
HINT Sentence scores in quiet (%, n = 78)	76.1 (30.2)	0–100
AzBio Sentences in quiet (%; n = 185)	81.2 (23.0)	0–100
AzBio Sentence scores in noise at +10 dB SNR (%; n=121)	64.3 (27.5)	0–100

Speech recognition scores were obtained by the subject from their audiologist. N represents the number of subjects who were able to obtain that score. AzBio: sentence recognition test performed in quiet or in noise at +10 dB SNR.
CI indicates cochlear implant; CNC, consonant-nucleus-consonant word recognition test performed in quiet; HINT, Hearing in Noise Test sentence recognition test performed in quiet; SNR, signal to noise ratio.

were used to define geographic region (Bureau 2010). Subjects were asked to self-identify the developed area where they lived as urban, suburban, or rural. The hearing and CI history collected are included in Tables 2 and 3. If subjects had bilateral CIs, duration of CI use was based on when the subject had their first CI activated. Subjects received their most recent best-aided speech recognition scores from their audiologist and entered them into the questionnaire. Consonant-nucleus-consonant words, Hearing in Noise Test sentences in quiet, and Arizona Biomedical (AzBio) sentences in quiet and in noise at a +10 dB signal to noise ratio were selected as these are part of the minimum standard test battery (Centers for Medicare and Medicaid Services: Quality Strategy 2011). Consonant-nucleus-consonant word test is an open-set 50 monosyllabic word recognition task performed in quiet (Peterson and Lehiste 1962). Hearing in Noise Test and AzBio are open-set sentence recognition tasks consisting of 10 and 20 sentences per list, respectively (Nilsson et al. 1994; Spahr et al. 2012). At least one of these speech recognition scores were available for 236 subjects (63.6%). Subjects were not excluded from analyses if they were unable to obtain their scores.

The development of the initial item pool for an adult CIQOL instrument has been previously described (McRackan et al. 2017) and follows the first and second steps in the PROMIS guidelines described earlier. Briefly, a systematic literature search was used to develop a protocol for three adult CI recipient focus groups. Participants in the three focus groups (n = 23) were representative

TABLE 3. Subject CI Device Information

	n (%)
CI company	
Advanced Bionics	43 (11.5)
Cochlear	216 (58.2)
MED-EL	112 (30.1)
Listening modality	
Unilateral CI with no contralateral HA	87 (23.4)
Unilateral CI with contralateral HA	96 (25.8)
Bilateral CI	188 (50.6)
Combined electro-acoustic hearing (Hybrid)	
No	358 (96.4)
Yes	12 (3.2)
No response	1 (0.2)

CI indicates cochlear implant.

of the adult CI population in terms of demographics, communication abilities, and listening modalities and were stratified based on speech recognition ability (McRackan et al. 2017). The development, execution, and analysis of the focus group protocol was based on grounded theory (Ralph et al. 2015), and the Consolidated Criteria for Reporting Qualitative Research (COREQ-32) was followed (Tong et al. 2007). In accordance with COREQ-32, sample size was determined based on data saturation. The 101-item pool was created based on the central and minor themes identified from transcript coding of our CI patient focus groups (McRackan et al. 2017). The majority of items were derived from direct quotes from focus group participants, whereas others were synthesized from similar comments made by multiple focus group participants. The items were reviewed for content validity and clarity by three fellowship-trained neurotologists who routinely perform CI surgeries, two adult CI audiologists, a PhD public health researcher with expertise in community engagement research, a PhD hearing research scientist with expertise in adult hearing loss, and a PhD psychometrician. Afterward, cognitive interviews were performed with 20 additional adults with CIs in order to ensure the clarity of the 101-item pool (McRackan et al. 2017). Item response options used one of the scales recommended by PROMIS: Never, Rarely, Sometimes, Often, Always.

Based on these results, the initial item pool was separated into 7 hypothesized domain constructs, including communication (receptive and expressive communication ability in different situations), emotional (impact of hearing ability on emotional well-being), entertainment (enjoyment and clarity of TV, radio, music, etc.), environmental (ability to distinguish and localize environmental sounds), independence (ability to function without assistance from others), listening effort (degree of effort and resulting fatigue associated with listening), and social (ability to interact in groups and to attend and enjoy social functions). Figure 2 displays the number of items per construct.

Data Analysis

Descriptive statistics were used to describe the subject demographics. CFA was used to determine the degree to which items in each domain represented a single unidimensional construct. Item residual correlations were used to determine if the responses to each item were unrelated to responses to other items (local independence). The assumption of unidimensionality for each domain was analyzed with an ordered-category CFA with diagonal weighted least squares estimation using the package "lavaan" in the statistical software R (Rosseel et al. 2017). It is best practice in CFA to examine multiple types of fit indicators, including those who are reflective of absolute fit (standardized root mean square residual <0.08), those who have parsimony corrections (root mean square error of approximation <0.08), and comparative fit indicators (comparative fit index >0.95; Tucker-Lewis index >0.95). Acceptable CFA model fit was defined a priori by standardized root mean square residual, root mean square error of approximation, comparative fit index, and Tucker-Lewis index (Brown 2015). In addition, standardized item factor loadings were examined. A minimum factor loading of 0.32 was chosen as the level of significance, as this equates to approximately 10% overlapping variance with other items in that factor (Tabachnick & Fidell 2013). Item residual correlations were examined for local dependence, with correlations >0.2 indicating dependence (Pilkonis et al. 2011).
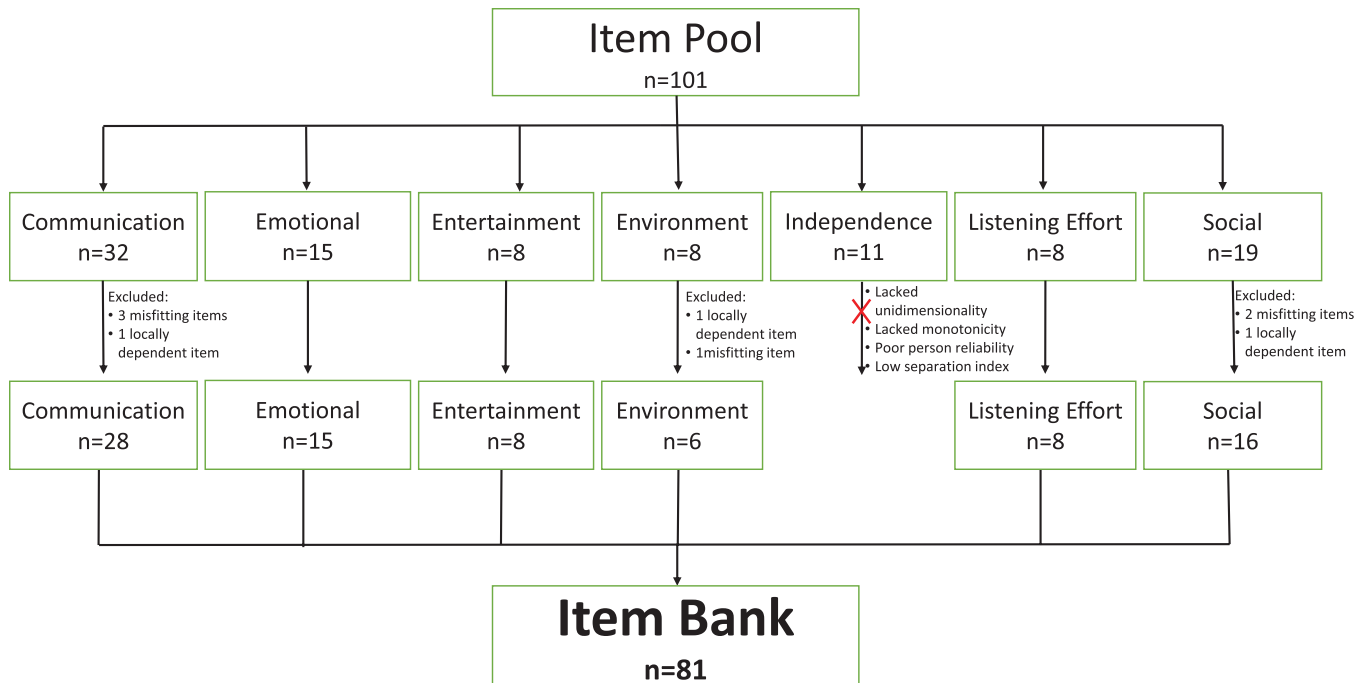
Figure 2. Flowchart showing the results of psychometric analysis of the item pool to develop the final item bank. The X indicates that the independence domain was not included in the final item bank.

Once a unidimensional set of items for each domain was identified, a one-parameter logistic IRT analysis was conducted. A rating scale model with joint maximum likelihood estimation was conducted using WINSTEPS, version 3.90.0 (Linacre 2016b). Results of the IRT analysis were examined using a multi-step approach. First, the appropriateness of the rating scale was evaluated using the following criteria (Linacre 2002): (1) at least 10 observations of each category, collapsed across all items; (2) monotonicity of rating scale categories (i.e., 0–4) as evidenced by an increase in average category difficulty with increasing category value; and (3) outfit mean square is <2.0. Second, the fit of the items and persons to the IRT model was evaluated by examining infit and outfit mean squares and standardized z values (Linacre 2002). Mean square values >1.70, as well as standardized z values greater than 2.0, were considered indicative of misfit to the IRT model (Wright et al. 1994). Third, reliability indicators were examined including (1) person reliability, which represents the reproducibility of person ordering and was interpreted such that values ≥0.5 were considered adequate, ≥0.80 were considered good, and ≥0.90 were considered high (Linacre 2016a) and (2) the separation index was used to calculate the number of statistically distinct ability strata in the sample (Wright & Masters 2002). The number of person strata is calculated according to the formula $\frac{(4G+1)}{3}$, where G is the person separation index and is an indicator of the number of statistically distinct person measures with centers three calibration errors apart. Test targeting, test coverage, and item hierarchy were examined visually using person-item maps.

RESULTS

Confirmatory Factor Analysis

A summary of CFA results used to evaluate unidimensionality and item local independence is provided in Table 4. Fit indices reflected adequate-to-good model fit for the communication, emotional, and social domains. Root mean square error of approximation indicated poor model fit for the entertainment, environment, independence, and listening effort domains. However, as these domains had fewer degrees of freedom due to smaller item pools, we focused on other indicators of model fit for these domains as evidence suggests that root mean square error of approximation may not be a reliable indicator of fit in these cases (Kenny et al. 2015). Remaining fit indices reflected good model fit for the entertainment, environment, and listening effort domains. However, standardized root mean square revealed poor model fit for the independence domain.

All items had standardized factor loadings of ≥0.32 on their respective domains. Examination of residual correlation matrices revealed no local dependence for the emotional, entertainment, independence, and listening effort domains. The communication domain had two items that were locally dependent upon one another (residual correlation >0.2). These items were "I have to ask people to look at me when they speak" and "I have to ask a lot of questions about what is being said in a conversation." The environment and social domains each had three items that demonstrated local dependence. For the environmental domain, the item "Noises from household appliances are bothersome" demonstrated dependence upon the items "Noises in my car are bothersome" and "I am able to hear cars approaching in traffic." In the social domain, the item "I am able to communicate with my family and friends" demonstrated dependence with the items "I avoid socializing with friends, relatives, or neighbors due to my hearing loss" and "I avoid social situations due to my hearing loss." For each of these items, the one item was locally dependent with two other items, but the two other items were not dependent upon each other. Therefore, the one item that displayed local dependence from each domain was excluded from subsequent analyses.

TABLE 4. Indicators of Unidimensionality and Item Local Independence

	Communication (n = 32)	Emotional (n = 15)	Entertainment (n = 8)	Environment (n = 8)	Independence (n = 11)	Listening Effort (n = 8)	Social (n = 19)
RMSEA	0.09	0.06	*0.17*	*0.14*	*0.20*	*0.15*	0.07
SRMR	0.06	0.05	0.08	0.08	*0.12*	0.07	0.05
CFI	0.99	0.99	0.99	0.98	0.99	0.98	0.99
TLI	0.99	0.99	0.98	0.96	0.99	0.98	0.99
Locally dependent items	2	0	0	3	0	0	3

n indicates the number of items in each domain. Better fit is denoted by lower values for RMSEA and SRMR, but higher values for CFI and TLI. Fit indices in italics indicate poor model fit.
CFI indicates comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker-Lewis index.

Item Response Theory Analysis

A summary of the IRT results are provided in Table 5. The rating scales of all domains met our three *a priori* criteria, with the exception of the independence domain. The rating scale for the independence domain did not demonstrate monotonicity or outfit mean square <2.0, indicating poor performance of the rating scale. All items of the emotional, entertainment, and listening effort domains fit the IRT model, whereas the communication, independence, social, and environment domains each had at least one misfitting item. The misfitting items were excluded from the final item bank.

The number of subjects who misfit the model ranged from n = 22 (5.9%; independence) to n = 74 (19.9%; communication) across the domains. After removing the three misfitting items from the communication domain, the number of misfitting subjects decreased to 46 (12.4%). The mean person ability was greater than two logits from the mean item difficulty for the independence domain, indicating a poor match between item difficulty and person ability. Separation indices ranged from 0.63 (independence domain) to 4.51 (communication domain) with a larger number representing greater capacity to separate individuals. Minimal ceiling and floor effects were observed for all domains except the independence domain, which showed a 36.1% ceiling effect. All domains demonstrated strong person reliability (>0.80) with the exception of the independence domain (0.29). The number of person strata, which represents the number of statistically distinct ability levels that the items can differentiate, ranged from 1.17 (independence domain) to 6.35 (communication domain).

In summary, the independence domain was not included in the final item bank because it lacked unidimensionality, monotonicity, had poor person reliability and low separation index. After removal of the locally dependent and misfitting items, all other domains and remaining items were included in the final item bank. In total, 9 other items were removed from other domains leaving 81 items in the final item bank and 6 domains (Fig. 2). The full CIQOL item bank is provided as supplemental digital content, http://links.lww.com/EANDH/A497 (online Appendix 1).

DISCUSSION

Although many hearing- and CI-specific PROMs have been developed, the current research differs in that it followed stringent and widely accepted guidelines and applied rigorous psychometric methods. This combined use of qualitative and quantitative approaches provides hypothesis-driven methods to develop PROMs. Based on the results of the CI patient focus groups (McRackan et al. 2017), we hypothesized domain constructs that encompass QOL for CI recipients. In addition, given that actual CI patients (rather than expert panels of providers) contributed to item development, we hypothesized that the item pool would appropriately measure the ability range of adult CI patients, providing content validity.

Several domains of the legacy QOL instruments used in the adult CI population align with our results. For example, similar to the CIQOL item bank, the NCIQ and HHIA/E include domains related to social/social interaction and emotional/psychological function. Other domains are unique to the CIQOL item bank, including entertainment, environment, and listening effort. Thus, using focus groups rather than expert panels to create the item bank may uncover topics that have been previously unknown or ignored. Although the SSQ is typically separated into speech, spatial, and qualities of hearing domains, there is some evidence for an additional listening effort domain, which aligns with our findings. However, it is important to note that the listening effort domain did not meet all criteria for a

TABLE 5. Summary of IRT Results

	Communication	Emotional	Entertainment	Environment	Independence	Listening Effort	Social
No. Items	31	15	8	7	11	8	18
Misfitting items	3	0	0	1	2	0	2
Misfitting persons (%)	46 (12.4%)*	34 (9.2%)	26 (7.0%)	23 (6.2%)	22 (5.9%)	30 (8.1%)	39 (10.5%)
Subjects reaching ceiling (%)	0 (0%)	6 (1.6%)	5 (1.3%)	3 (0.8%)	134 (36.1%)	0 (0%)	5 (1.3%)
Subjects reaching floor (%)	0 (0%)	0 (0%)	3 (0.8%)	0 (0%)	1 (0.3%)	0 (0%)	0 (0%)
Mean Person ability (logits)	0.73	1.64	0.82	1.32	2.39	–0.17	1.62
Person separation	4.51	3.79	2.71	2.16	0.63	2.79	3.36
Person reliability	0.95	0.93	0.88	0.82	0.29	0.89	0.92
No. person Strata	6.35	5.39	3.94	3.21	1.17	4.05	4.81
Cronbach α	0.96	0.95	0.91	0.86	1.00	0.90	0.95

*The value presented represents the number of misfitting persons after the misfitting items were removed.
IRT indicates item response theory.

unidimensional construct in a post hoc analysis of the SSQ, and this domain, along with the entire SSQ, has not been validated in the CI population (Akeroyd et al. 2014).

The relationship between the CIQOL communication domain and the NCIQ is more complex. CFA results from the current study clearly demonstrate that the communication domain represents a unidimensional construct. In contrast, the NCIQ separates communication into three domains: basic sound perception, advanced sound perception, and speech production, but a comparable CFA has not been performed. In addition, IRT analyses provide the difficulty level for each item, which can be used to develop subsequent PROMs. As discussed later, quantitative analysis determined whether an item is "basic" versus "advanced," rather than relying on the assumptions of expert panels.

The importance of the methodology used to develop the CIQOL can be contrasted to existing hearing- and CI-specific PROMs—using classical test theory or by simply evaluating the test–retest reliability in samples of patients with hearing. A critical step in developing a new QOL instrument is to establish a clear QOL construct to conceptualize the values of the affected population. As noted earlier, this is accomplished using results from patient focus groups, which provide a means for patients to participate directly in the item development process, rather than relying on expert panels (Hays et al. 2007; Velozo et al. 2012). This ensures content validity of the items used in the PROMs and allows interpretation of results within a meaningful QOL framework.

The manner in which domains were developed and evaluated is another advantage of the current methods. Here, thematic analysis of the focus group transcripts was used to create hypothesized constructs that compose QOL. CFA was then applied to ensure that the items in each domain contribute to a single construct (unidimensionality). This quantitative approach to confirm that each domain is measuring a unique construct has rarely been applied in our field. Rather, domains have been traditionally selected based on the research team's or expert panel's opinions and accepted without thorough analysis or input from the affected population (Hinderink et al. 2000), as described earlier for the NCIQ. The application of CFA in this manner allows researchers to have greater certainty of what is actually being measured within each domain.

The importance of this methodology is highlighted in the results of the current study where the independence domain did not meet criteria for inclusion in the final item bank. Focus group participants identified independence as an important QOL theme as it relates to their CI. However, our analysis showed that the items in this domain were important, but they represented more than one distinct QOL construct (multidimensional), did not show monotonicity, poorly stratified subjects with regard to ability, and had a ceiling effect for 36.1% of subjects. In contrast, the emotion domain demonstrated the next highest ceiling effect, where 1.6% of subjects had maximum scores.

In addition, previous hearing- and CI-specific PROMs have relied on Cronbach α as a marker of internal consistency and reliability (Newman et al. 1990; Hinderink et al. 2000), although it is known to over-estimate both (Sijtsma 2009). Here, the importance of more in-depth analysis is seen as Cronbach α of the independence domain was 1.0, but person reliability was well below acceptable standards (0.29). One may argue that these data suggest that items in this domain should be changed to be made more difficult to provide a better fit with the CI population. However, these items were developed from patient focus group responses on how their implants have impacted their lives. To add more difficult items to this domain as a means to improve psychometric qualities would sacrifice content validity, resulting in an instrument that is not relevant to the population of interest.

Rather than relying on traditional methods (such as classical test theory and Cronbach α), the application of CFA and IRT in this analysis provides a means to test these hypotheses at the item and group levels to create an item bank that represents and stratifies adult CI patients with regard to QOL. This analysis resulted in six domains with 81 psychometrically sound items in the item bank. Our findings show that each of these domains represents a single construct, which has not been determined for previous hearing- and CI-specific PROMs. This aids in interpretation of results because it is understood that each domain represents a singular latent trait. In addition, we know that the item bank has the capacity to measure the full range of functional abilities within the adult CI population, which has also not been established in legacy hearing- and CI-specific PROMs, such as the HHIA/E, SSQ, Abbreviated Profile of Hearing Aid Benefit, and NCIQ. With the psychometric properties of the item bank, domain constructs, and individual items known, these data can then be used to develop a suite of new QOL instruments.

Future Directions

The analysis resulted in a psychometrically sound CIQOL item bank with domains and items removed that did not meet accepted standards. We will use the data from this psychometric analysis to select the optimal items for new short-form and profile CIQOL instruments and to develop a CAT. Short-form PROMs provide a global evaluation of a CI user's QOL and are important for routine use in the busy clinical setting where clinicians and patients may not have sufficient time to complete longer instruments. In addition, short-form instruments are ideal for inclusion in research protocols to minimize the effect of questionnaire fatigue when multiple PROMs are used (Porter et al. 2004). In contrast, profile instruments are longer but expand upon the short-form evaluation and provide additional domain-specific QOL data. CAT, the most advanced and efficient method to administer PROMs, incorporates real-time IRT to select subsequent items based on a patient's responses to previous items. As such, items presented are dynamically selected for an individual patient based on their ability. Relative to short forms, CAT minimizes floor and ceiling effects and allows greater differentiation among individual patients relative to static instruments. Moreover, because items are individually selected based on ability, differentiating patient responses can be accomplished using a minimal number of items, thus reducing patient and practitioner burden (Fries et al. 2014). The information obtained through the psychometric evaluation of the item pool to create the CIQOL item bank will guide the selection of items for short-form and profile instruments and development of the CIQOL CAT.

Study Limitations

The study's limitations are inherent to the online design. Overall, the study population tended to be relatively well educated and have high socioeconomic status. Nevertheless, benefits of enrolling a large population from multiple CI centers (representing diverse geographic locations), which is required

for CFA and IRT analyses, outweighed this limitation. In addition, the items were developed using focus groups that were more racially diverse (McRackan et al. 2017). The current study's sample does not match the ethnic and racial demographics of the general public. However, no data are available regarding ethnic and racial demographics of the US adult CI population. Even large studies of CI utilization (Sorkin 2013; Agabigum et al. 2018) and CI outcomes (Gifford et al. 2008; Chung et al. 2012; Holden et al. 2013; Reeder et al. 2014; Wanna et al. 2014; Angelo et al. 2016) routinely fail to report race and ethnicity data. The lack of racial/ethnic data for the adult CI population makes it difficult to determine the effects of this study's demographics on our results. Although lack of racial/ethnic diversity was a potential limitation, a diverse sample of subjects was enrolled with respect to household income, living environments, employment status, age at implantation, duration of hearing loss, duration of CI use, device types, and CI listening modality.

An additional limitation was speech recognition measures, which were not available for all subjects (unavailable for 36.4%). In addition, because subjects were recruited from a large number of institutions, the specific conditions in which speech recognition tasks were performed could not be controlled. Nevertheless, the results reported revealed that subjects' speech recognition outcomes were consistent with published data and represented full range of speech recognition abilities (Gifford et al. 2008). Given the known weak correlation between speech recognition ability and self-reported QOL in the CI population (Capretta & Moberly 2016; McRackan et al. 2018a, 2018b; Moberly et al. 2018), recruiting an adequate sample size for robust psychometric analyses was more important than obtaining speech recognition data for all subjects. However, determining how speech recognition ability, patient demographics, and hearing/CI history impact CIQOL is important and will be addressed in future studies.

CONCLUSIONS

This study builds on our prior work in the development of adult CIQOL PROMs using rigorous established guidelines. We have completed the systematic review, patient focus groups, cognitive interviews, and psychometric analyses to create a CIQOL item bank, toward the overall goal of developing a suite of CIQOL instruments. The value of using these methodologies is seen through the elimination of a domain and items that were not psychometrically valid, ensuring that the items included in the bank covered the ability range of the adult CI population. We also determined each item's difficulty level, which will be used to optimally select items for other CIQOL instruments. Our future work will develop and validate a suite of short-form, profile, and CAT-based CIQOL instruments and compare results using these new instruments to legacy CIQOL instruments and functional outcome measures. Following additional research, these CIQOL instruments are anticipated to have significant impact on patient outcomes by revealing the communication, emotional, and social benefits of CIs in adults, which will provide tools to determine CI candidacy, set appropriate expectations, identify domain-specific therapies, and determine how device technologies, listening modalities, and novel processing strategies impact CIQOL outcomes.

REFERENCES

Agabigum, B., Mir, A., Arianpour, K., et al. (2018). Evolving trends in cochlear implantation: A critical look at the older population. *Otol Neurotol*, *39*, e660–e664.

Akeroyd, M. A., Guy, F. H., Harrison, D. L., et al. (2014). A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale). *Int J Audiol*, *53*, 101–114.

Angelo, T. C., Moret, A. L., Costa, O. A., et al. (2016). Quality of life in adult cochlear implant users. *Codas*, *28*, 106–112.

Arnoldner, C., Lin, V. Y., Bresler, R., et al. (2014). Quality of life in cochlear implantees: Comparing utility values obtained through the Medical Outcome Study Short-Form Survey-6D and the Health Utility Index Mark 3. *Laryngoscope*, *124*, 2586–2590.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, United Kingdom: The Guilford Press.

Bureau, U. S. C. (2010). Geographic Terms and Concepts - Census Divisions and Census Region [Online]. Verified November 2014 from https://www.census.gov/geo/reference/gtc/gtc_census_divreg.html. Retrieved January 10, 2018.

Capretta, N. R., & Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *Laryngoscope*, *126*, 699–706.

Cella, D., Yount, S., Rothrock, N., et al; PROMIS Cooperative Group. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Med Care*, *45*(5 Suppl 1), S3–S11.

Centers for Medicare and Medicaid Services: Quality Strategy. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf. Retrieved January 3, 2018.

Chia, E. M., Wang, J. J., Rochtchina, E., et al. (2007). Hearing impairment and health-related quality of life: The Blue Mountains Hearing Study. *Ear Hear*, *28*, 187–195.

Chung, J., Chueng, K., Shipp, D., et al. (2012). Unilateral multi-channel cochlear implantation results in significant improvement in quality of life. *Otol Neurotol*, 33, 566–571.

Cox, R. M., & Alexander, G. C. (1995). The Abbreviated Profile of Hearing Aid Benefit. *Ear Hear*, 16, 176–186.

Damen, G. W., Beynon, A. J., Krabbe, P. F., et al. (2007). Cochlear implantation and quality of life in postlingually deaf adults: Long-term follow-up. *Otolaryngol Head Neck Surg*, 136, 597–604.

DeWalt, D. A., Rothrock, N., Yount, S., et al; PROMIS Cooperative Group. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Med Care*, 45(5 Suppl 1), S12–S21.

Fries, J. F., Witter, J., Rose, M., et al. (2014). Item response theory, computerized adaptive testing, and PROMIS: Assessment of physical function. *J Rheumatol*, 41, 153–158.

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol*, 43, 85–99.

Gifford, R. H., Shallop, J. K., Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiol Neurootol*, 13, 193–205.

Hays, R. D., Liu, H., Spritzer, K., et al. (2007). Item response theory analyses of physical functioning items in the medical outcomes study. *Med Care*, 45(5 Suppl 1), S32–S38.

Hinderink, J. B., Krabbe, P. F., Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The Nijmegen Cochlear Implant Questionnaire. *Otolaryngol Head Neck Surg*, 123, 756–765.

Holden, L. K., Finley, C. C., Firszt, J. B., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear*, 34, 342–360.

Hughes, S. E., Hutchings, H. A., Rapport, F. L., et al. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: A grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear*, 39, 922–934.

Kenny, D. A., Kaniskan, B., McCoach D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociol Methods Res, 44*, 486–507.

Klem, M., Saghafi, E., Abromitis, R., et al. (2009). Building PROMIS item banks: librarians as co-investigators. *Qual Life Res*, 18, 881–888.

Klop, W. M., Boermans, P. P., Ferrier, M. B., et al. (2008). Clinical relevance of quality of life outcome in cochlear implantation in postlingually deafened adults. *Otol Neurotol*, 29, 615–621.

Kobosko, J., Jedrzejczak, W. W., Pilka, E., et al. (2015). Satisfaction with cochlear implants in postlingually deaf adults and its nonaudiological predictors: Psychological distress, coping strategies, and self-esteem. *Ear Hear*, 36, 605–618.

Linacre, J. (2016a). Rasch Measurement Computer Program: User's Guide [Online]. Verified November 22, 2017 from http://www.winsteps.com/winman/principalcomponents.htm. Retrieved December 10, 2018.

Linacre, J. (2016b). Rasch Measurement Computer Program [Online]. Verified November 14, 2017 from http://www.winsteps.com/. Retrieved December 4, 2018.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *J Appl Meas*, 3, 85–106.

Looi, V., Mackenzie, M., Bird, P., et al. (2011). Quality-of-life outcomes for adult cochlear implant recipients in New Zealand. *N Z Med J*, 124, 21–34.

MacCallum, R. C., Widaman, K. F., Zhang, S., et al. (1999). Sample size in factor analysis. *Psychol Methods*, 4, 84–99.

McRackan, T. R., Bauschard, M., Hatch, J. L., et al. (2018a). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope*, 128, 982–990.

McRackan, T. R., Bauschard, M., Hatch, J. L., et al. (2018b). Meta-analysis of cochlear implantation outcomes evaluated with general health-related patient-reported outcome measures. *Otol Neurotol*, 39, 29–36.

McRackan, T. R., Velozo, C. A., Holcomb, M. A., et al. (2017). Use of adult patient focus groups to develop the initial item bank for a cochlear implant quality-of-life instrument. *JAMA Otolaryngol Head Neck Surg*, 143, 975–982.

Moberly, A. C., Harris, M. S., Boyce, L., et al. (2018). Relating quality of life to outcomes and predictors in adult cochlear implant users: Are we measuring the right things? *Laryngoscope*, 128, 959–966.

MSTB: The new minimum speech test battery. http://auditorypotential.com/MSTB.html. Retrieved March 27, 2018.

Newman, C. W., Weinstein, B. E., Jacobson, G. P., et al. (1990). The Hearing Handicap Inventory for Adults: psychometric adequacy and audiometric correlates. *Ear Hear*, 11, 430–433.

Nilsson, M., Soli, S. D., Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*, 95, 1085–1099.

Olze, H., Szczepek, A. J., Haupt, H., et al. (2011). Cochlear implantation has a positive influence on quality of life, tinnitus, and psychological comorbidity. *Laryngoscope*, 121, 2220–2227.

Patrick, D. L., Burke, L. B., Powers, J. H., et al. (2007). Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health*, 10(Suppl 2), S125–S137.

Peterson, G. E., & Lehiste, I. (1962). Revised CNC lists for auditory tests. *J Speech Hear Disord*, 27, 62–70.

Pilkonis, P. A., Choi, S. W., Reise, S. P., et al; PROMIS Cooperative Group. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment*, 18, 263–283.

Pilkonis, P. A., Yu, L., Dodds, N. E., et al. (2014). Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *J Psychiatr Res*, 56, 112–119.

Porter, S. R., Whitcomb, M. E., Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *NDIR. 2004*, 63–73.

Prieto, L., Alonso, J., Lamarca, R. (2003). Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes*, 1, 27.

ROMIS: Instrument Development and Validation Scientific Standards. http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf. Retrieved January 10, 2018.

Ralph, N., Birks, M., Chapman, Y. (2015). The methodological dynamism of grounded theory. *Int J Qual Methods*. 14, 4.

Reeder, R. M., Firszt, J. B., Holden, L. K., et al. (2014). A longitudinal study in adults with sequential bilateral cochlear implants: Time course for individual ear and bilateral performance. *J Speech Lang Hear Res*, 57, 1108–1126.

Reeve, B. B., Hays, R. D., Bjorner, J. B., et al; PROMIS Cooperative Group. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*, 45(5 Suppl 1), S22–S31.

Rose, M., Bjorner, J. B., Becker, J., et al. (2008). Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*, 61, 17–33.

Rosseel, Y., Byrnes, J., Vanbrabant, L., et al. (2017). Package 'lavaan' [Online]. Verified November 12, 2017 from https://cran.r-project.org/web/packages/lavaan/lavaan.pdf. Retrieved November 20, 2018.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107–120.

Sorkin, D. L. (2013). Cochlear implantation in the world's largest medical device market: Utilization and awareness of cochlear implants in the United States. *Cochlear Implants Int*, 14(Suppl 1), S4–S12.

Spahr, A. J., Dorman, M. F., Litvak, L. M., et al. (2012). Development and validation of the AzBio sentence lists. *Ear Hear*, 33, 112–117.

Stevenson, R. A., Sheffield, S. W., Butera, I. M., et al. (2017). Multisensory integration in Cochlear implant recipients. *Ear Hear*, 38, 521–538.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Boston, MA: Pearson Education.

Tong, A., Sainsbury, P., Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *Int J Qual Health Care*, 19, 349–357.

Velozo, C. A., Seel, R. T., Magasi, S., et al. (2012). Improving measurement methods in rehabilitation: Core concepts and recommendations for scale development. *Arch Phys Med Rehabil*, 93(Suppl 8), S154–S163.

Ventry, I. M., & Weinstein, B. E. (1982). The hearing handicap inventory for the elderly: A new tool. *Ear Hear*, 3, 128–134.

Vermeire, K., Brokx, J. P., Wuyts, F. L., et al. (2005). Quality-of-life benefit from cochlear implantation in the elderly. *Otol Neurotol*, 26, 188–195.

Wanna, G. B., Noble, J. H., Carlson, M. L., et al. (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *Laryngoscope*, 124(Suppl 6), S1–S7.

Wright, B., Linacre, J., Gustafson, J., et al. (1994). Reasonable mean-square fit values. In: *Rasch Measurement Transactions 8*, 370.

Wright, B., & Masters, G. N. (2002). Number of Person or Item Strata. In: *Rasch Measurement Transactions 16*, 888.