Published in final edited form as:

J Speech Lang Hear Res. 2009 October ; 52(5): 1302–1320. doi:10.1044/1092-4388(2009/07-0275).

Developing the Communicative Participation Item Bank: Rasch Analysis Results From a Spasmodic Dysphonia Sample

Carolyn R. Baylor, Kathryn M. Yorkston, Tanya L. Eadie, Robert M. Miller, and Dagmar Amtmann

University of Washington, Seattle

Abstract

Purpose—The purpose of this study was to conduct the initial psychometric analyses of the Communicative Participation Item Bank—a new self-report instrument designed to measure the extent to which communication disorders interfere with communicative participation. This item bank is intended for community-dwelling adults across a range of communication disorders.

Method—A set of 141 candidate items was administered to 208 adults with spasmodic dysphonia. Participants rated the extent to which their condition interfered with participation in various speaking communication situations. Questionnaires were administered online or in a paper version per participant preference. Participants also completed the Voice Handicap Index (B. H. Jacobson et al., 1997) and a demographic questionnaire. Rasch analyses were conducted using Winsteps software (J. M. Linacre, 1991).

Results—The results show that items functioned better when the 5-category response format was recoded to a 4-category format. After removing 8 items that did not fit the Rasch model, the remaining 133 items demonstrated strong evidence of sufficient unidimensionality, with the model accounting for 89.3% of variance. Item location values ranged from -2.73 to 2.20 logits.

Conclusions—Preliminary Rasch analyses of the Communicative Participation Item Bank show strong psychometric properties. Further testing in populations with other communication disorders is needed.

Keywords

Rasch analyses; participation; spasmodic dysphonia

Participation in most life roles requires communicating with other people. Whether individuals are at work, taking care of their families and households, or relaxing with friends, communication is a key component to achieving what they need and want to do. The term *communicative participation* refers to the communication aspects of people's involvement in their life roles. Communicative participation is defined as "taking part in life situations where knowledge, information, ideas or feelings are exchanged" (Eadie et al., 2006, p. 309; also see Yorkston et al., 2008). When communication disorders interfere with participation in life roles, many negative consequences may follow, such as loss of employment, social isolation, and difficulty pursuing services, including health care. Understanding communicative participation is critical for understanding how well people

with communication disorders meet their communication needs in their daily lives and for documenting how intervention helps people to better meet their communication needs.

Communicative participation is a critical component of a biopsychosocial approach to communication disorders. Biopsychosocial frameworks—most notably the World Health Organization's (2001) International Classification of Functioning, Disability, and Health—draw attention to the multifactorial contributors to the consequences of health conditions, including communication disorders (Eadie, 2001, 2003; Threats, 2002). One of the consequences of framing our work from biopsychosocial perspectives is that we have to consider more complex hypotheses regarding the consequences of communication disorders. For example, how do various biological, personal, and social–environmental factors mediate the impact of communication disorders on participation in and on fulfillment of life roles? Exploration of these hypotheses requires measurement tools specifically designed for each construct included in the question.

The speech-language pathology field has traditionally focused more on the physical impairments underlying communication disorders, as well as on the ability of clients to perform basic speech and communication tasks, than on the consequences of communication disorders in daily life (Eadie et al., 2006; Threats, 2002; Worrall, McCooley, Davidson, Larkins, & Hickson, 2002). Outcome measurement tools that address physiologic function and/or isolated communication activities (i.e., acoustic measures of voice, accuracy of word production, and calculation of speech intelligibility) are critical to understanding the impact of intervention on function of the speech mechanism and the capacity of the individual to perform speech tasks. These measures do not, however, capture the translation of these activities into real-life situations. Participation must be measured directly and not inferred from the degree of physical impairment or performance of basic skills (Bickenbach, Chatterji, Badley, & Ustun, 1999; Cardol et al., 1999).

More recently, several questionnaires designed to measure the psychosocial impact of speech disorders, particularly voice disorders, have been implemented in research and clinical practice. These include the Voice Handicap Index (VHI; Jacobson et al., 1997), the Voice-Related Quality of Life Scale (Hogikyan & Sethuraman, 1999), the Voice Activity and Participation Profile (Ma & Yiu, 2001), and the Voice Symptoms Scale (Deary, Wilson, Carding, & MacKenzie, 2003). Although these tools have greatly advanced our understanding of the psychosocial consequences of voice disorders, none of these can function as a tool to specifically measure communicative participation because a variety of constructs are represented in each questionnaire (Eadie et al., 2006). Examples of these different constructs include physical symptoms, performance of basic speech tasks, and personal emotional coping. Although these questionnaires reflect the multifaceted nature of disability associated with voice disorders, these multidimensional instruments cannot be used effectively to evaluate participation without the confounding influence of other constructs. For example, one important question to ask about communicative participation would be the following: To what extent is communicative participation affected by severity of speech symptoms versus social obstacles, such as lack of social support? The answer to this question would provide meaningful information about how and where to invest intervention resources. Answering this question would require comparison of data from three instruments, each intended to measure a separate construct of communicative participation, speech symptoms, and social support, respectively. When all three constructs are mixed together in the summary score from one questionnaire, such comparisons are not possible. Dividing existing questionnaires into subscales is an option, but there is often disagreement on subscale structures. For example, one-factor (Rosen, Lee, Osborne, Zullo, & Murry, 2004; Wilson et al., 2004), two-factor (Bogaardt, Hakkesteeft, Grolman, & Lindeboom, 2007), and three-factor (Jacobson et al., 1997) structures have all been proposed

for the most common psychosocial voice questionnaire: the VHI (Jacobson et al., 1997). As the field of speech-language pathology moves toward investigating the complex relationships among biopsychosocial factors associated with communication disorders, it is critical that psychometrically sound, interpretable, and clinically meaningful instruments be developed to measure single constructs.

There are many possible approaches to measuring participation, including assessing what people do when they participate, how often they do it, how well they do it, and how satisfied they are with their participation. Although each of these options provides useful information about participation, one of the most critical measures of successful participation is the satisfaction of the individual with his or her own participation—not an external normative standard about what participation “should” be (Brown et al., 2004; Law, 2002). Internal perceptions, such as individuals’ judgments about the interference they experience in communicative participation, are latent traits that cannot be observed directly but instead can only be inferred from behaviors, such as how the individual responds to questions about the topic. An individual’s interpretation of his or her participation involves the interaction of complex and sometimes subjective physical, behavioral, personal, and contextual variables. These complex and subjective variables create challenges in measuring latent traits, such as self-perceived interference with participation.

Advances in modern psychometric methods have made it possible to develop a new generation of self-report instruments using protocols that can yield highly precise estimates of latent traits (Cella et al., 2007; Reeve et al., 2007). *Item response theory* (IRT) is central to the development of these measurement tools, and although it is fairly new to speech-language pathology, it has been used for several decades in the education fields. IRT is model-based measurement (Embretson & Reise, 2000), in that mathematical models are used to link observed behaviors (e.g., responses to questionnaires) to estimates of underlying latent traits. IRT incorporates both person and item characteristics to model the relationships between observed behaviors and latent traits (Bond & Fox, 2001; Embretson & Reise, 2000). Person characteristics may include the latent trait as well as other characteristics that might influence measurement, such as age, gender, or any number of other participant descriptors. Item characteristics include item difficulty and item discrimination. Item difficulty, referred to as *item location* for the remainder of this article, refers to the level of the underlying trait (or location on the trait scale) that the item measures most precisely. Item location is determined by the frequency with which an item is endorsed or is answered in a particular way. For example, in this study, items on which people frequently report interference with participation indicate a different level of the trait (different location on the interference range) than items on which people rarely report interference. Another item characteristic is *item discrimination*, or the degree to which an item can differentiate between respondents who have different levels of the latent trait. Items that are more sensitive to small changes in the latent trait have higher item discrimination.

There are many different types of IRT models. A Rasch model was used for this study for reasons that are discussed in the Method section. Several assumptions are critical to Rasch models (Bond & Fox, 2001; Embretson & Reise, 2000). First, Rasch models are one-parameter models, in that the only item characteristic allowed to vary is item location. Item discrimination is assumed to be constant across all items. Second, Rasch models assume that the item set is unidimensional, meaning that all items measure one underlying construct or trait. The final assumption is that of local independence, which means items are not related to each other beyond the shared underlying construct. When the assumption of local independence is met, an individual’s response to any item depends only on his or her underlying trait level and not on his or her responses to or understanding of any of the other items in the questionnaire.

A variety of methods can be used to examine how well items meet these assumptions and conform to the properties of a Rasch model. When these conditions have been adequately met, the model offers several advantages for measurement. The items can be ordered along an equal interval scale (the logit scale) according to the level of the latent trait that each item represents. This sequential ordering along an interval scale yields measurement properties that can generate valid and reliable estimates of the underlying latent trait. Another advantage of IRT, when the data fit the model adequately, is the separation of item parameters and person characteristics. This means that item parameters (e.g., item locations on trait range) are not sample dependent but are invariant across different samples. This makes it much easier to use IRT instruments across a variety of different participant groups. A key application of IRT is the possibility of creating large item banks from which statistically equivalent item subsets can be extracted to target individual assessment needs. These item banks might be created by compiling an original set of items, as with the Communicative Participation Item Bank in this study, or by combining items from existing related instruments into a single item bank. IRT-generated item banks are often administered by computerized adaptive testing (CAT; Cook, O'Malley, & Roddey, 2005; Fries, Bruce, & Cella, 2005; Ware, Gandek, Sinclair, & Bjorner, 2005). In CAT, presentation of items is individualized in that after the first item, presentation of subsequent items is determined by software algorithms based on the participant's responses to previous items. CAT provides measurement efficiency by yielding similarly precise measurement as traditional non-IRT instruments with fewer items and, hence, lower respondent burden (Cook et al., 2005). Large item banks provide a wide range of available items from which the most relevant or most targeted items can be selected to improve measurement precision for any individual.

IRT has been used in the field of speech-language pathology to evaluate existing instruments (Bogaardt et al., 2007; Donovan, Rosenbek, Ketterson, & Velozo, 2006; Donovan, Velozo, & Rosenbek, 2007; Donovan, Velozo, Rosenbek, Okun, & Sapienza, 2004; Hula, Doyle, McNeil, & Mikolic, 2006) and to compile item sets using items from different existing instruments (Doyle, Hula, McNeil, Mikolic, & Matthews, 2005). Although these analyses provide new information about the psychometric qualities of these scales, IRT analyses of existing scales limit some of the IRT applications. For example, IRT analyses of existing instruments will reveal characteristics of existing items, such as their location along the trait range. If there are gaps along the trait range where no items are located, existing questionnaires do not have a reservoir of additional items to draw from to ensure an adequate number of items distributed evenly across the range of desired measurement. Hence, existing instruments may not offer as complete coverage (precise measurement) of the trait as is possible. When instruments are developed from their inception with IRT, a large reservoir of candidate items can be created and tested, and those items that provide the most complete coverage of the trait range with the best fit to the model can be chosen for the final item bank.

This article introduces a new instrument to measure communicative participation: the Communicative Participation Item Bank (CPIB). This item bank should fill a gap in currently available instruments by providing an item bank that is dedicated to the construct of communicative participation and that has been developed with some of the most up-to-date guidelines for self-report scales (Fries et al., 2005; Reeve et al., 2007). The CPIB has been designed from its inception using IRT methodologies to take full advantage of IRT measurement properties.

The CPIB is intended to be a dynamic, self-report outcome measurement tool appropriate for clinical trials, research, and clinical practice. It is designed for community-dwelling adults across a variety of communication disorders, including motor speech, voice, and mild-to-moderate cognitive-communication disorders. Other instruments already exist to

address communication in adults with more severe language impairments—for example, the American Speech-Language-Hearing Association (ASHA) Functional Assessment of Communication Skills (Frattali, Thompson, Holland, Wohl, & Ferketic, 1995) or the ASHA Quality of Communication Life Scale (Paul et al., 2004). The items in the CPIB are restricted to speech communication to preserve sufficient unidimensionality. Unidimensionality is critical for measurement purposes because it allows all of the items in an item bank to be placed along a single scale (Bond & Fox, 2001). Prior research has suggested that including different communication modalities (reading, writing, and sign language) would violate assumptions of sufficient unidimensionality (Doyle et al., 2005). Future research is needed to address communicative participation in other communication modalities.

The current set of candidate items was generated by a multidisciplinary team on the basis of literature review (Eadie et al., 2006), prior qualitative studies of the experiences of people living with communication disorders (Baylor, Yorkston, & Eadie, 2005; Baylor, Yorkston, Eadie, & Maronian, 2007; Yorkston, Klasner, & Swanson, 2001), and the clinical and research experience of members of the team. The items underwent qualitative review by speech-language pathologists and by people with multiple sclerosis (Yorkston et al., 2007) and with spasmodic dysphonia (SD; Yorkston et al., 2008); the items were then modified on the basis of feedback from participants. The items ask about communication activities in various life domains (home, leisure, work, health care, community, and personal relationships), different communication contexts (e.g., face-to-face, phone, one-on-one, groups, familiar communication partners, and unfamiliar communication partners), and different purposes or goals (e.g., to schedule an appointment, to give instructions to a repair person, to have a casual conversation, and to share personal feelings). By specifying these various contexts and situations, the candidate items are designed to reflect the communicative participation that occurs as people engage in their various life roles. A sample item is presented in Figure 1, and a complete list of the current candidate items can be obtained by request from the first author. For each item, the participant is asked to rate the extent to which his or her condition (communication disorder) interferes with participation in that situation on a 5-point scale (*not at all*, *a little*, *quite a bit*, *a lot*, and *extremely*).

The long-term goal of this research is to create an item bank to measure communicative participation across different communication disorder populations. The current set of over 140 candidate items will be reduced to a core bank of approximately 60–80 items that (a) have the strongest psychometric properties with invariance across the variety of communication disorders for which the tool is intended and (b) cover a wide range of communicative participation levels. This core item bank will then be available for generating short forms or CAT applications for use in research and clinical settings. This article presents the first in a series of planned IRT psychometric analyses of the candidate item set for the CPIB. This study reports on the analysis of the item parameters in a sample of adults with SD. A single population was chosen for this study to reduce variability that would be introduced into the data by combining different communication disorders into one analysis. Furthermore, the focus on a single population contributes to long-term goals of the research program to study communicative participation in depth within separate communication disorder groups as well as across communication disorders. The results of this study will be combined with the results of similar future analyses in other communication disorder populations to determine the invariance of item parameters across communication disorder groups and to compile the final version of the CPIB.

Methods

Methods were approved by the Institutional Review Board at the University of Washington.

Participants

Spasmodic dysphonia (SD) is a chronic neurologic voice disorder treated most commonly by injections of Botox into the laryngeal muscles for temporary symptomatic relief of dysphonia (Duffy & Yorkston, 2003; Sulica, 2004). The SD population was chosen as the initial population for analyzing the CPIB for several reasons. People with SD cover a wide age range and are active in a wide range of life roles, allowing for exploration of communicative participation in a variety of life domains. Furthermore, SD is not systematically associated with other impairments that might influence participation. This allows exploration of the impact of a communication disorder independently from other conditions, such as mobility limitations.

Participants were adults 18 years of age or older who had been diagnosed with SD by an otolaryngologist. Participants were included regardless of type of SD (adductor, abductor, tremor, or mixed), severity of symptoms, duration of SD, or treatment history. Participants were recruited through multiple sources, including the National Spasmodic Dysphonia Association (NSDA; www.dysphonia.org), ASHA Special Interest Division 3: Voice and Voice Disorders, and local voice clinics. The purpose of the broad inclusion criteria was to recruit participants representing a wide diversity of demographic and condition characteristics. Targeted enrollment was 200 participants because that is the smallest sample size recommended for tenable latent trait modeling (Kline, 1998).

Questionnaire Administration

For this study, the CPIB consisted of 141 items that asked about the extent to which SD interferes with participation in a variety of communication situations. Participants were asked to rate the extent of interference on a 5-point scale (*not at all, a little, quite a bit, a lot, and extremely*). They were asked to answer the items according to how they feel when their speech is not at its best or is causing them problems, and to consider their experiences over the past year when responding to the items. The time frame of 1 year was chosen to ensure coverage of a wide range of situations, many of which might not arise within shorter time frames, such as weekly or even monthly, but may still be very relevant to people. For example, people may visit their health care provider or gather for large family social events only a few times per year, yet these few occasions might be very important to them and contribute to their overall judgments of their experiences with SD.

Participants were also asked to complete the VHI (Jacobson et al., 1997), the Levels of Speech Usage (a self-report categorical rating of speech usage patterns; Baylor, Yorkston, Eadie, Miller, & Amtmann, 2008), and a demographic information page that included questions about their SD history. Questionnaires were available in Web-based and paper versions. Online administration was chosen as a way to improve the efficiency and feasibility of collecting data from a large number of participants across the country. Out of concern, however, that some potential participants would not have access to an online questionnaire, a paper version was also made available. This helped to ensure the representation of a wide range of demographic variables without any potential systematic exclusion of people on the basis of use of online data collection only. Participants selected the method of administration that they preferred. The Web-based questionnaires were administered using WebQ, an online survey research tool developed by the University of Washington (www.catalyst.washington.edu/web_tools/webq.html). Participants who chose the Web-based administration option received an e-mail instruction page along with an

individual access code to sign onto the Web site. Participants were then able to go online independently and complete the questionnaires. The first author provided technical support for any participants who had questions about using the Web site. Participants who chose to complete the questionnaires on paper were mailed a packet consisting of the same questionnaires that the online participants received as well as a stamped return envelope.

Data Analysis

Data from the online and paper questionnaires were combined into one group for analysis. Raw data from the online questionnaires were available in an Excel spreadsheet download from WebQ. Mplus (Version 4.2; Muthén & Muthén, 1998) was used for an exploratory factor analysis (EFA), and Winsteps was used for the Rasch analyses (Linacre, 1991).

A Rasch (one-parameter) model was used for the IRT analyses for this study. The item parameter modeled in a Rasch analysis is item location along the trait range. Other item parameters, such as the discrimination ability of items, are held constant. The Rasch model is recommended for item bank construction for various reasons, including (a) separation of item and person parameters resulting in independence of the item parameters from the particular test sample (provided that the data fit the model adequately), (b) distinct ordering of items in terms of their location on a trait range, and (c) maintenance of consistent interval unit scaling across the range of items (Bezruczko, 2005).

Within the one-parameter family of models there are additional models to choose from depending on response format (i.e., dichotomous vs. polytomous responses). The partial credit model (PCM; Wright & Masters, 1982) was chosen for this analysis because it is appropriate for polytomous (Likert-type) items. The PCM does not impose a uniform structure on the response categories across all items in terms of the number of response categories or the trait thresholds for crossing from one category to another. The PCM allows the number and structure of response categories to vary across items. This model is very helpful, particularly for the initial analyses of items in which the response patterns of participants are unknown. The PCM allows the investigator to determine whether there are too many or too few response categories so that the categories can be revised if needed. Descriptions of the specific Rasch analyses conducted for this study follow.

Response categories—Well-functioning response categories must meet several criteria. Each category should cover its own region of the trait scale where it is the most likely response option. This is assessed visually on a plot of category probability curves in which each category should have a unique peak and adequate distribution along the logit scale to distinguish it from neighboring categories. The distance between category thresholds should be at least 1.4 logits to show distinct categories but not larger than 5.0 logits, which would suggest gaps in the categories (Bond & Fox, 2001; Linacre, 2004). The categories should advance monotonically, meaning that the curves should follow their intended category order along the logit scale (i.e., the response curves should be in the same sequence as the categories ranging from *not at all* to *a little* to *quite a bit* etc.; Bond & Fox, 2001; Linacre, 2004). Finally, each category should have a minimum of 10 respondents endorsing that category on each item (Linacre, 2004).

Item fit—Item fit is evaluated in terms of the degree to which the observed participants' responses correspond to expected responses on the basis of the IRT model. Infit and outfit statistics are two types of fit statistics common in IRT to assess the residual differences between actual and expected participants' responses. The infit mean square statistic is information-weighted, giving more weight to large residuals (unexpected responses) that are close to the item's location on the logit scale. Outfit statistics are not weighted and are

therefore more sensitive to more extreme outlying scores (Bond & Fox, 2001). Infit statistics are often used in Rasch analyses because unexpected responses for participants whose trait levels are expected to be close to an item's location raise more concern about the functioning of an item than frank outlying scores (Bond & Fox, 2001). For this study, fit was assessed using the infit mean square statistic, which is calculated as the squared deviation of observed responses from model expectations summed across participants, divided by the expected variance of these deviations summed across participants. On the basis of recommendations from Bond and Fox (2001), items with infit mean square statistics within the range of 0.75–1.3 were retained as well-fitting items. Following recommendations by Linacre (1991), items with poor fit were removed from the item set in an iterative manner starting with the items with the highest infit values. Reanalysis after removal of each item is recommended because the removal of items may result in changes in the parameters of remaining items on subsequent analyses.

Unidimensionality—Prior to conducting IRT analyses, EFA can be used to determine whether evidence of sufficient unidimensionality exists to proceed with IRT (Embretson & Reise, 2000; Smith, 2004). EFA was conducted using Mplus (Muthén & Muthén, 1998) with un-weighted least squares for the estimation method and pairwise deletion of missing data. The EFA was conducted on the full set of 141 items, and no items were removed as part of the EFA.

Poorly fitting items were removed during the Rasch analysis, and after removal of these items, unidimensionality was assessed again using the principal components analysis of the residuals within Winsteps. The principal components analysis of the residuals is an analysis of the residual variance that remains after accounting for the variance that can be attributed to the unidimensional Rasch model (Linacre, 1991; Smith, 2004). The Rasch model should account for a large proportion of variance (>60%; Linacre, 1991), and the residual variance should be random (the first principal contrast in the residuals should account for less than 5% of the unexplained variance; Linacre, 1991).

Local independence—Local independence exists when there are no significant relationships among items in an item bank after accounting for the dominant unidimensional construct, as described in the prior section (Reeve et al., 2007). When items are related in ways beyond the core construct, this creates local dependence. Local dependence may indicate the presence of additional constructs that influence participants' responses. Items may also be dependent if the response to one item depends in some way on a response to a prior item (perhaps a prior item provided additional content that is helpful in a later item or increases the likelihood of answering with a similar response to the later item). One common example of local dependence from educational testing is when several reading comprehension questions reference a single reading passage (Wainer & Thissen, 1996). In such a situation, individuals may have prior experience or familiarity with the subject of the reading passage. This knowledge, and not just the respondents' reading comprehension, could impact the probability of correctly responding to the items anchored to the passage. Thus, something beyond the trait being measured (reading comprehension) is affecting responses to all of the items related to that passage. Local dependence is problematic in constructing item banks because items that are dependent contain redundancies and, therefore, do not contain as much psychometric information as the IRT model might predict (Chen & Thissen, 1997). This reduces reliability, and then more items (longer instruments) are required to achieve desired reliability levels (Wainer & Thissen, 1996). Local dependence may also affect estimation of item parameters (Chen & Thissen, 1997).

Local dependence may be evaluated in several ways. Evidence of essential unidimensionality provides support for the assumption of local independence because if all

items measure the same underlying construct, this construct accounts for any relationships among items, and other relationships among items are unlikely or sufficiently weak as to not affect measurement (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). Local dependence also can be evaluated by examining the residual interitem correlations. These are the correlations among item responses after accounting for the core unidimensional construct. High residual inter-item correlations suggest local dependency. For this study, residual interitem correlations were obtained from Mplus after finding a one-factor EFA solution. Correlations of item pairs greater than 0.2 were flagged as items possibly demonstrating local dependence (Reeve et al., 2007). The correlations were then converted to standardized correlations using Fisher's z transformation in SPSS (SPSS, 2003). Fisher's z uses a logarithmic transformation to convert Pearson's correlations (r s) to a normal distribution. Item pairs whose Fisher's z scores were greater than 2 SD s away from the mean of Fisher's z values were classified as locally dependent (Shen, 1996). These items will be examined closely in future research but will not be discarded from the item bank on the basis of this study involving a single population. Final decisions regarding removing items on the basis of local dependency will be made in future research after data from multiple populations are available.

Measurement range—The CPIB is intended to be applicable across a wide range of people in terms of the extent of interference they experience in communicative participation. Items should be distributed evenly across as much of this range as possible to provide adequate measurement along the continuum of interference in participation. Each item's location on the trait range is designated by its logit value and specifies the point on the trait range that the item measures most precisely. Items are ordered along the logit scale according to response probabilities.

In IRT, information functions provide graphical illustrations of the measurement precision of an item set across the trait range. Test information functions provide this information for the item set as a whole, whereas item information functions provide this information for individual items. Measurement is most precise in those regions where the peak of the curve is highest. The information function is the inverse of the standard error of measurement, so in those regions where the information peak is highest, measurement error is lowest (Linacre, 1991). Instruments, such as the CPIB, that are intended to be applicable across a wide range of people with different levels of the trait should have a test information function that is broad enough to provide equally precise measurement across a wide range of the trait (Hambleton et al., 1991).

The test information function was obtained from Winsteps. The standard error of measurement curve was obtained in Excel using the data from the test information function in the following formula: standard error = $1/\sqrt{\text{test information}}$ (Embretson & Reise, 2000).

Reliability—Cronbach's alpha was obtained with SPSS. Cronbach's alpha provides a measure of the correlation among items in a set. High values for Cronbach's alpha suggest that items are closely related and that inclusion in the same item set is appropriate. Cronbach's alpha can be inflated, however, when a very large number of items are in the set (Pett, Lackey, & Sullivan, 2003). Another method of assessing reliability is through inter-item correlations, which were obtained from Mplus.

Correlation with VHI—Correlation between scores on the CPIB and the total VHI score was calculated using Spearman's correlation coefficient. The VHI is currently the most commonly referenced voice psychosocial questionnaire (Eadie et al., 2006) and provides a standard against which to compare new measurement tools in populations with voice disorders.

Results

A total of 322 prospective participants contacted the primary investigator and were sent the needed information to complete the questionnaires. Of these, 208 (64.6%) returned completed, usable questionnaires. Of these 208 participants in the study, 168 (80.8%) completed the questionnaires online, and 40 (19.2%) completed the questionnaires on paper.

Of the total number of possible responses (141 items $\times$ 208 participants), 7.6% were missing. The items were analyzed to determine whether some items had higher rates of missing data than others. Of the 46 items with more than 10 missing responses, 18 items were related to work and employment; 10 items were related to communication with people who live with you and would not be applicable to people living alone; 5 items were related to communicating with young children; 3 items were related to communicating with pets; 2 items addressed giving formal presentations; and the remaining items dealt with a variety of lifestyle issues, such as romance/flirting, participating in religious activities, and using public transportation, such as buses and taxis. These items were not expected to be applicable in all life situations. Although all of the original 141 items were included in the IRT analyses for this study, future recommendations for the item bank will include removal of items with high rates of missing data. Subgroups of items, such as work-related items, may form separate testlets if appropriate as the item bank is constructed.

Participant Information

Demographic characteristics of the sample are presented in Table 1. The sample was predominately female (77.9%) and Caucasian (94.2%); the average age of the participants was 55.4 years ($SD = 11.0$), and 62.0% of the participants were in either full- or part-time paid employment. Thirty-seven U.S. states were represented, and there were 4 participants from Canada and the United Kingdom. Participants were asked whether they had any other medical conditions that in their opinion influenced their participation in their daily life activities. Sixty-seven (32.2%) participants reported other medical conditions that they felt influenced their daily activities. The most commonly reported conditions were depression/anxiety/panic disorder (15 participants), other forms of dystonia (9 participants), arthritis (9 participants), and nonvoice tremor (8 participants). In terms of treatment status for SD, the largest proportion of participants was currently receiving Botox injections (59.1%). The next largest group was participants who were not receiving any treatment at the time of the study. The remaining participants reported a variety of interventions, including voice therapy, surgical intervention and medications. As a self-rating of voice severity, participants were asked to rate their voices compared with typical peers without SD on the day that they completed the questionnaires. The majority of participants rated their voices as “very different” than their peers without SD.

The characteristics of the sample, particularly the predominance of women and the age range, were consistent with SD prevalence data (Sulica, 2004). To provide an indication of how well this sample represented a broader SD population, demographic characteristics were compared with data from an online survey conducted by the NSDA in 2006 with 758 respondents (Feeley, 2008). In the NSDA study, 77.4% of the respondents were female, 93.1% were Caucasian, 29.4% were not in paid employment, and 65.0% reported that Botox treatments had been the most helpful treatment that they had tried for SD (Feeley, 2008).

The mean VHI score in the current study was 83.4 ($SD = 18.5$). (The VHI has a possible range of 0–120, with lower scores indicating less voice-related handicap.) This mean is higher than the pre-Botox treatment VHI scores in the following studies: Wingate et al. (2005) reported a mean of 63.5 ($SD = 27.2$); Benninger, Gardner, and Grywalski (2001) reported a mean of 67.6 ($SD = 14.7$); Eadie et al. (2007) reported a mean of 68.1 ($SD =$

21.8); and Anari, Carding, Hawthorne, Deakin, and Drinnan (2007) reported a mean of 70.9 (95% confidence interval = 65.8–76.1). These differences may be due to methodological variations among the studies as well as characteristics of the sample for this study, which did not include treatment status as an inclusion criterion.

IRT Results

The following sections present the results for the psychometric item analyses.

Response categories—The results suggest that participants did not differentiate between the third response option (*quite a bit*) and the fourth response option (*a lot*). Analyses of the distance between response category thresholds revealed that 96.5% of the items had category distances below the 1.4 lower limit (Bond & Fox, 2001) for one or both of these categories, suggesting that these two categories did not function independently of each other. All items were recoded to a four-category response set by combining the *quite a bit* and *a lot* categories. An example of response category function is presented in Figures 2A and 2B for the item, “Taking a phone message for someone who lives with you.” Figure 2A shows the original five-category response set. The third and fourth response options are not independent of neighboring categories, as indicated by the considerable overlap between the curves. Figure 2B shows the revised four-category response set for the same item. Four clearly distinct peaks, representing four distinct response categories, are evident. All of the following analyses were conducted using the revised four-category response format.

Item fit—In the initial analysis, 12 of the 141 items were identified as having poor fit statistics using the mean square infit statistic boundaries of 0.75–1.3 (Bond & Fox, 2001). Table 2 summarizes the item fit data for poorly fitting items. For comparison, Table 3 presents the 15 items with the best fit to the model.

Of the misfit items, 10 had infit mean square values greater than 1.3, and the other 2 items had infit values less than 0.75. Items with poor fit are typically flagged for removal from the item bank or for revision and retesting in the future. Out of concern for data from this single population exerting undue influence over the removal of items from the item bank, a conservative approach was taken for removing items. The misfitting items were analyzed to determine whether misfit might be due to high rates of missing data. Items that misfit because of high rates of missing data may be items that simply do not apply to a large number of people. Six items with poor fit had high rates of missing data, and these appeared to be items that would pertain only to people with specific lifestyles (e.g., participating in religious activities or owning pets). These lifestyle issues were expected to be similar across different communication disorder groups. For that reason, poorly fitting items with high rates of missing data (or not applicable responses) were flagged for consideration for removal from the item bank.

The six items with high rates of missing data were removed in an iterative manner starting with the item with the highest mean square infit statistic. Items were removed one at a time in the order presented in Table 2 with the data reanalyzed after each item removal (although all three pet-related items were removed at the same time because of very similar content). As items were removed and the data reanalyzed, two additional items rose above the upper fit statistic limit and were also removed because of high rates of missing data. This resulted in removal of 8 items from the set with 133 items remaining for further analysis.

Because high rates of missing data were a suspected reason for item misfit for some items, the item set was analyzed to determine whether any well-fitting items had missing data at rates of 25% or higher—similar to the missing data levels for the eight misfit items that were removed. Eleven items did have missing data rates of 25% or higher but still had acceptable

item fit statistics. Five of these items were related to work situations; four of the items were related to communication with young children, one item was related to using public transportation (such as buses and taxis), and the final item addressed participation in clubs or other organized social groups.

For the six misfit items with missing responses for 2% or fewer of the participants (see Table 2), missing data probably did not account for the misfit. Three of these items had less than 10 responses in a response category, which might contribute to poor item fit. Poor item fit might also be attributed to poor item discrimination. Although the Rasch model treats all items as having the same discrimination, in reality item discrimination does vary across items. Estimates of item discrimination were calculated as a post hoc analysis in Winsteps, which provides a value of 1.0 as the expected discrimination value. Greater deviations in discrimination from 1.0 generally correspond to poorer model fit (Linacre, 1991). Discrimination values less than 1.0 indicate lower discrimination than would be expected by the model and are typically associated with high infit mean square values (Linacre, 1991). The item discrimination values (on the basis of the original analysis of all 141 items) for all of the poorly fitting items are listed in Table 2. Poor item discrimination may have been a contributor to poor item fit in these items on the basis of the observation that the discrimination values of the items in Table 2 composed 12 of the 14 lowest discrimination values and 2 of the 5 highest discrimination values among all 141 items. These deviations from the assumption of equal discrimination across items might call into question the applicability of the Rasch model. As mentioned above, however, the current bank of candidate items is large enough that elimination of additional items because of poor fit does not raise concern at this time. Further exploration of these items in future research is warranted before making final decisions about removal of items from the item bank.

Unidimensionality—The scree plot for the EFA on the original set of 141 items is presented in Figure 3. The scree plot suggested the presence of one dominant factor, in that there is a steep declining slope between the first and second factors (steep decline in eigenvalues accounted for), and the second and all subsequent factors are fairly level on the plot (there is a very gradual or no decline in eigenvalues accounted for by subsequent factors). The Rasch principal components analysis of the residuals was conducted on the 133-item set after the removal of misfitting items. The Rasch model accounted for 89.3% of the variance, and the first principal contrast of the residuals accounted for 0.9% of the unexplained variance. This evidence suggests sufficient unidimensionality.

Local independence—The evidence of sufficient unidimensionality in the prior section supports the assumption of local independence. Local independence was also assessed by examining the residual interitem correlations using Mplus for the 133 items remaining after removal of poorly fitting items. Residual correlations ranged from -0.210 to 0.394 . Residual correlations greater than 0.2 (Reeve et al., 2007) were observed in 97 (1.1%) of the pairs. The correlations were transformed to standardized correlations using Fisher's z transformation (Shen, 1996), with the recommendation to flag any pairs with standardized correlations extending beyond 2 SDs . Using this criterion, 3.3% of the pairs had standardized correlations greater than 2.0, and 0.8% of the pairs had standardized correlations less than -2.0 . Ideally, the residual correlations would be as close to zero as possible, showing that the items are not related or dependent on each other beyond the shared unidimensional construct of the item bank. As examples of item pairs with extreme correlations, Table 4 presents the 10 item pairs with the highest standardized Fisher's z correlations and the 10 pairs with the lowest standardized correlations. The reason for most of the high correlations was due to similarity in topics or wording. We intentionally included items that addressed the same topic but with different wording to evaluate how wording might affect item characteristics, such as item location in the trait range. The impact of

wording may be very important, particularly for populations with mild-to-moderate language or cognitive–communication impairments who are among the intended future users of this item bank. These nearly duplicate items will be considered in future research, and locally dependent items will be eliminated before formulation of the final item bank. These decisions will be based on maximizing model fit and obtaining even measurement coverage of the trait range once data from additional populations are available.

Measurement range—To evaluate measurement range, the response categories for rating extent of interference were scored as follows: *not at all* = 4, *a little* = 3, *quite a bit/a lot* = 2, and *extremely* = 1. With this scoring format, higher scores are more favorable, in that a higher score represents less interference in participation. The item location values for the 133-item set ranged from -2.73 logits (Item: “Being polite when talking to people”) to 2.20 logits (Item: “Making a telephone call to get information”). Items are distributed according to response probabilities. In this scoring format, people are more likely to report interference on items with high logit values, and they are less likely to report interference on items with low logit values.

The items were distributed fairly evenly across this range, with wider gaps at the extreme ends of the range. Figure 4 presents the distribution of items along the logit scale with each item represented by a single horizontal line. Figure 4 also presents five sample items for illustration of the types of items that fell into different regions of the item range. There is even coverage from the range of approximately -2.0 to 1.5 logits. The area from 0.5 to 1.5 logits is relatively more dense with items, and the regions beyond ± 2.0 logits are more sparsely populated with items. In IRT, when multiple items have the same item location, some of these items may be removed from the bank to reduce measurement redundancy and to pare down the number of items (Bond & Fox, 2001). No decisions were made about removing items on the basis of redundancy in item location during this study for two reasons. First, items that are redundant in location (measure the same trait level) might ask about very different communication situations and therefore provide unique content information even if they provide redundant quantitative measurement. Second, the results of this study require replication across other communication disorders before final decisions are made about removing items on the basis of their location on the trait range.

The test information function in Figure 5 compares measurement precision with levels of the trait being measured. Embretson and Reise (2000) have noted that “... if the conditional information is around 10, then the conditional standard error is about 0.31 [and] a standard error of 0.31 corresponds to a reliability coefficient of 0.90” (p. 270). Figure 6 presents the standard error of measurement across the trait range generated from the data in the test information function. Figures 5 and 6 suggest that the item bank has good measurement precision even beyond the range of -3 to 3 logits, which is a common range for logit scale values (Embretson & Reise, 2000).

Participants’ theta values (scores) can be reported on the same logit scale as item location values. A participant’s theta is that point on the logit scale at which he or she has an equal chance of reporting no interference versus extreme interference on items at that location. A participant is more likely to report high interference on items at higher logit levels than his or her theta, and a participant is more likely to report low interference on items at lower logit levels than his or her theta. For example, on Figure 4, if hypothetical Participant 1 has a theta value of -1.2 logits, he or she is more likely to report interference on Items A–C and is less likely to report interference on Items D and E. If hypothetical Participant 2 has a theta value of 1.8 logits, he or she is more likely to report interference on Item A and is less likely to report interference on Items B–E. Participant 2 has a higher theta than Participant 1, so

Participant 2 in general experiences less interference in communicative participation than Participant 1.

The range of participants' scores was from -8.48 to 4.83 logits. Figure 5 includes the distribution of individual participants' scores below the test information curve for a comparison of the distribution of participants across the measurement range of the item set. When comparing the range of participants' scores with the test information function, it is evident that this item bank does not provide equally precise measurement across the range of participants' scores observed in this sample. Scores for participants at the extreme ends of the range have greater measurement error, although only those scores extending beyond approximately ± 6 logits have estimated reliability less than 0.90 .

Reliability—Cronbach's alpha was $.99$, an estimate likely inflated by the large number of items (Pett et al., 2003). The average interitem correlation was 0.623 , with a range of correlations from 0.194 to 0.962 . Item pairs having the highest interitem correlations (>0.9) could be explained by evident relationships in item content. For example, there were high interitem correlations among four items related to using the telephone: "Making appointments over the phone for personal services" (e.g., haircut, dentist); "Making a phone call to schedule services or repairs" (e.g., home or car); "Making a phone call for household business"; and "Making a telephone call to get information." Another trend for high correlations was found among items related to communicating with health care providers: "Communicating with familiar health care providers about your health care," "Answering questions from a doctor or health care provider who you know," "Telling a doctor or health care provider who you know about your symptoms or medical history," and "Asking a familiar doctor or health care provider questions." Examples of other highly correlated items included situations such as communicating at home ("Communicating with people who live with you" and "Communicating at home") and communicating with unfamiliar partners ("Having a casual conversation with someone you do not know very well" and "Communicating at social gatherings where you do not know most of the people"). Elimination of some of the items in each of these highly correlated sets will be considered in future research once data from additional communication disorder populations are available and once patterns across populations can be compared. Elimination of redundant item content will reduce the number of items while still retaining a range of important communication situations. Decisions about eliminating items will balance the need to maximize model fit, obtain good coverage of the trait range, and retain items that are widely applicable across populations. For example, in the pair of items about communicating at home, the item "Communicating with people who live with you" will not be applicable to individuals who live alone, but the item "Communicating at home" has a broader interpretation and might be more widely applicable for any communication situations that occur in the home.

Correlation with VHI—The correlation between participants' scores on the CPIB and total VHI scores using Spearman's correlation was -0.678 .

Discussion

The availability of measurement tools designed specifically for communicative participation is critical to a biopsychosocial approach to communication disorders. A measure of communicative participation is important for understanding the impact that a communication disorder (or other health condition) has on an individual's participation in and fulfillment of the communication aspects of his or her life roles. By measuring participation, we gain insight into what people actually experience in real-life situations as they face their daily communication needs.

Although there are many physical, personal, and environmental factors that will affect communicative participation, one single measurement tool cannot encompass all of these constructs. A measure of communicative participation should be sufficiently unidimensional to avoid confounding measurement of participation with measurement of other components of a biopsychosocial framework, such as physical function or environmental influences. Exploration of the interactions among communicative participation and contributing variables should be conducted by comparing data from different measurement tools, each designed to provide unidimensional measurement of individual constructs. The CPIB is an instrument currently under development to provide a unidimensional measure of speech-related communicative participation. When completed, the item bank will be available for use in future research to explore the construct of communicative participation, particularly the roles of various contributing factors and the effects of different types of intervention on participation. The CPIB is also intended to be a clinically useful outcome measurement instrument. The purpose of this study was to conduct the first in a series of psychometric analyses of the CPIB. This study provided item calibrations for 141 candidate items in a sample of people with SD.

IRT analyses of the candidate items suggested that participants did not differentiate between two of the five response categories. The response categories were recoded to a four-category set currently labeled as *not at all*, *a little*, *quite a bit*, and *extremely*. The improved functioning of a four-category response format over formats with higher numbers of categories is consistent with findings from IRT analyses of other speech-language pathology instruments (Donovan et al., 2004, 2006, 2007; Doyle et al., 2005).

In terms of item fit to the IRT model, 8 of the 141 items failed to meet infit statistic criteria and were flagged for removal from the item bank because the reason for misfit appeared to be a high rate of *not applicable* responses on the basis of life situations. An additional six misfitting items whose reasons for misfit were less clear or potentially disorder-specific were retained for future analysis in other populations. Retention of these items at this time did not diminish the psychometric strengths of the item bank. The remaining 133 items form a sufficiently large enough set of items for an item bank. The size of the item bank is expected to decrease with future analyses with other communication disorder populations. At present, the 133 items in the bank demonstrate evidence of sufficient unidimensionality. There is some evidence of local dependence, which would be expected in an item bank of this size. The presence of some dependent items may also have been created by the inclusion of items that use slightly different wording to ask about similar content. These items were included to allow observation of the effects of wording variations on item parameters. This is of concern because this item bank is intended for people across a range of communication disorders, including people with mild-to-moderate language or cognitive-communication disorders, and if wording variations cause some items to be more “difficult” than other items or to affect parameters in other ways, this information would be useful in selecting items for the bank. Examples of such variations have already been found by Doyle et al. (2005) in their analysis of an item set compiled from several questionnaires for people who have had strokes. Their Rasch analysis showed that the item “Talk on the telephone” had an item logit value of -0.50 logits, whereas a similar item, “Have a conversation on the phone,” had an item logit value of -0.10 . The item “Talk with a group” had a logit value of 0.06 , whereas the item “Participate in a group conversation” had a logit value of -0.32 . When additional data are available across populations for use in compilation of the final CPIB, decisions will be made about eliminating items with redundant content with consideration for maximizing item fit, for coverage of the trait range, and for applicability across participants.

The 133 items remaining in the item bank demonstrate good coverage of the trait range in terms of item location. IRT orders items along the logit scale in terms of response

probability, and the distribution of items is consistent with findings from prior qualitative analyses of the items (Yorkston et al., 2008). Items on which participants were more likely to report participation interference in the current study were the items that participants in the prior cognitive interviews reported as more difficult communication situations. These situations typically involved unfamiliar communication partners, situations in which they were rushed, use of the telephone, or more serious conversational situations. Items on which participants were less likely to report participation interference in the present study were the items that participants in the prior cognitive interviews reported to be easier communication situations. These situations typically involved familiar communication partners, communication in familiar or comfortable settings (such as home), and more relaxed or casual situations.

The test information function and standard error curve suggested that the item bank provides good measurement precision across the desired trait range. Measurement is not as precise at the extreme ends of the trait range, as is often the case for measurement tools. Future research should include efforts to add or modify items that will increase the psychometric information available at the extreme ends of the range.

The degree of correlation between the VHI and the CPIB found in this study suggests that the two instruments cover related but not entirely redundant content. As reported by Eadie et al. (2006), the VHI does contain some items that measure communicative participation, and this would account for the degree of correlation observed between the two tools. The lack of complete overlap, however, stems from the observation that the CPIB is designed to contain only items addressing communicative participation, whereas the VHI contains items covering a wider array of constructs, including physical symptoms and personal or emotional consequences. The choice between using the CPIB or the VHI will depend on the purpose of the assessment. For example, if the objective is to obtain a multidimensional overview of the effects of a voice disorder, the VHI will continue to be a useful tool because it samples multiple biopsychosocial domains relevant to voice disorders. If, however, precise unidimensional measurement of communicative participation is needed, particularly for purposes in which relationships among participation and contributing variables need to be explored, the CPIB may be the better choice. The CPIB may be useful in future research to develop theoretical models of participation, to understand the roles that various contributing factors play in communicative participation, and to understand the impact of various types of intervention on participation.

Future Steps in Item Bank Development

The present study is very promising in that the candidate items for the CPIB appear to function well according to Rasch principles. These results are preliminary, however, and several additional steps are needed before the item bank will be ready for everyday clinical applications (Fries et al., 2005). Because this tool is intended for use in a wide variety of communication disorder populations, both cognitive interviews and the IRT analyses need to be replicated in additional populations to evaluate the invariance of item characteristics across different communication disorders. The results from this study will be combined with the results from future replication studies in other communication disorder populations to guide the finalization of the item bank. The goal is to retain a core item bank of 60–80 items that demonstrate strong and invariant psychometric properties across multiple communication disorders. After a core item bank is established, the next stages of research will involve generating CAT applications to make using the item bank easy and efficient for clinicians and researchers.

Conclusions

In conclusion, this study offers two contributions to the literature. First, it presents a new measurement tool to address a gap that currently exists in the ability to measure self-reported participation restrictions associated with communication disorders (Eadie et al., 2006). If the Communicative Participation Item Bank continues to show strong psychometric properties in future analyses, it will provide a tool for researchers and clinicians to use in a genuinely biopsychosocial approach to working with people with communication disorders. Second, this study is among the first reports of using IRT for the development of new speech-language pathology instruments. IRT has been used in a small number of studies (a) to evaluate existing speech-language pathology instruments that address psychosocial components of communication disorders (Bogaardt et al., 2007; Donovan et al., 2004, 2006, 2007) and (b) to compile a combined item set from multiple instruments (Doyle et al., 2005). Using IRT to develop speech-language pathology instruments will advance the measurement science for many aspects of communication disorders, particularly self-report constructs that historically have been difficult to measure with psychometric rigor.

Acknowledgments

This project was funded by National Institutes of Health (NIH) Planning Grant 1R21 HD 45882-01, with additional technical support from personnel at the University of Washington Center on Outcomes Research in Rehabilitation funded by NIH Patient-Reported Outcomes Measurement Information System (PROMIS) Grant 5U01AR052171-03. We wish to express appreciation to all of the participants who volunteered their time to this project and who have been so willing to share their experiences with us. We also wish to thank the staff, clinicians, and researchers associated with the University of Washington Medical Center Voice Clinic, the National Spasmodic Dysphonia Association, and the American Speech-Language-Hearing Association Special Interest Division 3 Voice listserv, who helped us with participant recruitment. There are too many of you to thank individually here, but your contributions are appreciated. The contributions of Jean Deitz and Brian Dudgeon—members of the development team for the CPIB—are gratefully acknowledged, as are the contributions of Karon Cook for input on statistical analyses. Thank you to Jody Guariz and Ashlie Stegeman for help with data entry.

References

- Anari S, Carding PN, Hawthorne MR, Deakin J, Drinnan MJ. Nonpharmacologic effects of botulinum toxin on the life quality of patients with spasmodic dysphonia. *Laryngoscope*. 2007; 117:1888–1892. [PubMed: 17690610]
- Baylor CR, Yorkston KM, Eadie TL. The consequences of spasmodic dysphonia on communication-related quality of life: A qualitative study of the insider's experiences. *Journal of Communication Disorders*. 2005; 38:395–419. [PubMed: 15963338]
- Baylor CR, Yorkston KM, Eadie TL, Maronian N. The psychosocial consequences of Botox injections for spasmodic dysphonia: A qualitative study of patients' experiences. *Journal of Voice*. 2007; 21:231–247. [PubMed: 16564675]
- Baylor CR, Yorkston KM, Eadie TL, Miller R, Amtmann D. Levels of Speech Usage: A self-report scale for describing how people use speech. *Journal of Medical Speech-Language Pathology*. 2008; 16:191–198.
- Benninger MS, Gardner G, Grywalski C. Outcomes of botulinum toxin treatment for patients with spasmodic dysphonia. *Archives of Otolaryngology—Head and Neck Surgery*. 2001; 127:1083–1085. [PubMed: 11556856]
- Bezruczko, N. Rasch model essentials. In: Bezruczko, N., editor. *Rasch measurement in health sciences*. Maple Grove, MN: JAM Press; 2005. p. 35-71.
- Bickenbach JE, Chatterji S, Badley EM, Ustun TB. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Social Science and Medicine*. 1999; 48:1173–1187. [PubMed: 10220018]
- Bogaardt HCA, Hakkesteegt MM, Grolman W, Lindeboom R. Validation of the Voice Handicap Index using Rasch analysis. *Journal of Voice*. 2007; 21:337–344. [PubMed: 16504469]

- Bond, TG.; Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Erlbaum; 2001.
- Brown M, Dijkers MPJM, Gordon WA, Ashman T, Charatz H, Cheng Z. Participation objective, participation subjective: A measure of participation combining outsider and insider perspectives. *Journal of Head Trauma Rehabilitation*. 2004; 19:459–481. [PubMed: 15602309]
- Cardol M, Brandsma JW, De Groot IJM, van den Bos GAM, de Haan RJ, de Jong BA. Handicap questionnaires: What do they assess? *Disability and Rehabilitation*. 1999; 21:97–105. [PubMed: 10206348]
- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Medical Care*. 2007; 45(Suppl 1):S3–S11. [PubMed: 17443116]
- Chen WH, Thissen D. Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*. 1997; 22:265–289.
- Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: Time to let the CAT out of the bag? *Health Services Research*. 2005; 40:1694–1711. [PubMed: 16179003]
- Deary IJ, Wilson JA, Carding PN, MacKenzie K. VoiSS: A patient-derived voice symptom scale. *Journal of Psychosomatic Research*. 2003; 54:483–489. [PubMed: 12726906]
- Donovan NJ, Rosenbek JC, Ketterson TU, Velozo CA. Adding meaning to measurement: Initial Rasch analysis of the ASHA FACS social communication subtest. *Aphasiology*. 2006; 20:362–373.
- Donovan NJ, Velozo CA, Rosenbek JC. The Communicative Effectiveness Survey: Investigating its item-level psychometric properties. *Journal of Medical Speech-Language Pathology*. 2007; 15:433–447.
- Donovan, NJ.; Velozo, CA.; Rosenbek, JC.; Okun, MS.; Sapienza, CM. Communicative effectiveness survey measurement properties for individuals with Parkinson's disease. Paper presented at the annual convention of the American Speech-Language-Hearing Association; Philadelphia, PA. 2004 November.
- Doyle PJ, Hula WD, McNeil MR, Mikolic JM, Matthews C. An application of Rasch analysis to the measurement of communicative functioning. *Journal of Speech, Language, and Hearing Research*. 2005; 48:1412–1428.
- Duffy JR, Yorkston KM. Medical interventions for spasmodic dysphonia and some related conditions: A systematic review. *Journal of Medical Speech-Language Pathology*. 2003; 11:ix–lviii.
- Eadie TL. The ICIDH-2: Theoretical and clinical implications for speech-language pathology. *Journal of Speech-Language Pathology and Audiology*. 2001; 25:181–200.
- Eadie TL. The ICF: A proposed framework for comprehensive rehabilitation of individuals who use alaryngeal speech. *American Journal of Speech-Language Pathology*. 2003; 12:189–197. [PubMed: 12828532]
- Eadie TL, Nicolici C, Baylor CR, Almand K, Waugh P, Maronian N. Effect of experience on judgments of adductor spasmodic dysphonia (ADSD). *Annals of Otology, Rhinology and Laryngology*. 2007; 116:695–701.
- Eadie TL, Yorkston KM, Klasner ER, Dudgeon BJ, Deitz J, Baylor CR, et al. Measuring communicative participation: A review of self-report instruments in speech-language pathology. *American Journal of Speech-Language Pathology*. 2006; 15:307–320. [PubMed: 17102143]
- Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah, NJ: Erlbaum; 2000.
- Feeley, KA. Easier done than said: Living with a broken voice. Itasca, IL: National Spasmodic Dysphonia Association; 2008.
- Frattali, CM.; Thompson, CK.; Holland, A.; Wohl, CB.; Ferketic, MM. The American Speech-Language-Hearing Association Functional Assessment of Communication Skills for Adults. Rockville, MD: American Speech-Language-Hearing Association; 1995.
- Fries JF, Bruce B, Cella D. The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. *Clinical and Experimental Rheumatology*. 2005; 23(Suppl 39):S53–S57. [PubMed: 16273785]
- Hambleton, RK.; Swaminathan, H.; Rogers, HJ. Fundamentals of item response theory. Newbury Park, CA: Sage; 1991.

- Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice*. 1999; 13:557–569. [PubMed: 10622521]
- Hula W, Doyle PJ, McNeil MR, Mikolic JM. Rasch modeling of Revised Token Test performance: Validity and sensitivity to change. *Journal of Speech, Language, and Hearing Research*. 2006; 49:27–46.
- Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS, et al. The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology*. 1997; 6:66–70.
- Kline, RB. Principles and practices of structural equation modeling. New York: Guilford Press; 1998.
- Law M. Participation in the occupations of everyday life. *The American Journal of Occupational Therapy*. 2002; 56:640–649. [PubMed: 12458856]
- Linacre, JM. Winsteps 3.63.0: Multiple choice, rating scale, and partial credit Rasch analysis [Computer software]. Chicago: MESA Press; 1991.
- Linacre, JM. Optimizing rating scale category effectiveness. In: Smith, EV., Jr; Smith, RM., editors. *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press; 2004. p. 258-278.
- Ma EPM, Yiu EML. Voice Activity and Participation Profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language, and Hearing Research*. 2001; 44:511–524.
- Muthén, LK.; Muthén, BO. Mplus (Version 4.2) [Computer software]. Los Angeles: Author; 1998.
- Paul, DR.; Frattali, C.; Holland, AL.; Thompson, CK.; Caperton, CJ.; Slater, SC. *The American Speech-Language-Hearing Association Quality of Communication Life (QCL) Scale manual*. Rockville, MD: American Speech-Language-Hearing Association; 2004.
- Pett, MA.; Lackey, NR.; Sullivan, JJ. *Making sense of factor analysis*. Thousand Oaks, CA: Sage; 2003.
- Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks. *Medical Care*. 2007; 45(Suppl 1):S22–S31. [PubMed: 17443115]
- Rosen CA, Lee AS, Osborne J, Zullo T, Murry T. Development and validation of the Voice Handicap Index–10. *Laryngoscope*. 2004; 114:1549–1556. [PubMed: 15475780]
- Shen L. Quantifying item dependency. *Rasch Measurement Transactions*. 1996; 10(1):485.
- Smith, EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. In: Smith, EV., Jr; Smith, RM., editors. *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press; 2004. p. 575-600.
- SPSS. SPSS 12.0 for Windows [Computer software]. Chicago: Author; 2003.
- Sulica L. Contemporary management of spasmodic dysphonia. *Current Opinion in Otolaryngology Head and Neck Surgery*. 2004; 12:543–548. [PubMed: 15548915]
- Threats TT. Evidence-based practice research using a WHO framework. *Journal of Medical Speech-Language Pathology*. 2002; 10:xvii–xxiv.
- Wainer H, Thissen D. How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*. 1996; 15(1):22–29.
- Ware JE, Gandek B, Sinclair SJ, Bjorner JB. Item response theory and computerized adaptive testing: Implications for outcomes measurement in rehabilitation. *Rehabilitation Psychology*. 2005; 50:71–78.
- Wilson J, Webb A, Carding P, Steen IN, Mackenzie K, Deary IJ. The Voice Symptom Scale (VoiSS) and the Vocal Handicap Index (VHI): A comparison of structure and content. *Clinical Otolaryngology*. 2004; 29:169–174. [PubMed: 15113305]
- Wingate JM, Ruddy BH, Lundy DS, Lehman J, Casiano R, Collins SP, et al. Voice Handicap Index results for older patients with adductor spasmodic dysphonia. *Journal of Voice*. 2005; 19:124–131. [PubMed: 15766857]
- World Health Organization. *International Classification of Functioning, Disability, and Health: ICF*. Geneva, Switzerland: Author; 2001.
- Worrall L, McCooley R, Davidson B, Larkins B, Hickson L. The validity of functional assessments of communication and the activity/participation components of the ICIDH-2: Do they reflect what

really happens in real-life? *Journal of Communication Disorders*. 2002; 35:107–137. [PubMed: 12036147]

Wright, BD.; Masters, GN. Rating scale analysis. Chicago: MESA Press; 1982.

Yorkston KM, Baylor CR, Deitz J, Dudgeon BJ, Eadie TL, Miller RM, et al. Developing a scale of communicative participation: A cognitive interviewing study. *Disability and Rehabilitation*. 2008; 30:425–433. [PubMed: 17943516]

Yorkston KM, Baylor CR, Klasner ER, Deitz J, Dudgeon BJ, Eadie TL, et al. Satisfaction with communicative participation as defined by adults with multiple sclerosis: A qualitative study. *Journal of Communication Disorders*. 2007; 40:433–451. [PubMed: 17125785]

Yorkston KM, Klasner ER, Swanson KM. Communication in context: A qualitative study of the experiences of individuals with multiple sclerosis. *American Journal of Speech-Language Pathology*. 2001; 10:126–137.

How much does SD interfere with making a phone call for household business?
— Not at all
— A little
— Quite a bit
— A lot
— Extremely

Figure 1.
Sample item. SD = spasmodic dysphonia.

Figure 2.

A: Category response curves for a sample item using the original five-category response structure. Note that the curves for *quite a bit* and *not at all* are not distinctly separated from neighboring categories. B: Category response curves for the same item as in 2A using the recoded four-category response structure. Note that each of the four categories now presents a curve with a peak distinctly separate from neighboring categories.

Figure 3.

Scree plot for exploratory factor analysis of the original 141 items. The plot demonstrates the presence of one dominant factor by the steep slope between the first and second factor, and the minimal change in slope between subsequent factors.

Figure 4.

Distribution of 133 items along the logit scale according to item location. Each horizontal line represents one item. Sample items are provided to illustrate item content at different item location levels. Example participants are provided to illustrate the relationship between participants' scores (theta values) and items that are calibrated on the same logit scale. Participants are more likely to report interference on items located higher than their theta values, and they are less likely to report interference on items located lower than their theta values. Participant 2 has a more favorable score than Participant 1 (higher scores are better).

Figure 5.

The curve is the test information function for the final 133-item set. Measurement is most precise for trait regions under the highest regions of the curve. The horizontal line indicates an information level of 10, which corresponds approximately to a reliability coefficient of 0.90 (Embretson & Reise, 2000). Below the test information curve, the distribution of participants' theta values (scores) across the same logit scale is presented. Each vertical tick mark represents 1 participant. Comparison of the distribution of participants' scores with the test information curve reveals those participants for whom measurement was more precise.

Figure 6.
The standard error of measurement curve calculated as the inverse of the square root of the test information function.

Table 1

Participant demographic information.

Variable	%	<i>M</i>	<i>SD</i>	Range
Gender				
Women	77.9			
Men	22.1			
Age (years)		55.4	11.0	27–83
Race				
Caucasian	94.2			
Occupation				
Working full time	47.6			
Working part time	14.4			
Working as volunteer, homemaker, or care provider for family	21.6			
Not working/retired	14.9			
No response	1.4			
Education				
High school	9.6			
Attended some college/graduated from community college	30.8			
Bachelor's degree	29.3			
Graduate degree	30.3			
Time since SD diagnosis (years)		9.3	6.9	<1–32
Current treatment				
Botox	59.1			
No treatment	32.2			
Other treatment categories	8.7			
Voice self-rating compared with peers without SD				
No different/slightly different	13.9			
Somewhat different	22.1			
Very different	63.0			
No response	1.0			
VHI total scores		83.4	18.5	31–119 ^a

Note. SD = spasmodic dysphonia; VHI = Voice Handicap Index.

^aPossible range = 0–120, with high scores indicating greater handicap.

Table 2

Items with fit statistics outside the acceptable limits (0.77–1.30).

Item	Original mean square infit	Mean square infit when item was removed	Response categories with <10 responses	% of participants with no response/NA	Original item discrimination estimate
Items removed from item bank:					
Flirting or being romantic	1.89	1.89	NA	23	−0.08
Participating in religious activities	1.69	1.70	NA	25	0.14
Communicating with a pet	1.55	1.57	Extremely	32	0.35
Talking to a pet	1.55	1.57	Extremely	32	0.40
Training a pet	1.45	1.47	Extremely	34	0.44
Applying for a new job	1.34	1.36	Not at all	27	0.59
Giving a presentation to large audience	1.29	1.32	Not at all	32	0.71
Talking to supervisor about goals for work	1.29	1.31	NA	41	0.73
Items retained in item bank:					
Being polite when talking	1.96	NA	Extremely	<1	−0.29
Influencing people's decisions	1.51	NA	NA	2	0.48
If you needed to communicate during an emergency	1.43	NA	NA	0	0.42
Getting your point across when you are upset	1.37	NA	Not at all	1	0.59
Telling someone how to do something	0.71	NA	Not at all	<1	1.26
Having a casual conversation with someone you know well	0.65	NA	NA	0	1.38

Note. Items in the first section were removed from the item bank because the high rate of missing data suggested that these lifestyle questions were not applicable to large numbers of people and that these trends were expected to continue in future studies in other populations. Items in the second section were retained for further analysis because the reason for item misfit was either unclear or was possibly due to disorder-specific characteristics. The ideal item discrimination value is 1.0, with lower values suggesting lower discrimination. NA = not applicable.

Table 3

The 15 items with the best fit statistics (the best fit to the Rasch model).

Fit statistic	Item
1.02	Negotiating
1.02	Telling a story at a party
1.02	Reporting to your work supervisor or teacher
1.01	Making a phone call for household business
1.01	Communicating when you do NOT need to talk quickly
1.01	Receiving and discussing feedback about work or school projects
1.01	Correcting a child who is misbehaving
1.00	Telling a doctor or health care provider who you know about your symptoms or medical history
0.99	Talking to a store clerk who is in a hurry
0.99	Convincing a friend or family member to change his/her mind
0.99	Greeting someone by name
0.99	Asking for help from a child you are around frequently
0.98	Making appointments over the phone for personal services (e.g., haircut, dentist)
0.98	Giving directions to someone who is lost and has stopped you to ask for your help
0.98	Talking with someone while you are walking together outside

Note. A mean square infit value of 1.0 indicates ideal model fit.

Table 4

The 10 item pairs with the highest standardized residual interitem correlations and the 10 item pairs with the lowest standardized correlations.

<i>z</i> value	Item 1	Item 2
Item pairs with highest residual standardized correlations		
4.0	Does your condition interfere with teaching a young child a new skill?	Does your condition interfere with telling a story to a child?
3.9	Does your condition interfere with letting someone you live with know what you need at the store?	Does your condition interfere with communicating with people you live with about things that need to get done around the house?
3.8	In general, does your condition interfere with taking part in communication activities that are important to you?	In general, does your condition interfere with communicating about serious topics?
	Does your condition interfere with making comments to family or friends about a TV show or movie you are watching together at home?	Does your condition interfere with having a conversation with the TV or radio on at home?
	In general, does your condition interfere with casual conversations?	Does your condition interfere with making small talk?
	If you needed to communicate during an emergency, do you think your condition would interfere with this?	If you needed to give important information over the telephone in an emergency, do you think your condition would interfere with this?
3.7	Does your condition interfere with having a casual conversation with someone you do not know very well?	Does your condition interfere with communicating at social gatherings where you do NOT know most of the people?
	If you were with someone you know and needed to ask them for help right away, how much would SD interfere with doing this?	If you needed to communicate during an emergency, do you think your condition would interfere with this?
	Does your condition interfere with communicating with people who live with you?	Does your condition interfere with communicating at home?
	In general, does your condition interfere with communicating about serious topics?	In general, does your condition interfere with communicating when you do NOT need to talk quickly?
Item pairs with lowest residual standardized correlations		
-2.6	Does your condition interfere with making plans for an outing with family or friends?	Does your condition interfere with communicating at social gatherings where you do NOT know most of the people?
	Does your condition interfere with starting a conversation with someone you know?	If you needed to communicate during an emergency, do you think your condition would interfere with this?
	Does your condition interfere with asking for help from a child you are around frequently?	Does your condition interfere with communicating in situations where there is a large group of people?
-2.7	Does your condition interfere with having a conversation with someone you know about a book, movie, show, or sports event?	Does your condition interfere with talking to a taxi driver or bus driver about where you need to go?
	If you needed to give important information over the telephone in an emergency, do you think your condition would interfere with this?	Does your condition interfere with chatting with coworkers or colleagues at work or school?
-2.8	Does your condition interfere with having a conversation with someone you know about a book, movie, show, or sports event?	If you had to ask for help from a stranger when you are in a hurry, how much would SD interfere with this?
	In general, does your condition interfere with having a conversation in a quiet place?	If you needed to communicate during an emergency, do you think your condition would interfere with this?
	Does your condition interfere with having a conversation with someone you know about a book, movie, show, or sports event?	If you needed to give important information over the telephone in an emergency, do you think your condition would interfere with this?
	Does your condition interfere with asking for help from a family member?	Does your condition interfere with using the telephone at work?
-3.0	Does your condition interfere with interfere with arguing your points?	Does your condition interfere with using the telephone at work?

Note. Correlations with Fisher's *z* values greater than 2.0 suggest local dependence of items, and removal of one of the items in each pair should be considered (Shen, 1996).