

The effect of noise on relationships between speech intelligibility and self-report measures in tracheoesophageal speakers

Devon Sawin Otero

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington

2014

Committee:

Tanya Eadie

Kristie Spencer

Kathryn Yorkston

Program Authorized to Offer Degree:

Medical Speech-Language Pathology

©Copyright 2014  
Devon Sawin Otero

University of Washington

Abstract

The effect of noise on relationships between speech intelligibility and self-report measures in tracheoesophageal speakers

Devon Sawin Otero

Chair of the Supervisory Committee:  
Associate Professor Dr. Tanya Eadie  
Speech and Hearing Sciences

As clinicians and researchers, it is our ultimate goal to improve our patients' quality of life. This goal is achieved through reliable measurements such as speech intelligibility, which is a standard assessment of a patient's impairment level. Attempts to correlate this measurement with a patient's daily communication outside the clinic have been weak or uncertain. In this study we explore the correlation between speech intelligibility and self-report measures in a population of head and neck cancer patients, alaryngeal tracheoesophageal speakers (TEP). Participants: 24 individuals using TEP, 66 naïve listeners who performed intelligibility ratings of the speakers in quiet and in noise. The strength of these relationships was compared across the quiet and noise conditions. There was a weak correlation ( $r = 0.201$  and  $r = 0.003$ ) between speech intelligibility in quiet and self-report measures. A slightly stronger correlation ( $r = 0.435$  and  $r = 0.311$ ) was found between speech intelligibility in noise and self-report measures. The results of the study suggest that intelligibility in noise is a better predictor of self-rated communication function than intelligibility in quiet.

## INTRODUCTION

One of the primary treatments for head and neck cancer, specifically laryngeal cancer, is a total laryngectomy (TL). After TL, the larynx is removed, and the trachea is cut and sutured at the base of the neck, creating an opening called a tracheostoma. The procedure results in an altered airway and a need for a new alaryngeal voice source, as well as leaving the individual to cope with other physical, social, and psychological consequences (Doyle & Keith, 2005).

Alaryngeal communication is achieved through use of electrolaryngeal (EL), esophageal (ES), or tracheoesophageal prosthetic (TEP) restorative speech. EL speech is produced by placing an artificial laryngeal device against the neck or cheek and articulating the sound transmitted by the artificial laryngeal device through the tissues. In contrast, both ES and TEP speech are produced using sound sources within the body. To produce ES, air must be injected below the pharyngoesophageal (PE) segment and then released back up, causing the upper esophageal sphincter tissue to vibrate and become the sound source. The individual then articulates similar to the method used by typical speakers. Finally, TEP speech is acquired after a puncture is surgically created between the posterior wall of the trachea through the anterior wall of the esophagus. A tracheoesophageal prosthesis is then inserted within the puncture, allowing air to move from the trachea to the esophagus when the tracheostoma (stoma) is occluded, causing the upper esophagus sphincter to vibrate and create voicing. Similar to ES, the sound emitted from the upper esophageal sphincter is modified by the articulators to produce

speech. However, unlike ES, the respiratory support for the sound source in TEP speech is provided by the lungs, which offers advantages in terms of air flow and pressures needed for longer, fluent speech (Doyle & Keith, 2005).

In the past decade, TEP speech has been considered the 'gold standard' in post-laryngectomy voice rehabilitation (Kazi, Sayed, & Dwivedi, 2010). Advantages of using a voice prosthesis include: nearly immediate voice production following surgery, the possibility of sustained speech (i.e., beyond the limits of ES), the ability to reverse the surgery and use another speech method, and complication rates that are relatively low (Kazi et al., 2010). In addition, TEP speech has been shown to be more similar to laryngeal speech than other alaryngeal modes on parameters such as fundamental frequency, speaking rate, and maximum phonation time (Kazi et al., 2010). Finally, TEP speech appears to be a preferred method of alaryngeal speech in that it is considered more intelligible and fluent than its counterparts, ES and EL speech (D'Alatri et al., 2012).

Despite increased intelligibility among TEP speakers, there is some controversy about whether speakers who use this method actually report better outcomes in everyday contexts than other alaryngeal speakers (Doyle & Keith, 2005). For example, Law and colleagues (2009) examined relationships among speech intelligibility, acceptability, and communication-related quality of life (QOL) in 49 alaryngeal speakers. They found that speech intelligibility and communication-related QOL was poorest for ES and EL speakers, and best for a subgroup of speakers who used pneumatic devices (a specific type of artificial larynx). They noted that while TEP speakers demonstrated the second highest

speech intelligibility and acceptability scores in their study, they showed the best functional QOL outcomes. These results suggest that high speech intelligibility does not always translate to high QOL. This result is important because it suggests that a communication partner's ability to understand a speaker in quiet does not always predict how well that person perceives his or her own communication abilities in everyday contexts.

One reason past research may have shown uncertain relationships between intelligibility and self-reported outcomes may be due to the approach used in measuring speech outcomes. Specifically, typical listener-rated measures used in research (e.g., measures of intelligibility, acceptability etc.) are most often assessed under quiet conditions. In contrast, self-report outcomes such as voice-related QOL scales (e.g., Jacobson et al., 1997) and the *Communicative Participation Item Bank* (Baylor et al., 2013) often consider background noise when assessing communication because it is regularly encountered in everyday environments. Consequently, the overall purpose of this study is to examine relationships between speech intelligibility and patient-reported outcomes, and to determine whether the addition of noise to a speech intelligibility assessment will increase the strength of the relationships among these measures. Before these relationships can be understood, a review of typical outcomes after TL needs to be summarized.

#### *Outcomes after Total Laryngectomy*

*Speech intelligibility.* One common measure of alaryngeal speech with a long history of use in the clinical setting includes ratings of intelligibility performed by both clinicians and unfamiliar listeners (Bennett & Weinberg, 1973; Law, Ma, & Yiu,

2009; Miralles & Cervera, 1995; Tardy-Mitzell, Andrews, & Bowman, 1985). Speech intelligibility has been defined as the percentage of speech items that are correctly identified by the listener (Hillman, Walsh, & Heaton, 2005). Intelligibility in TEP speakers can be impacted by poor voice quality, altered speech rate, as well as the noise associated with improper timing and occlusion of tracheostoma (van As, Koopmans-van Beinum, Pols, & Hilgers, 2003). TEP voice quality, while more natural and fluent than ES, is still deviant from “normal” or laryngeal voice quality (van As et al., 2003). Low speech intelligibility may influence communication abilities of an alaryngeal speaker and have a negative impact on conversation (Law, Ma, & Yiu, 2009). For this reason, a measure of intelligibility may be helpful in gauging the success of one’s alaryngeal voice rehabilitation.

The intelligibility of TEP speakers has been shown at least equal to ES and better than EL speech, implying that TEP speech is one of the best options for voice rehabilitation for people who have had a laryngectomy (Tardy-Mitzell, Andrews, & Bowman, 1985). In one study, 46 unfamiliar listeners understood 15 TEP speakers on average 93% of the time (Tardy-Mitzell et al., 1985). Intelligibility judgments were based on transcription of randomized word lists consisting of 50 items. This result indicates that TEP speakers are considered to be highly intelligible to unfamiliar listeners, the type of listeners they would be most likely to encounter on a day-to-day basis (Tardy-Mitzell et al., 1985).

Recently, Eadie, Day, Sawin, Lamvik and Doyle (2013) examined speech intelligibility of 25 laryngectomized individuals (20 male, 5 female) who used a variety of alaryngeal speech methods, including 16 TEP speakers. Thirty-three

listeners with no prior experience or exposure to alaryngeal speech judged TL speakers intelligibility by transcribing sentences from the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Tice, 1996). The mean speech intelligibility score for all speakers was 89.74% (SD = 8.90%). The mean intelligibility score for the 16 TEP speakers was 92.39%, with a range of 72.88-99.57%. The two ES speakers' mean was 96.58% (slightly higher than TE). The 7 EL speakers had a mean of 81.74% intelligibility, the lowest of the 3 methods of alaryngeal speech.

Together, these results show that TEP speakers usually demonstrate high intelligibility scores as a group, but that there is variability within the group. Listener-rated intelligibility is important to consider when measuring post-laryngectomy outcomes because it may influence the way a communication partner interacts with a TE speaker, resulting in positive or negative changes to the speaker's QOL. However, to understand its impact, we first need to examine how an alaryngeal speaker judges his or her own speech in everyday settings. To do this, we must examine self-reported outcomes in this patient population.

*Self-report measures: Voice-related quality of life and communicative participation.* After a total laryngectomy, individuals not only need some type of restorative communication method, but the surgery also leaves the patient with psychological, nutritional, social and physical ramifications of their now altered anatomic structures (Terrell, Fisher & Wolf, 1998). Due to the effects of these changes, individuals often report a severe impact on quality of life (QOL) (Sherman et al., 2000). Per the World Health Organization, QOL is defined as "the individual's perception of their position in life, in the context of culture and value systems in

their life, and in relation to their goals, expectations, standards, and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychosocial state, level of independence, social relationships, and their relationship to salient features of the environment" (WHOQOL Group, 1995, p. 1405).

This definition reflects the multifaceted nature of QOL, and the wide variety of factors that influence it. By standard, QOL measures may be disease-specific (e.g., relating to all symptoms of head and neck cancer) or symptom-related (e.g. voice-related QOL: measuring voice-related symptoms of various voice disorders). Because disease-specific QOL measures may only ask one or two specific questions regarding certain functions, a voice- or speech-related QOL measure is better suited to capture QOL as it relates to the voice symptoms experienced by a person using alaryngeal speech.

One voice-related QOL measure that is commonly used in clinical practice is the *Voice Handicap Index* (VHI; Jacobson et al., 1997). The VHI assesses the impact of a voice disorder on an individual by means of a self-administered questionnaire. Individuals rate their perceived voice handicap across three dimensions: functional, emotional, and physical. There are 30 items on the VHI; scores range from 0 (no voice handicap) to 120 (high voice handicap). This measure was validated using participants exhibiting a variety of voice disorders, including laryngectomy. There is also a validated short form with 10 items (VHI-10; Rosen et al., 2004) that has been used with individuals who have undergone TL (Oridate et al., 2009). Therefore, the VHI (and VHI-10) serve as viable methods of measuring voice-specific outcomes in this population.

While the TEP method provides a patient with a laryngectomy the means of generating sound needed for speech production, the use of TEP speech has been found to adversely affect the quality of communication interactions and voice-related QOL (Day & Doyle, 2010). In a study by Evans, Carding, and Drinnan (2009), the VHI was administered to a group of male patients (n=53) who had undergone TL. The mean score of the men using surgical voice rehabilitation (i.e., TEP speech) (n=26) was 44.7, which indicated a moderate voice handicap. Similarly, Azevedo, Montoni, Fliho, Kowalski, and Angelis (2011) showed that patients using TEP speech after TL (n=17) reported scores that reflected a mild to moderate voice handicap (mean=31). More recent studies have used the short form of the VHI, the VHI-10, to assess similar outcome domains (Rosen et al., 2004). For example, Oridate et al. (2009) found laryngectomees (n=27) to have a mean VHI-10<sup>1</sup> score of 11.26 (SD=7.17), which also represents a mild to moderate handicap.

Another type of patient-reported outcome, similar to voice-related QOL, is communicative participation. Communicative participation is defined as “taking part in life situations in which knowledge, information, ideas or feelings are exchanged” (Baylor et al., 2013, p.1190). Patients with communication disorders, in this case voice disorders, are at risk for limited communicative participation if they are unable to “successfully navigate communication situations in the real world and therefore fulfill their life roles” (Baylor et al., 2013, p.1191). While the concept of communicative participation is reflected in several voice-related QOL instruments (including the VHI), the construct is intertwined with other dimensions such as

---

<sup>1</sup> Scores on the VHI-10 range from 0 (normal) to 40 (most severely limited) voice handicap

emotional, physical, and social functioning. As a result, voice-related QOL instruments do not solely measure communicative participation and do not allow us to determine contributions of other variables to participation.

Communication participation can be measured specifically by the *Communicative Participation Item Bank (CPIB)*. The CPIB was designed to measure how a condition, such as a total laryngectomy, interferes with communication in a wide range of speaking situations (e.g., on the phone, in a restaurant, in noise, in a large group) (Baylor et al., 2013). One advantage of the CPIB is that the items include other variables that affect participation, such as environmental features (e.g., communication partners, background noise) that may impact an individual's ability to communicate in everyday environments (Baylor et al., 2013). It also includes items that are sensitive to functional speech difficulties, as opposed to being only focused on voice-related function, which is one limitation of an instrument such as the VHI. The CPIB was developed on a basis of literature review, qualitative studies of people with communication disorders including patients treated for head and neck cancer, including those who had undergone TL. Head and neck cancer patients and alaryngeal speakers have difficulty participating and communicating in their daily lives (Baylor et al., 2011).

The results of all of these studies reveal the negative impact of voice- and speech-related symptoms on functional communication outcomes in alaryngeal speakers, including those who use TEP (Eadie et al., 2014). One hallmark of these studies is the wide range of scores reported by the participants. For example, in the study by Evans et al. (2009), the VHI scores of laryngectomees ranged from 4 to 106.

A recent study by Eadie et al. (2014) investigated relationships between self-reported QOL and the CPIB in a broad group of participants with head and neck cancer. Among this group were 122 individuals who had undergone TL, including 66 individuals who primarily used TEP. The TEP speakers reported on average a score of 19 (SD = 7) for the VHI-10, and on average a score of 0.20 (SD = 0.82) for the CPIB. In this context, a score of 0 on the CPIB represents the mean of the sample used for item calibration, and higher scores are better (see the methods section for more detailed information about these scales). In contrast, speakers who also were treated for HNC but who used natural speech reported an average VHI-10 score of 12 (SD = 8) and an average CPIB score of 0.55 (SD = 1.04). The results show that on average, TEP speakers report moderate voice handicap or moderately restricted communicative participation compared to those who did not have a TL but who were also treated for HNC. While a variety of factors affect communication in everyday settings, one potential predictor includes how much is understood by the communication partner (i.e., speech intelligibility). These relationships will be examined next.

#### *Relationships between Speech Intelligibility and Self-Report Measures*

Some studies in the head and neck cancer literature have examined the relationship between speech intelligibility or voice quality with self-report measures, such as head and neck cancer specific- or voice-related QOL (Meyer et al, 2004; Eadie & Doyle, 2004). In general, these studies have found weak to moderate relationships between these measures.

For example, Meyer et al. (2004) explored whether there was a strong correlation between speech intelligibility and QOL in 62 head and neck cancer survivors. The 62 survivors were separated into subgroups that included 16 laryngectomees and 48 non-laryngectomees. Speech intelligibility was measured with word and sentence intelligibility portions of *Assessment of Intelligibility Dysarthric Speech* (Yorkston & Beukelman, 1981). Three experienced speech-language pathologists transcribed the words and sentences to derive a mean intelligibility score for each speaker. Four self-reported measures were completed, including disease-specific QOL scales. The results revealed long-term head and neck cancer survivors (post 5 years) continued to have deficits in QOL and speech intelligibility. Interestingly, for the 48 members of nonlaryngectomy subgroup, sentence intelligibility scores were significantly associated with the QOL domains of speech ( $p=0.005$ ). However, in the TL subgroup, there were no significant association between sentence intelligibility and QOL speech domains ( $p=0.08$ ), suggesting there may be different relationships depending upon the patient population.

A previous study in our lab (Eadie et al., 2013) investigated whether there was a correlation among alaryngeal speech intelligibility with disease-specific and voice-specific QOL measures for 25 laryngectomees. The results of this study revealed weak correlations among the measures. The correlation between speech intelligibility and the VHI-10 was weak ( $r = 0.042$ ), as was the correlation between speech intelligibility and one item measuring self-rated speech understandability on a disease-specific scale (UW-QOL;  $r = 0.222$ ). These results contrast those reported

by Meyer et al (2004) for their nonlaryngectomy group. However, despite the significant correlation reported by Meyer and colleagues, it is difficult to determine the strength of this relationship because statistical significance was the only reported value.

The only study investigating relationships between intelligibility and self-reported communication function (i.e., beyond voice-related QOL) in the TL population was performed by Law, Ma, and Yiu (2009). Using a sample of Chinese alaryngeal speakers (n=49), the authors collected speech samples using the *Cantonese Sentence Intelligibility Test* (Lo, 1999), based on the *AIDS* (Yorkston & Beukelman, 1981) protocol and a self-report measure called the *Communication Activity and Participation After Laryngectomy* (CAPAL) QOL questionnaire. Six untrained listeners provided intelligibility and speech acceptability ratings. Results indicated that the laryngectomees with the highest speech intelligibility also scored better on the communication-related participation measure, which is contrary to the findings by Meyer et al. (2004) for their laryngectomy group. Unfortunately, Law et al. (2009) did not report the strength of the relationship between their measures, and so results are difficult to evaluate. In addition, results need replication in English-speaking adults because the tonality of the Chinese language could significantly and differentially impact results.

Beyond the TL population, results show that intelligibility measured by listeners and self-reported measures have uncertain results. In the motor speech literature, researchers have similarly shown weak relationships between measures of speech intelligibility with self-report measures of communication success. For

example, Donovan et al. (2008) reported a weak correlation ( $r = .35$ ) between speech intelligibility measured by unfamiliar listeners with self-reported communicative effectiveness in individuals with Parkinson's Disease. McCauliffe, Carpenter, and Moran (2010) similarly reported weak correlations between intelligibility and communicative effectiveness in individuals with chronic dysarthria secondary to traumatic brain injury. Collectively, these studies suggest that measures of intelligibility do not strongly predict how well an individual perceives his or her communication success in everyday settings, and that these results appear to be similar across patient populations.

*Relationships between Speech Intelligibility and Self-Report Measures: The effect of noise?*

In all of the reported studies examined above, researchers measured speech intelligibility in a quiet environment. Yet, in daily life, events often occur in sub-optimal listening conditions, which may negatively affect a communication partner's ability to process the speech signal during a communication exchange. Presence of noise adversely affects normal speech intelligibility (Van Engen & Bradlow, 2007). How background noise affects speakers with speech and voice disorders has had limited study (e.g., Bunton, 2006; McColl, Fucci, Petrosino, Marin, & McCaffery, 1998; Tjaden, Sussman, & Wilding, 2014). However, measuring speech intelligibility in adverse conditions has been suggested as a future area of research (Yorkston, Hakel, Beukelman, & Fager, 2007).

For example, in the dysarthria literature, one study found that for three speakers with a variety of dysarthria types, background noise appeared to

differentially penalize dysarthric speakers more than typical speakers (McAulliffe, Schafer, O'Beirne, & Lapointe, 2009). For alaryngeal speakers, one study examined the intelligibility of ES and EL speech as well as typical speech in three signal-to-noise ratio conditions: quiet, + 3 dB, and -1 dB (Holley, Lerman, & Randolph, 1983). They found that intelligibility for all 3 speech types decreased as the signal-to-noise ratios worsened (i.e., as the conditions became “noisier”).

TEP speech intelligibility in noise was investigated by McColl and colleagues (1998). In that study, McColl et al. (1998) recorded speech samples from one typical and one TEP speaker, who were rated as superior by three speech-language pathologists. The speech samples included recordings of two sentences from a reading passage. Fifty listeners then were asked to subjectively judge how well they understood the speech sample using a rating scale. The speech samples were presented at nine signal-to-noise ratio levels, with the noise being multi-talker babble. The results of the study revealed that the TEP speaker was significantly less intelligible than the typical speaker, and that the TEP speaker had significantly lower intelligibility in noise. The results also found a noteworthy interaction: the TEP speaker appeared to be more significantly affected in the presence of noise than the typical speaker. These differences appeared to be greatest around a +6 dB signal to noise ratio. However, the results of this study must be interpreted with caution since it only involved one TEP speaker and intelligibility was judged by using a subjective rating scale with relatively low interrater reliability.

### *Purpose of the Study*

The results of all of these studies appear to suggest that individuals with voice or speech disorders may be more susceptible to speech degradation in noise than typical speakers. This is an important factor to consider because most TEP speakers report difficulty communicating in noise. For example, noise was the “biggest limiting factor,” according to laryngectomy participants in Baylor et al.’s (2011) study. The laryngectomy participants in that study reported that white noise was regarded as a “physical barrier” because it was particularly challenging to speak loudly and clearly enough to be heard. Other researchers have also reported background noise is a problem. For example, Op de Coul et al. (2005) found that 50 out of 80 (63%) of their laryngectomy patients reported “serious problems” with speaking in noisy environments. As a result, self-reported outcome measures that take noise into account, such as the VHI-10 (with a few questions) or the CPIB (more so) may not strongly relate to measures of speech intelligibility, which are typically measured in quiet. Consequently, if we present the speech samples in noise and obtain speech intelligibility using that method, it is hypothesized that relationships between self-reported outcome measures and speech intelligibility in noise will be stronger.

The purpose of this study, therefore, is to investigate how speech intelligibility of TEP speech relates to self-report measures, including voice-related QOL and communicative participation, when speech intelligibility is measured in both quiet and in noise.

## METHODS

### *Overview and Design*

This study included three groups of participants: (1) 24 individuals (20 men, 4 women) who underwent TL provided speech samples using their primary method of communication, tracheoesophageal speech, as well as standard self-report measures that included the VHI-10 and the CPIB; (2) 48 naïve listeners (35 women, 13 men) performed intelligibility ratings of the speakers in quiet (with speaker scores derived from an average of 3 listeners) and; (3) 18 naïve listeners (15 women, 3 men) performed intelligibility ratings of the speakers in noise (with speaker scores derived from an average of 3 listeners). The reason for the larger number of listeners performing intelligibility rating of speakers in quiet was because they participated in a broader study that included judgments on all 3 types of alaryngeal speakers (Eadie et al., 2013). All participants were native English speakers, and none reported any other significant health conditions that might affect speech (beyond TL for that group of participants). The University of Washington Human Subjects Committee approved the procedures used in this study; all participants were paid for their participation.

This study used an exploratory correlational design to investigate the relationship between speech intelligibility in quiet and noise with voice-related QOL using the VHI-10 (Rosen et al., 2004), as well as communicative participation using the CPIB-10 (Baylor et al., 2009; Baylor et al., 2013; Eadie et al., 2014). The strength of these relationships was compared across the quiet and noise conditions.

*Speakers:* Speakers included 24 adults who underwent total laryngectomy secondary to cancer. The average age was 63.93 years with an age range of 39-86. Participants were at least one-year status-post total laryngectomy to allow time for adjustment and adaption of their new speech method. They were recruited from a national support group for individuals who underwent TL, called the International Association of Laryngectomees, as well as using professional contacts and advertisements within Washington State. The speakers used TEP speech as their primary method of speech. Exclusion criteria included individuals with previously altered upper aerodigestive tract, pre-existing dysphagia, speech impairments unrelated to the laryngeal cancer, or neurologic disorders.

*Listeners:* 66 naïve listeners were recruited from among the student population at the University of Washington and surrounding Seattle community. The average age of the listeners was 24.26 with an age range of 19-45. They were all 'naïve listeners,' characterized as individuals with no prior experience with or coursework in alaryngeal speech. All listeners passed hearing screening tests (25dB at octave frequencies of 250-4000Hz) and were native English speakers.

#### *Speaker Data Collection*

Speakers completed a consent form and then provided demographic information including age, gender, ethnicity, education level, living situation, date of TL and type of treatment(s), and primary speech method. They completed a set of self-report measures that include the *Communicative Participation Item Bank* (CPIB; Baylor et al., 2013), and the *Voice Handicap Index-10* (VHI-10; Rosen, Lee, Osborne,

Zullo, & Murry, 2004). These questionnaires were filled out at time of the speech recording, or returned within a two-month window of recording.

*Voice Handicap Index – 10 (VHI-10)*. The VHI-10 (appendix A) is a shortened version of the original 30-item VHI, combining three subscales measuring functional, physical, and emotional domains into one total scale (Rosen et al., 2004). The VHI-10 consists of 10 items measuring self-rated voice limitations by asking patients how often they have had each experience described by a set of statements (e.g. “My voice makes it difficult for people to hear me”). These 10 statements are summed to derive a composite VHI-10 score that ranges from “0” (minimal voice handicap) to “40” (significant/severe voice handicap). The VHI-10 has concurrent validity ( $r = .90$ ) with the original 30-item VHI (Rosen et al. 2004), and has a strong relationship ( $r = .82$ ) with another psychosocial voice instrument, the *Voice-Related Quality of Life* scale (Hogikyan & Sethuraman, 1999; Portone, Hapner et al. 2007).

*Communication Participation Item Bank-10*. The CPIB-10 (Appendix B) was used to measure communicative participation. The CPIB-10 is a 10 item short form derived from the CPIB (Baylor et al., 2009; Baylor et al., 2013). The CPIB is an instrument designed to measure a community dwelling individual’s participation in a wide range of speaking situations. It is applicable across multiple communication disorders and life situations. The final item bank has a total of 46 items including interference talking to unfamiliar people, ordering a meal at a restaurant and talking in groups of people. The CPIB was validated using Item Response Theory, which establishes psychometric properties for each item in the item bank. Baylor et al. (2013) reported the correlation between scores on the CPIB full item bank and the

short form (CPIB-10) to be .971 (significant at  $p < .001$ ), which suggests that the scores from the short form are strongly related to the scores generated using the full item set. The CPIB-10 item has a summary score generated by adding up the values of each item with a Likert scale (Not at all = 3, A little = 2, Quite a bit = 1, and Very much = 0) (Baylor et al., 2013). The total summary can range from 0 to 30, the high score being more favorable (e.g., higher level of participation).

*Speech Recordings.* Once the self-report measures were completed, the 24 speakers then provided speech recordings of sustained vowels, the Rainbow Passage (Fairbanks, 1960) and the *Sentence Intelligibility Test* (SIT; Yorkston, Beukelman, & Tice, 1996) in quiet. For the purposes of this study, only sentences from the SIT were used to assess speech intelligibility. An adapted method was used to assess intelligibility to increase the feasibility of the listening task. In this study, 6 sentences of increasing length were recorded (consisting of 5, 7, 9, 11, 13, and 15 words in length, resulting in 60 words per speaker). Speakers were allowed to hear the experimenter's model production of the sentence if they had any difficulty reading, but were encouraged to speak at their own comfortable rate.

Speech samples were recorded in a sound-treated room using a headset microphone (Shure PG-81, Sure Inc., Niles, Illinois; or AKG-C20, AKG Acoustics, Vienna, Austria) with a three-inch (offset) mouth-to-microphone distance connected to a preamplifier (M-Audio Fast Track Pro, Avid Technology Inc., Burlington, Massachusetts) and acquired on a laptop computer using a specialized sound card and acoustic software (Sona-Speech II, Model 3650, KayPENTAX, Montvale, New Jersey), as well as a portable digital audiotape recorder (TASCAM DAP1; TASCAM,

Montebello, California). All speech samples were recorded at a sampling rate of 44.1 kHz with 16-bit quantization. They were transferred to a computer and converted into WAV files using acoustic software (Sony Soundforge). Each sentence from the SIT (Yorkston et al., 1996) was edited using acoustic software.

Sentences were equated for peak amplitude (normalized) using the sound-editing software (Sony Soundforge). Then, one set of sentences from all the speakers was saved as the recordings in “quiet”. A second set of stimuli were then mixed with a 4-talker babble (one male, and three females; Audiotec of St. Louis) from the Sentences in Noise (SIN) test and saved as the second set of recordings in “noise”. Multi-talker babble was selected as the noise because previous research has found that meaningful speech competitors had a significantly more adverse effect on word recognition performance compared to non-meaningful competitors (e.g., white noise) (Sperry, Wiley, & Chial, 1997). In addition, multi-talker babble is representative of the most challenging adverse listening environment encountered in everyday speech communication situations (Gilbert, Tamati, & Pisoni, 2013).

Speech samples were mixed with multi-talker babble at a signal-to-noise-ratio (SNR) of +6dB for each sentence, with each sentence starting with 500 ms of noise alone, then the speech signal mixed with the noise, and finally 500 ms of noise at the end of each trial (Van Engen & Bradlow, 2007). The +6 dB SNR level was identified from the McColl et al. (1998) study, and was pilot-tested with a few listeners to ensure no ceiling or floor effects. This level is equivalent to ratio of 1.995 signal to noise. The WAV files were then entered into a software program (EcosWin,

Avaaz Innovations) that randomizes the presentation order of the speech samples and allows the listener to transcribe the speech sample (intelligibility).

### *Listener Procedures*

There were two groups of listeners in this study. Both groups performed similar procedures. However, one group was randomized to transcribe SIT sentences presented in “quiet”, and the other set transcribed SIT sentences presented at the +6 dB SNR (i.e., the “noise” condition).

Before the transcription task, both groups of listeners were provided instructions about the task. “You will be listening to adult speakers who have had total removal of their voice box due to cancer. These speakers are using a new method of speech called “tracheoesophageal speech”. We are interested in how well listeners can understand these speakers in both quiet and background noise. We will play some sentences, and we would like you to type out the words that you hear. You may listen to the sentences up to two times. Some of these sentences will be difficult to understand. Do your best, and guess when you need to. You may listen to each sentence 2 times.” They were presented several samples produced by speakers using TEP speech who were not otherwise included in the study to familiarize them with the general quality of TEP speech. For listeners in the noise group, they were familiarized with the speaker who they would eventually transcribe. The speech sample they heard included the 2<sup>nd</sup> sentence of the *Rainbow Passage*, “The rainbow is a division of white light into many beautiful colors” (Fairbanks, 1960). They were instructed that this same speaker would be presented in noise that included other speakers, such as what they might encounter at a party

(as found in the SIN test). This familiarization protocol was used to prepare the listener for the quality of TEP speech and the multi-talker babble.

Both groups of listeners were asked to transcribe 6 SIT sentences for 3 speakers, providing a score of speech intelligibility (% of words understood) for 18 sentences. The listeners transcribed the sentences using the protocol described by the *Assessment of Intelligibility of Dysarthric Speech (AIDS)* (Yorkston & Beukelman, 1981). Different speakers were randomly assigned to each listener (Hustad & Beukelman, 2002). The recordings were presented over headphones (Samson RH600), and listeners were asked to orthographically transcribe the sentences they heard into a software program on a desktop computer (EcosWin, Avaaz Innovations). The listeners were instructed to type exactly what they heard, and that they should make their best guess if uncertain (Hustad, 2006). Listeners were allowed to listen to the sentences in their entirety one time, and then again a second time, if necessary, providing a maximum of two exposures to the speech sample. Each speaker's intelligibility judgments were based on an average score derived from 3 listeners using a total word phonemic match model in scoring (Hustad & Cahill, 2003). Transcribed words were counted as correct if all phonemes included in the spelling match the target word, including homonyms and misspelled words. Analyses centered on the proportion of correct words compared to the possible words resulting from the transcription.

#### *Reliability of Listeners' Transcriptions*

Measures of intrarater reliability were not included in this study due to learning effects with presentation of a repeated sentence. However, to assess

interrater reliability of transcriptions for each set of 3 speakers evaluated by all listeners, inter-item correlations were calculated. For listeners in quiet, the mean correlation across all listeners was 0.68 (SD = 0.17). For listeners in noise, the mean correlation across listeners was 0.74 (SD = .14). These levels are consistent with prior research and are acceptable levels for further data analysis (Sussman & Tjaden, 2012).

### *Data Analysis*

The predictor variables in this study included speech intelligibility in quiet and in noise. The predicted variables included measures of VHI-10 (total score) and CPIB-10 (summary score). To calculate intelligibility scores for each speaker, the average of listeners' percent-correct transcribed words were determined (3 listeners per speaker). Total and summary scores for the VHI-10 and the CPIB-10 were then obtained by adding up all responses to items in each self-report measure for each speaker. To determine relationships between the variables, four Pearson's correlations were calculated. Specifically, relationships were examined between intelligibility in quiet and VHI-10, intelligibility in quiet and CPIB-10, as well as between intelligibility in noise and CPIB-10. The strength of relationships was then assessed using variance scores ( $R^2$ ).

## **RESULTS**

### *Demographics of Tracheoesophageal speakers.*

The 21 male and 3 female TEP speakers were on average 6.91 years (range 1-18 years) post-laryngectomy. Sixty-seven percent (n = 16) received radiation treatment with surgery, 75% had completed at least some college education, 79%

(n=19) lived with family, and the majority (92%) were white/Caucasian (n=22).

Additional demographic information is presented in **Table 1**.

**Table 1.** Demographics of Tracheoesophageal speakers

Characteristic	No. (%)	Mean (SD)	Range
Sex			
• Male	21 (87.5)		
• Female	3 (12.5)		
Ethnicity	22		
• White/Caucasian	(91.66)		
• American Indian/Alaskan Native	1 (4.17)		
• Other	1 (4.17)		
Age (years)		63.75 (10.06)	39-86
Living situation			
• Alone	5 (21)		
• With family	19 (79)		
Education			
• High school grad	5 (21)		
• Some college	11 (46)		
• College grad	6 (25)		
• Post-grad	1 (4)		
• Not reported	1 (4)		
Time since total laryngectomy (years)		6.91 (5.56)	1-18
Cancer treatment			
• Surgery alone	2 (8)		
• Surgery + radiation	16 (67)		
• Surgery + radio(chemo)therapy	5 (21)		
• Not reported	1 (4)		

### *VHI-10 Total and CPIB-10 Summary Scores*

Self-report measures from the TEP speakers included VHI-10 total and CPIB-10 summary scores. Overall, the speakers reported an average VHI-10 total score of 16.52 (SD = 7.02), indicating a moderate voice handicap. One speaker did not complete the VHI-10, so this total score was based on 23 speakers. Data from all 24 speakers were obtained for the CPIB-10. On average, they reported scores of 20.67 (SD = 6.17) for the CPIB-10 (higher scores are better; scores range from 0 to 30). Average scores and data for individual speakers are presented in Table 2.

### *Intelligibility Scores in Quiet and in Noise*

The mean (SD) speech intelligibility score for TEP speakers in quiet was 92.55% (6.11%); for speech intelligibility in noise, the mean (SD) was 68.82% (15.38%). Scores revealed an average decrease of 23.73% in intelligibility with the introduction of background noise. For 4 out of 24 speakers (M, N, P, SO), intelligibility did not appear to change (less than 5% decrease) in noise. Data from all 24 speakers are presented in **Table 2**.

**Table 2.** Measures across speakers, including intelligibility in quiet (mean %), intelligibility in noise (mean %), and self-report scores (VHI-10 and CPIB-10).

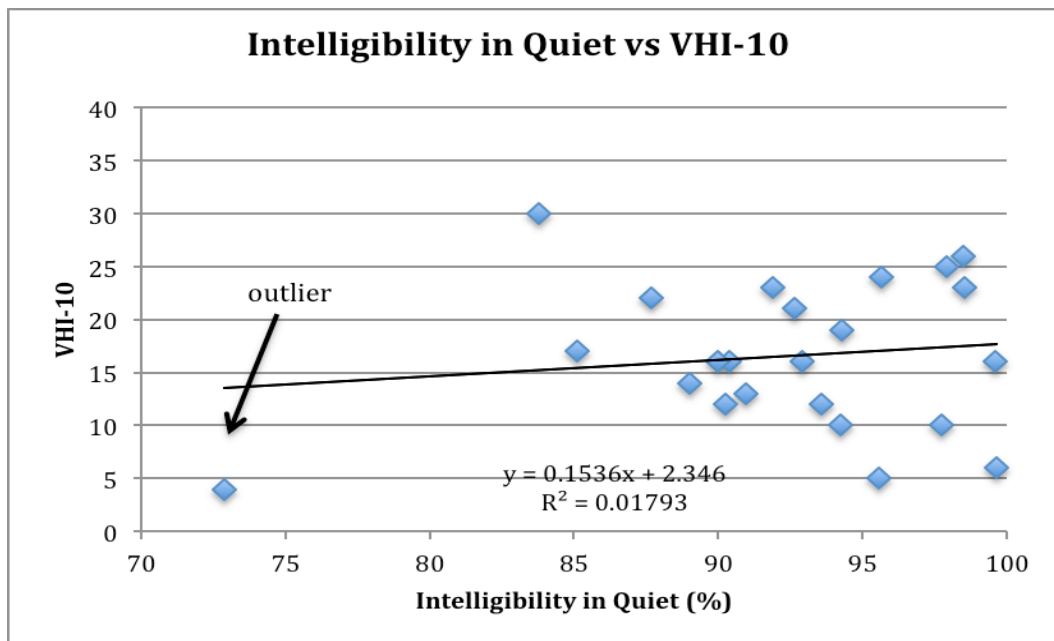
Speaker	Intelligibility in Quiet, Mean % (SD)	Intelligibility in Noise, Mean % (SD)	VHI-10 Total	CPIB-10 Summary
D	98.52 (4.88)	60.09(38.98)	23	15
G	91.87(14.13)	49.89(37.71)	23	10
GG	91.56(17.61)	77.84(28.95)	13	25
I	95.64(6.91)	72.15(28.90)	24	10
II	97.90(4.85)	73.79(23.08)	25	13
J	97.72(4.71)	62.06(35.15)	10	27
JJ	99.57(1.81)	55.81(33.10)	16	28
L	92.88(7.67)	60.60(27.39)	16	20
M	93.54(6.55)	90.69(10.07)	12	29
N	90.37(13.15)	88.70(16.49)	16	24
O	92.61(11.13)	83.85(16.68)	21	20
P	94.23(11.72)	93.79(11.78)	10	26
*Q	72.88(20.51)	48.31(26.68)	4	28
S	87.67(15.27)	45.47(32.04)	22	19
SC	95.56(12.94)	83.55(29.95)	5	30
SE	89.01(14.80)	59.97(34.71)	14	17
SG	94.29(13.17)	66.72(25.00)	19	20
SH	85.11(26.46)	53.07(40.49)	17	18
SM	90.25(11.97)	49.43(30.80)	12	27
SN	98.48(3.49)	65.61(36.92)	26	16
SO	89.95(21.29)	86.62(22.40)	16	20
SP	99.63(1.57)	91.31(16.67)	6	24
T	98.34(3.88)	74.90(26.54)	n/a	11
V	83.78(23.11)	57.48(21.94)	30	19
Average (SD)	92.55(6.11)	68.82(15.38)	16.52(7.02)	20.67(6.17)

\*Denotes the outlier speaker, whose data were removed from figures 2-7.

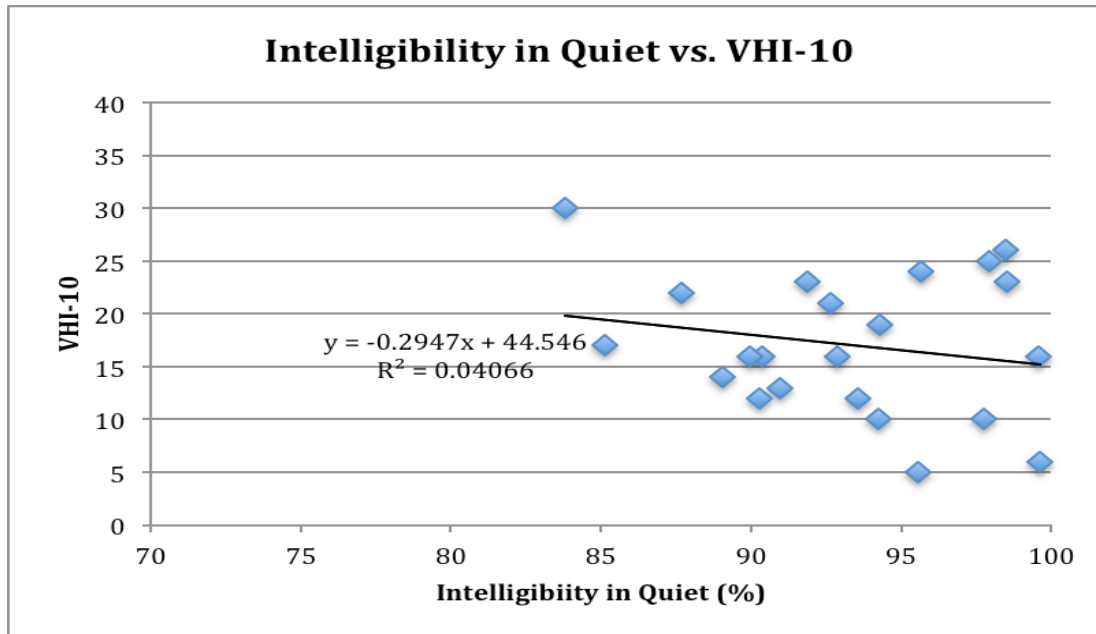
*Relationships between Intelligibility Measures and Self-Report Measures (VHI-10 and CPIB-10)*

*Relationships between intelligibility in quiet with VHI-10 and CPIB-10.* The relationships between intelligibility in quiet with both self-report measures (VHI-10 and CPIB-10) are shown in **Figures 1-3**. First, a scatterplot showing the relationship between intelligibility in quiet and VHI-10 scores revealed a spread of scores (see Figure 1). Note that **Figure 1** includes a speaker who was a significant outlier from the rest of group. A similar plot between intelligibility and CPIB-10 scores also revealed that the same speaker (Q) was also an outlier for those relationships. As a consequence, his data were removed for all subsequent analyses.

**Figure 1.** *Relationship between Intelligibility in Quiet vs. VHI-10 total scores. Each dot represents data for a single speaker. A line of best fit and variance scores are also reported on the graph below.*

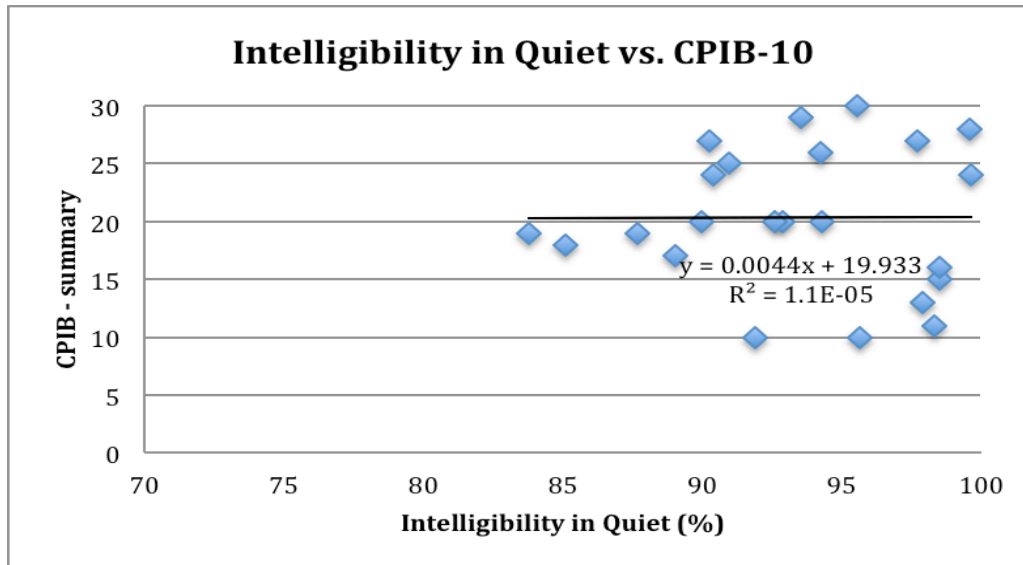


**Figure 2.** Relationship between Intelligibility in Quiet vs. VHI-10 total scores. Each dot represents data for a single speaker. A line of best fit and variance scores are also reported on the graph below.



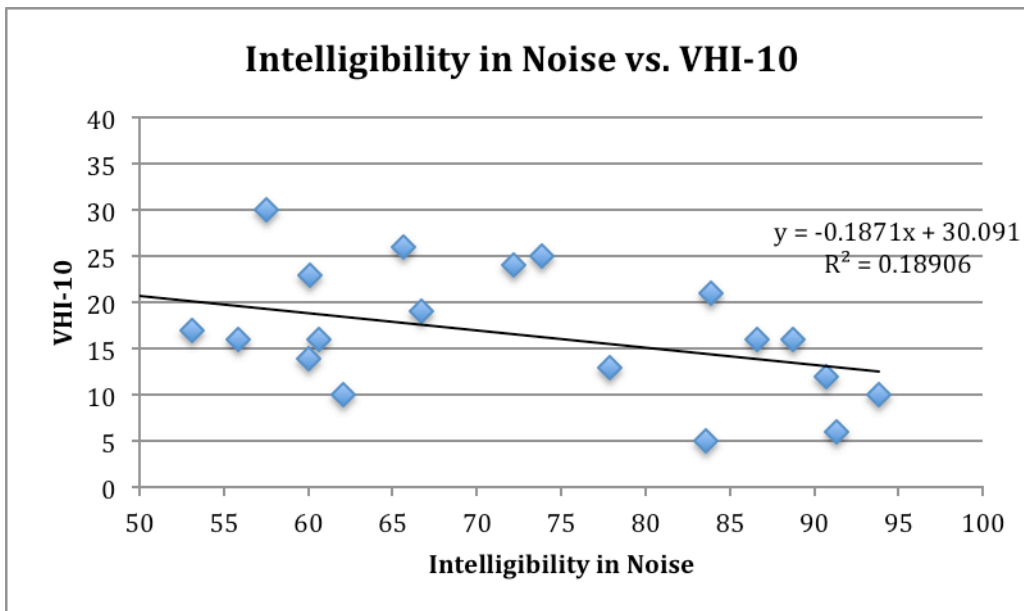
Figures 2 and 3 show the relationships between intelligibility in quiet and self-report scores for the remaining 23 TEP speakers. Overall, the relationships between intelligibility in quiet and VHI-10, as well as the relationships between intelligibility in quiet and CPIB-10 scores were weak in strength ( $r = -0.206$  and  $r = 0.0033$  respectively; see **figures 2 and 3**). The relationships were in the predicted directions; that is, as voice handicap increased, intelligibility scores decreased (see **figure 2**); similarly, as intelligibility increased, CPIB-10 summary scores also increased, although this was a very weak (almost nil) relationship (see **figure 3**).

**Figure 3** Relationship between Intelligibility in Quiet vs. CPIB-10 total scores. Each dot represents data for a single speaker. A line of best fit and variance scores are also reported on the graph below.

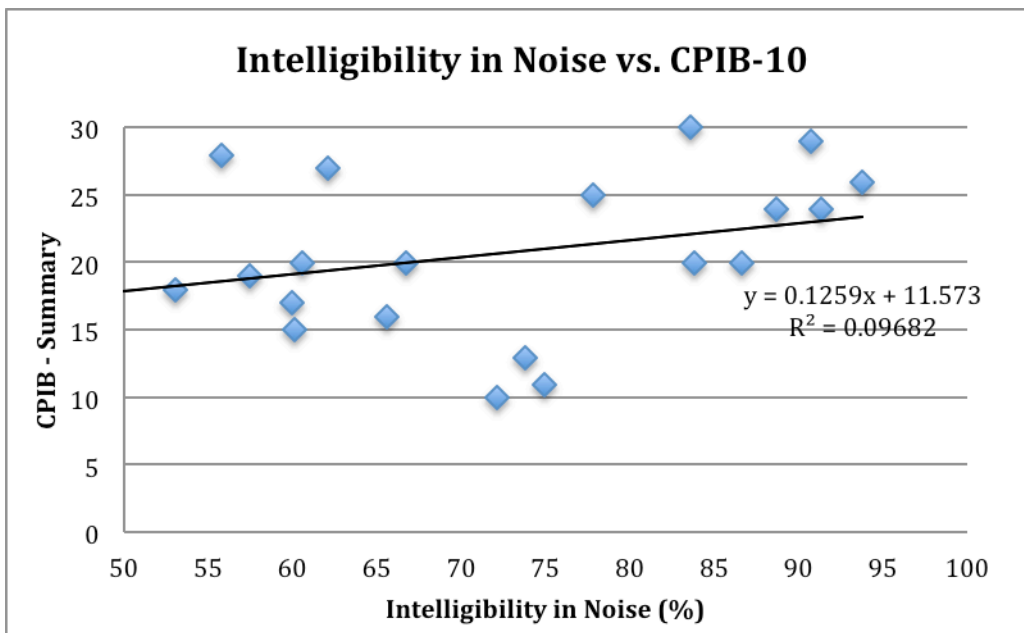


Relationships between intelligibility in noise with VHI-10 and CPIB. Intelligibility in noise scores were more strongly related to the VHI-10 and CPIB than relationships between intelligibility in quiet with the self-report scores. Intelligibility in noise and the VHI-10 showed a moderate, but statistically significant, relationship ( $r = -0.43$ ;  $p < .05$ ). Intelligibility in noise was also weakly related to CPIB-10 scores ( $r = 0.31$ ). The relationships were in the predicted directions; that is, as voice handicap increased, intelligibility scores decreased (see **figure 4**), and as CPIB increased, intelligibility scores increased (see **figure 5**).

**Figure 4.** Relationship between Intelligibility in Noise vs. VHI-10 total scores. Each dot represents data for a single speaker. A line of best fit and variance scores are also reported on the graph below.



**Figure 5.** Relationship between Intelligibility in Noise vs. CPIB-10 total scores. Each dot represents data for a single speaker. A line of best fit and variance scores are also reported on the graph below



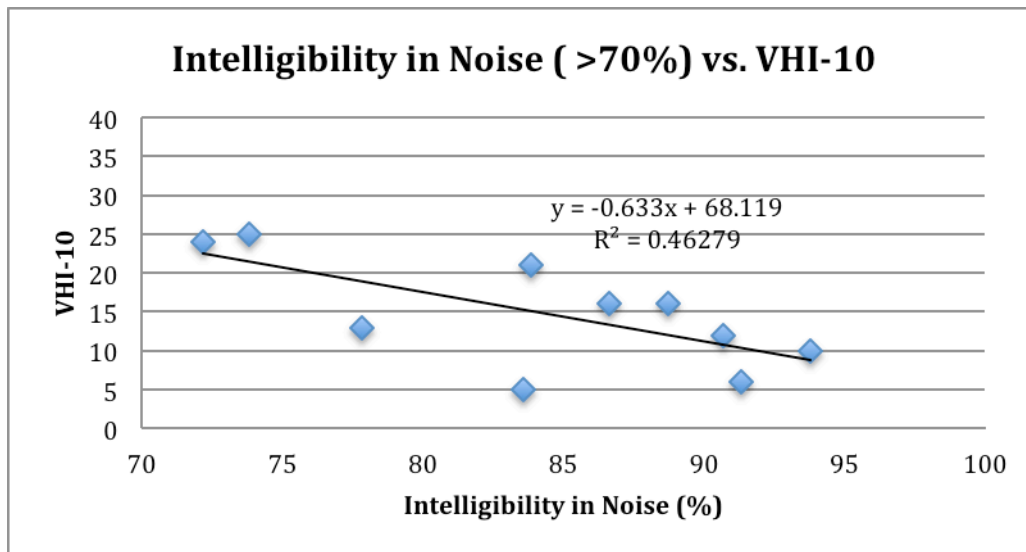
**Table 3.** Correlations between listeners' ratings of intelligibility in quiet (*Intell-Q*) and noise (*Intell-N*) with Self-Reported *VHI-10* and *CPIB-10* in 23 tracheoesophageal speakers.

	<b>Intell-Q</b>	<b>Intell-N</b>	<b>VHI-10</b>
<b>CPIB-summary</b>	0.003	0.311	<b>-0.768</b>
<b>VHI-10</b>	-0.201	-0.435	

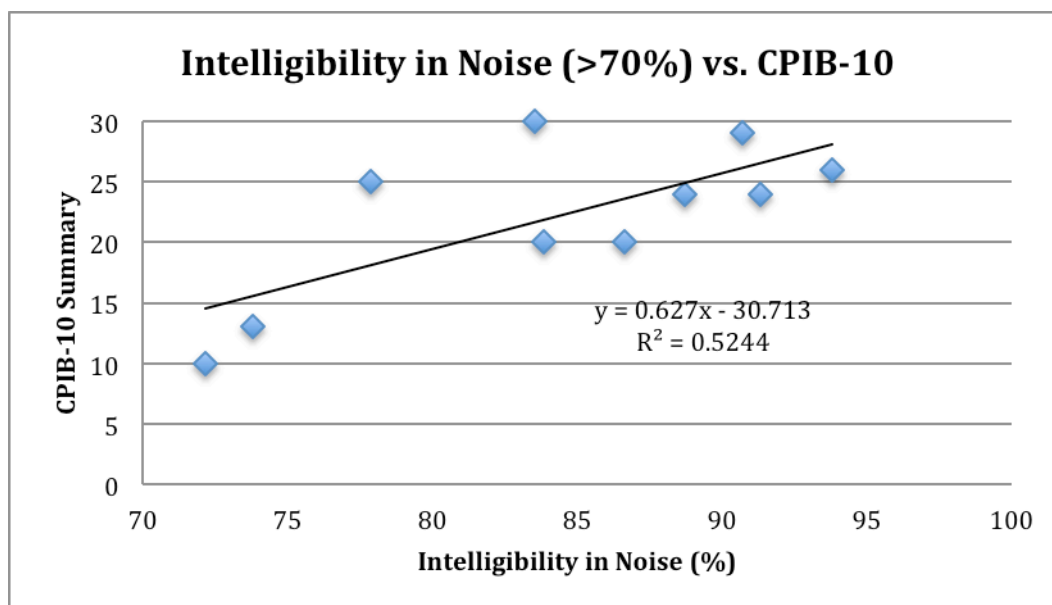
Relationships among the measures are summarized in **Table 3**. Note, as was predicted, the correlation between scores on the self-report measures (*VHI-10* and *CPIB-10*) was strong and statistically significant ( $r = -0.77, p < .05$ ).

After visual inspection of the data (scatter plots), a notable pattern was observed above 70% intelligibility in noise for both relationships with the *VHI-10* and the *CPIB-10* (**figure 6 and 7**). The data were re-analyzed excluding samples below 70% (criterion) to better describe relationships above this level of intelligibility. The relationships are shown in detail in the figures below.

**Figure 6.** Relationship between Intelligibility in Noise vs. VHI-10 total scores in speakers with greater than 70% intelligibility. Each dot represents data for a single speaker ( $n=10$ ). A line of best fit and variance scores are also reported on the graph below.



**Figure 7** Relationship between Intelligibility in Noise vs. CPIB-10 total scores in speakers with greater than 70% intelligibility. Each dot represents data for a single speaker ( $n=10$ ). A line of best fit and variance scores are also reported on graph below.



As observed in figures 6 and 7, the relationships between intelligibility in noise (above 70%) with the self-reported measures are much stronger than relationships for all speakers (**figures 4 and 5**). There was a moderately strong and statistically significant relationship shown between intelligibility in noise and VHI-10 scores ( $r = -0.67, p < .05$ ) for speakers with more than 70% intelligibility ( $n=10$ ). As the intelligibility in noise decreased, VHI-10 scores strongly increased (i.e., more voice handicap), as predicted. The relationship between intelligibility in noise and the CPIB-10 scores was also strong and statistically significant ( $r = 0.72, p < .05$ ) for speakers with more than 70% intelligibility ( $n=10$ ). As the level of intelligibility increased, communicative participation also strongly increased for these 10 TEP speakers.

## DISCUSSION

This study investigated the relationship between speech intelligibility with patients' self-reported voice handicap and communication outcomes. The main question was whether these relationships were stronger when intelligibility was measured in quiet, as is standard practice, or in noise, which might better reflect everyday life adverse conditions. Results revealed very weak relationships between intelligibility in quiet with the VHI-10 and the CPIB-10. As hypothesized, slightly stronger (but still weak) relationships were observed between measures of intelligibility in noise with both self-report measures. However, a subset of the data revealed a stronger pattern. Scores from speakers with intelligibility at 70% or more in noise predicted both VHI-10 and CPIB scores, with approximately 46% – 52% of the variance predicted. These results suggest that with the addition of

background noise, intelligibility scores could be a strong predictor of voice-related quality of life and communication outcomes in some tracheoesophageal speakers. How these results compare to previous findings, and what this means for both future research and clinical practice is discussed next.

*Self-report measures: Voice-related quality of life and communicative participation*

The sample of speakers who participated in this study first must be considered when interpreting the results. The sample of 24 tracheoesophageal speakers who provided self-report scores and speech recordings appeared to be representative of the laryngectomy population using TEP as their primary mode of communication. Demographics from this group are comparable to those reported by others (Eadie et al., 2012; Vilaseca, Chen, & Backscheder, 2006). For example, most subjects who participated were older Caucasian males, which is comparable to the laryngectomy populations who have participated in research in the past (Lundstrom, Hammarberg, & Munch-Wikland, 2009). Participants in this study were mostly recruited through support groups, and the majority (88%) had received radiation, lived with family (79%), and had some college education (75%). This is important to note as some of these factors could increase speech intelligibility and performance on self-report measures. For example, those who are involved in support groups, have higher education, or family support may be coping better or have better mental and physical health than those who do not choose to participate in research, and therefore may be more willing (and able) to participate in life events. Using education as a proxy for socioeconomic status, one must also consider that most of these individuals may have had good access to treatments and speech

rehabilitation. The results should therefore be interpreted with these demographics and potential biases in mind.

Overall, the results of this study revealed VHI-10 scores of moderate severity, 16.52 (7.02). These values are comparable to those reported by others who have studied individuals with TL, including a large number of TEP speakers. For example, 24 patients (56%) in a study by Lundstrom et al. (2009) reported a moderate voice handicap. Similarly, Evans et al. (2009) and Azevedo et al. (2011) reported mild to moderate voice handicaps in their groups of speakers who used TEP. In the Evans et al. study, 35 TEP participants who were predominantly male (male to female ratio, 6:1) reported a moderate voice handicap. Thus, the group recruited in this study appeared to function similarly to others studied in the literature.

In this study, the average CPIB summary score for the 24 TEP speakers was 20.67. This value is consistent with a larger independent group of 66 TEP speakers who completed the CPIB (entire item bank) in the study by Eadie et al. (2014). They reported an average theta value of 0.22, which is equivalent to a summary score of nearly 20 on the CPIB-10 (Baylor et al., 2013). Therefore, the subgroup of 24 TEP speakers who participated in this study appeared to function at similar levels of communicative participation as similar speakers in the larger validation studies. These results strengthen the external validity of the new tool. To put all of these scores into context, we can compare the average CPIB-10 summary score for these 24 TEP speakers with others who have completed the CPIB. For example, 73 individuals who were treated for head and neck cancer but who used natural speech scored about 22 on the CPIB (Eadie et al., 2014); in other words, those who had a

larynx but who underwent head and neck cancer treatments did better than those with a TEP, which is consistent with voice-related QOL outcomes (Oridate et al., 2009).

Finally, a strong correlation was found between VHI-10 and CPIB scores ( $r = -.77$ ), demonstrating a statistically significant relationship between voice-related quality of life and communicative participation in this group of TEP speakers. This relationship is similar to that reported by Eadie et al. (2014) in a group of 197 patients treated for different types of head and neck cancer ( $r = -.79$ ; CPIB full item bank vs. VHI-10). While this value is strong and shows the overlap between the measures, it may also reflect the population that was studied. In a group such as individuals who have undergone TL (primarily a voice, as opposed to a speech disorder), we might expect a strong relationship between these measures. Until the CPIB is compared to the VHI in a broader population with speech-focused disorders (e.g., individuals with oral or oropharyngeal-based cancers), the utility of the CPIB beyond voice-related QOL measures such as the VHI remains mostly untested.

#### *Speech Intelligibility in Quiet and in Noise*

In this study, speech intelligibility in quiet ranged from 72.88% to 99.57% (mean = 92.55%) across the speakers, which is consistent with previous studies (Eadie et al., 2013). For example, Tardy-Mitzell et al. (1985) reported that 46 unfamiliar listeners understood 15 TEP speakers on average 93% of the time (Tardy-Mitzell et al., 1985). As hypothesized, speech intelligibility in noise was lower than in quiet, similar to a previous study of other alaryngeal speakers (Holley

et al., 1983). Scores in that study ranged from 45.47% to 93.79% in noise (mean = 68.82%).

The drop in intelligibility (from 93% to 69%) in a +6dB SNR condition is a large decrease and may reveal the sensitivity of disordered speech to adverse conditions. For example, McAuliffe et al. (2009) showed that across similar noise conditions and with similar stimuli, 3 control adult male speakers did not show a decrease in intelligibility. Instead, their averages remained consistent (above 90%) when speakers were presented in a no noise and a +6dB SNR condition. However, similar to the present study, the 3 dysarthric speakers in their study were significantly affected in relatively low noise conditions (+6 dB SNR), and showed a large, significant decrease in intelligibility. These results support the contention that speakers with communication disorders, including those using TEP speech, may be differentially penalized in difficult listening environments.

Similar to the present study, McColl et al. (1998) also found that their one superior TEP speaker was adversely affected in noise. In particular, they noted large differences between the TEP speaker and a typical speaker's performances when noise was also presented at a +6dB signal to noise ratio. It is interesting to note that for 4 out of 24 speakers in this study (M, N, P, SO), intelligibility appeared to remain relatively constant from quiet to noise conditions. This result is similar to the control speakers in the study by McAuliffe et al. (2009). These results may indicate that there may be some TEP speakers who have speech characteristics more similar to typical speakers that may help listeners overcome the noisy conditions. Acoustic characteristics of these speakers need future study.

A common complaint of individuals who have had a laryngectomy is presence of noise (Baylor et al., 2011; Op de Coul et al., 2005) and its impact on intelligibility. While it may be assumed that more intelligible speakers in quiet would remain at higher intelligibility ratings in noise, there were some individual speakers in this study with greater than 95% intelligibility in quiet who dropped more significantly in intelligibility than those were less than 95%. Similar to speakers who could overcome noise factors, future studies should investigate characteristics of these speakers who are particularly vulnerable in adverse conditions. In addition to speaker factors, it is also important to consider how intelligibility measures are obtained. Factors related to the listener, including short-term memory, listener effort, listener reliability, and predictability of the sentences must also be considered. For example, in this study, sentences were selected at random for the *SIT* (Yorkston et al., 1996); yet, it is possible that some sentences were more predictable than others, which would allow listeners to use more top-down strategies to increase performance. This should be controlled in future studies. How listeners performed above and below certain levels of intelligibility will also be discussed when we examine relationships between intelligibility and self-report measures.

#### *Relationships between Intelligibility and Self-Report Measures*

This study investigated the relationship between listener-rated intelligibility in quiet and in noise with self-reported outcomes, as measured by the VHI-10 (Rosen et al., 2004) and the CPIB (Baylor et al., 2013) in TEP speakers. The relatively weak relationships found between intelligibility in quiet and VHI-10 ( $r = -$

0.201) and CPIB ( $r = 0.003$ ) were not entirely unexpected. Meyer et al. (2004) similarly found no significant association between sentence intelligibility and QOL speech domains in their group of laryngectomy patients. Likewise, Eadie et al. (2013) showed weak correlations between speech intelligibility and the VHI-10 ( $r = 0.042$ ). Finally, Donovan et al. (2008) reported a weak correlation ( $r = .35$ ) between speech intelligibility measured by unfamiliar listeners with self-reported communicative effectiveness in individuals with Parkinson's Disease.

Collectively, these studies suggest that an unfamiliar communication partner's ability to understand a speaker is not predictive of how a person with a communication disorder perceives everyday voice or communication function. Because these measures are derived from two different people (with different perspectives), one might anticipate that these relationships would be weak. A second reason for a weak relationship between measures may be that speaking (and listening) in quiet is not representative of daily life activities. Most speakers are asked to communicate with some kind of background noise or distraction present.

In this study, intelligibility in a +6 dB SNR noise condition showed slightly stronger relationships with voice handicap and communicative participation ( $r = -0.43$  and  $r = 0.31$ , respectively) than when the same relationships were examined for speech intelligibility in quiet. One possible reason for this increased strength in the relationship between measures is that speaking with noise in the background is more representative of TEP speakers' daily communication environments than speaking in quiet. In particular, both the CPIB (Baylor et al., 2013) and the VHI-10 (Rosen et al., 2004) ask specific questions about voice and communication in

background noise. This may explain the slight stronger correlation between intelligibility in noise with these self-report measures.

An interesting pattern was noted when examining relationships between intelligibility in noise scores with the self-report measures. When speakers who were more than 70% intelligible in noise were considered, relatively strong, linear relationships with self-report measures were found. Specifically, correlations increased to  $r = -0.67$  (VHI-10) and  $r = 0.72$  (CPIB).

Several reasons for the increase in strength of the correlations for these speakers may be postulated. First, one must consider the reliability of listeners' transcriptions below 70% intelligibility. An examination of interrater reliability values did not reveal a discernable pattern between intelligibility scores and reliability. All conditions appeared to be equally reliable. Second, one should consider strategies used by listeners at certain levels and how processes such as attention could affect scores. For example, Beukelman, Childes, Carrell, Funk, Ball and Pattee (2011) examined the relationship between listeners' ratings of attention allocation and speech intelligibility in speakers with amyotrophic lateral sclerosis (ALS). They found a strong negative correlation between perceived attention allocation and intelligibility ( $r = -.885$ ). Interestingly, they showed the highest ratings of attention allocation occurred for speakers with mean intelligibility scores between 75% and 85%. In contrast, attention allocation scores were much lower for speakers with lower (<70%) and higher (>85%) mean intelligibility. The authors concluded that these results may indicate that listeners try less hard when speech is either easy or very difficult to listen to. As in this study, when intelligibility was low

enough due to adverse conditions, listeners may not have paid as much attention because it was too frustrating or unproductive to do so. This could result in an unpredictable relationship between scores below 70% intelligibility with self-report measures.

A final reason why relationships remain variable between intelligibility (even in noise) and self-report measures is that one must consider that these types of measures are complementary. They measure different aspects of communicative function (e.g., impairment vs. participation vs. quality of life). In other words, relationships would never be expected to always be strong. Some individuals may have a mild deficit in intelligibility but rate their quality of life or communicative participation as severely affected. In this study, TEP speakers II and D had high intelligibility (>95%) and yet their self-report measures indicated moderate to severe handicap in VHI-10 and they had low CPIB scores. The opposite is also seen in many voice and speech disorders, in which someone with a severe speech intelligibility deficit may use supplemental (e.g., nonverbal cues) communicative strategies, familiar communication partners, and positive coping strategies to continue participating in everyday life situations that are meaningful to that person (Eadie 2003). In this study, TEP speaker Q, had significantly low intelligibility in quiet (72.88%) and his self reported measures indicated no to mild handicap in VHI-10 and high participation level noted in a higher CPIB score. This is why it is so important to use a multidimensional approach to assessment because one measure does not always predict another.

*Future Research and Clinical Implications*

Results from this study have a number of future research and clinical implications. The results of the study suggest that intelligibility in noise is a better predictor of self-rated communication function than intelligibility in quiet. While this needs to be explored further in other alaryngeal populations and other speech and voice disorders, there are some valuable clinical implications. Clinically, speech intelligibility is measured in quiet and often rated by experienced listeners (clinicians) (Bennett & Weinberg, 1973; Tardy-Mitzell et al., 1985). This is not representative of typical environments where patients and sometimes an unfamiliar communication partner will be communicating. Part of a good multidimensional approach to assessment and intervention with alaryngeal speakers should therefore include an evaluation of their communication in background noise. The addition of intelligibility ratings in background noise would be an asset when working on therapy goals. Clinical options could include playing background noise in a session and calculating the impact of noise on each individual's intelligibility. If the decrease in intelligibility in noise is significant, the clinician could work on compensatory strategies (environmental changes or external aids) with the client and their communication partners. However, one must also consider that some alaryngeal speakers' quality of life is not functionally impacted by reduced intelligibility. Thus, measurement of intelligibility alone is not enough, and must be considered in the context of self-reported function in everyday situations as well.

It is essential to consider other factors that may influence intelligibility and self-reported outcomes. Future studies should explore the demographic factors that might impact these variables: socioeconomic status, daily communication demands,

quality, types and level of background of noise. For example, this study examined only one level of background noise, +6db SNR. Each individual speaker may be exposed to a variety of levels and types of background noise in the daily situations. Further investigation into determining what type and level of noise that may decrease intelligibility most is warranted.

Another factor to consider is the acoustic qualities of different alaryngeal speakers, and how those qualities impact intelligibility. A previous study, by van As and colleagues (2003), discussed how intelligibility of TEP speakers can be impacted by poor voice quality, altered speech rate and noise associated with occlusion of the tracheostoma. Perhaps the more intelligible TEP speakers in noise have better speech rate and less noise from occluding their tracheostoma. There could also be an acoustic or vocal quality variable that made those speakers more intelligible. Further studies including an acoustic analysis of TEP speakers and other alaryngeal speakers may also help to determine other factors that improve or impact intelligibility in noise.

Finally, further research should also explore listener-rated outcome measures and the impact of noise. As Beukelman and colleagues noted, attention allocation and the cognitive load placed on listeners of speakers with disordered speech is significant. Short-term memory, attention and effort all play a substantial role in a listener's judgment of intelligibility. Further studies should investigate the cognitive load placed on listeners and determine methods that may increase reliability among listeners. In addition, different methods of assessing intelligibility

(e.g., use of rating scales vs. transcription methods) may also reveal underlying relationships.

Further investigation of intelligibility in noise, and speaker-rated outcome measures, such as the VHI-10 and the CPIB, and how they impact each other will offer an improved understanding of the impact and disability of alaryngeal speech. The results of these studies will be critical to understand when working with laryngectomy patients and caregivers, for providing counseling and strategies for intervention. There are now three commonly used alaryngeal speech method options for successful voice restoration after laryngectomy: esophageal, electrolarynx and tracheoesophageal speech. It is the role of the clinician to give the patient with laryngectomy sufficient information to make an informed choice based on his/her own individual needs and abilities (Ylvisaker, Jacobs, & Feeney, 2003). Assessing a patient by solely using measures such as intelligibility in quiet may not be representative of everyday communication interactions. As this study demonstrates, there is a stronger relationship between speech intelligibility in noise and self-reported measures than in quiet. When counseling alaryngeal speakers on their chosen speech method, we must therefore consider all the external (noise level, noise type and environment) and internal factors (personality, communication demands) that may impact quality of life and ability to participate in all daily activities. These considerations should ultimately enhance our approach to rehabilitation in this special population.

## References:

- Azevedo, E. H. M., Montoni, N., Gonçalves Filho, J., Kowalski, L. P., & Carrara-de Angelis, E. (2012). Vocal handicap and quality of life after treatment of advanced squamous carcinoma of the larynx and/or hypopharynx. *Journal of Voice*, *26*(2), e63-e71.
- Baylor, C., Burns, M., Eadie, T., Britton, D., & Yorkston, K. (2011). A qualitative study of interference with communicative participation across communication disorders in adults. *American Journal of Speech-Language Pathology*, *20*(4), 269-287.
- Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of speech, language, and hearing research : JSLHR*, *56*(August), 1-16. doi:10.1044/1092-4388(2012/12-0140)
- Baylor, C., Yorkston, K., Eadie, T., Miller, R., & Amtmann, D. (2009). Developing the Communication Participation Item Bank. *Journal of Speech, Language, and Hearing Research*, *52*, 1302-1320.
- Bennett, S., & Weinberg, B. (1973). Acceptability ratings of normal, esophageal, and artificial larynx speech. *Journal of Speech and Hearing Research*, *16*(4), 608.
- Beukelman, D., Childes, J., Carrell, T., Funk, T., Ball, L. J., & Pattee, G. L. (2011). Perceived attention allocation of listeners who transcribe the speech of speakers with amyotrophic lateral sclerosis. *Speech Communication*, *53*(6), 801-806.
- Bunton, K. (2006). Fundamental frequency as a perceptual cue for vowel identification in speakers with Parkinson's disease. *Folia phoniatrica et logopaedica*, *58*(5), 323-339.
- Chen, A.Y., Frankowski, R., Bishop-Leone, J., Hebert, T., Leyk, S., Lewin, J., & Goepfert, H. (2001). The development and validation of a dysphagia-specific quality-of-life questionnaire for patients with head and neck cancer: the M. D. Anderson dysphagia inventory. *Archives of Otolaryngology – Head and Neck Surgery*, *127*(7), 870-876.
- Cullinan, W., Brown, C., & Blalock, P. (1986). Ratings of intelligibility of esophageal and tracheoesophageal speech. *Journal of communication disorders*, *19*, 185-195.  
Retrieved from  
<http://www.sciencedirect.com/science/article/pii/0021992486900080>
- D'Alatri, L., Bussu, F., Scarano, E., Paludetti, G., & Marchese, M. R. (2012). Objective and subjective assessment of tracheoesophageal prosthesis voice outcome. *Journal of voice : official journal of the Voice Foundation*, *26*(5), 607-13. doi:10.1016/j.jvoice.2011.08.013
- Day, A. M. B., & Doyle, P. C. (2010). Assessing self-reported measures of voice disability in tracheoesophageal speakers. *Journal of otolaryngology head neck surgery Le Journal otorhinolaryngologie et de chirurgie cervicofaciale*, *39*(6), 762-768.

- Donovan, N. J., Kendall, D. L., Young, M. E., & Rosenbek, J. C. (2008). The communicative effectiveness survey: Preliminary evidence of construct validity. *American Journal of Speech-Language Pathology, 17*(4), 335-347.
- Doyle, P.C. and Keith, R.L. (eds.) (2005). Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech and swallowing. Pro-Ed: Austin, TX.
- Eadie, T. L. (2003). The ICF: a proposed framework for comprehensive rehabilitation of individuals who use alaryngeal speech. *American journal of speech-language pathology / American Speech-Language-Hearing Association, 12*(2), 189-97. doi:10.1044/1058-0360(2003/065)
- Eadie, T.L., & Doyle, P.C. (2004). Auditory-perceptual scaling and quality of life in tracheoesophageal speakers. *Laryngoscope, 114*(4), 753-759.
- Eadie, T. L., & Doyle, P. C. (2005). Scaling of voice pleasantness and acceptability in tracheoesophageal speakers. *Journal of voice : official journal of the Voice Foundation, 19*(3), 373-83. doi:10.1016/j.jvoice.2004.04.004
- Eadie, T. L. (2007). Application of the ICF in Communication after Total Laryngectomy. *Seminars in Speech and Language, 1*(212), 291-300. doi:10.1055/s-2007-986526.
- Eadie, T. L., Day, A. M., Sawin, D. E., Lamvik, K., & Doyle, P. C. (2013). Auditory-Perceptual Speech Outcomes and Quality of Life after Total Laryngectomy. *Otolaryngology--Head and Neck Surgery, 148*(1), 82-88.
- Eadie, T. L., Lamvik, K., Baylor, C. R., Yorkston, K. M., Kim, J., & Amtmann, D. (2014). Communicative Participation and Quality of Life in Head and Neck Cancer. *Annals of Otolaryngology, Rhinology & Laryngology, 123*(4), 257-264.
- Evans, E., Carding, P., & Drinnan, M. (2009). The Voice Handicap Index with post-laryngectomy male voices. *International journal of language & communication disorders, 44*(5), 575-586.
- Fairbanks, G. (1960). *Voice and articulation drillbook*. New York: Harper & Bros.
- Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, Reliability and Validity of PRESTO: A New High-Variability Sentence Recognition Test. *Journal of the American Academy of Audiology, 24*(1), 26.
- Hillman, R. E., Walsh, M. J., & Heaton, J. T. (2005). Laryngectomy speech rehabilitation: A review of outcomes. *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing, 75-90*.
- Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13*(4), 557-569.

- Holley, S. C., Lerman, J., & Randolph, K. (1983). A comparison of the intelligibility of esophageal, electrolaryngeal, and normal speech in quiet and in noise. *Journal of communication disorders, 16*(2), 143-155.
- Hustad, K. C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech Language Pathology, 15*(3), 268-277.
- Hustad, K. C., & Beukelman, D. R. (2002). Listener comprehension of severely dysarthric speech: effects of linguistic cues and stimulus cohesion. *Journal of Speech, Language and Hearing Research, 45*, 545-558.
- Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech Language Pathology, 12*, 198-208.
- Hustad, K. C., & Garcia, J. M. (2002). The influences of alphabet supplementation, iconic gestures, and predictive messages on intelligibility of a speaker with cerebral palsy. *Journal of Medical Speech Language Pathology, 10*, 279-285.
- Jacobson, B.H., Johnson, A., Grywalsky, C., Silbergleit, A., Jacobson, G., & Benninger, M.S. (1997). The Voice Handicap Index (VHI): development and validation. *American Journal of Speech Language Pathology, 6*(3), 66-70.
- Kazi, R., Sayed, S. I., & Dwivedi, R. C. (2010). Post laryngectomy speech and voice rehabilitation: past, present, and future. *ANZ Journal of Surgery, 80*(11): 770-71.
- Kazi, R., Pawar, P., Sayed, S. I., & Dwivedi, R. C. (2010). Perspectives on voice rehabilitation following total laryngectomy. *European journal of cancer care, 19*(6), 703-705.
- Kim, J., & Davis, C. (2013). Comparing the consistency and distinctiveness of speech produced in quiet and in noise. *Computer Speech & Language*.  
doi:10.1016/j.csl.2013.02.002
- Law, I. K. Y., Ma, E. P. M., & Yiu, E. M. L. (2009). Speech intelligibility, acceptability, and communication-related quality of life in Chinese alaryngeal speakers. *Archives of Otolaryngology-Head & Neck Surgery, 135*(7), 704-711.
- McAuliffe, M.J., Schaefer, M., O'Beirne, G.A., LaPointe, L.L. (2009) Effect of noise upon the perception of speech intelligibility in dysarthria. New Orleans, LA, USA: American-Speech-Language-Hearing Association (ASHA) Annual Convention 2009, 19-21 Nov 2009.
- McCull, D. a. (2006). Intelligibility of tracheoesophageal speech in noise. *Journal of voice : official journal of the Voice Foundation, 20*(4), 605-15.  
doi:10.1016/j.jvoice.2005.09.005

- McColl, D., Fucci, D., Petrosino, L., Martin, D. E., & McCaffrey, P. (1998). Listener ratings of the intelligibility of tracheoesophageal speech in noise. *Journal of Communication Disorders, 31*(4), 279–288; quiz 288–289.
- Meyer, T. K., Kuhn, J. C., Campbell, B. H., Marbella, A. M., Myers, K. B., & Layde, P. M. (2004). Speech intelligibility and quality of life in head and neck cancer survivors. *The Laryngoscope, 114*, 1977–1981.
- Miralles, J. L., & Cervera, T. (1995). Voice intelligibility in patients who have undergone laryngectomies. *Journal of Speech, Language, and Hearing Research, 38*(3), 564–571.
- Oridate, N., Homma, A., Suzuki, S., Nakamaru, Y., Suzuki, F., Hatakeyama, H., ... & Fukuda, S. (2009). Voice-related quality of life after treatment of laryngeal cancer. *Archives of Otolaryngology–Head & Neck Surgery, 135*(4), 363–368.
- Rogers, S. N., Gwanne, S., Lowe, D., Humphris, G., Yueh, B., & Weymuller, E. A. (2002). The addition of mood and anxiety domains to the University of Washington quality of life scale. *Head neck, 24*, 521–529.
- Rosen, C.A., Lee, A.S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the voice handicap index-10. *Laryngoscope, 114*(9), 1549–1556.
- Sherman, A.C., Simonton, S., Adams, D.C., Vural, E., Owens, B., & Hanna, E. (2000). Assessing quality of life in patients with head and neck cancer: cross-validation of the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Head and Neck module (QLQ-H&N35). *Archives of Otolaryngology – Head and Neck Surgery, 126*(4), 459–467.
- Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology, 8*(2).
- Tardy-Mitzell, S., Andrews, M. L., & Bowman, S. A. (1985). Acceptability and intelligibility of tracheoesophageal speech. *Archives of Otolaryngology, 111*(4), 213–215.
- Terrell, J.E., Fisher, S.G., & Wolf, G.T. (1998). Long-term quality of life after treatment of laryngeal cancer. The Veterans Affairs Laryngeal Cancer Study Group. *Archives of Otolaryngology – Head and Neck Surgery, 124*(9), 964–971.
- Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of Clear, Loud and Slow Speech on Scaled Intelligibility and Speech Severity in Parkinson’s Disease and Multiple Sclerosis. *Journal of Speech, Language, and Hearing Research*. Retrieved from [http://dx.doi.org/10.1044/2014\\_JSLHR-S-12-0372](http://dx.doi.org/10.1044/2014_JSLHR-S-12-0372)
- van As, C. J., Koopmans-van Beinum, F. J., Pols, L. C., & Hilgers, F. J. (2003). Perceptual evaluation of tracheoesophageal speech by naive and experienced judges through the use of semantic differential scales. *Journal of Speech, Language & Hearing Research, 46*(4).

- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native-and foreign-language multi-talker background noise). *The Journal of the Acoustical Society of America*, *121*(1), 519-526.
- WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine*, *41*(10), 1403–1409.
- Ylvisaker, M., Jacobs, H. E., & Feeney, T. (2003). Positive supports for people who experience behavioral and cognitive disability after brain injury: A review. *The Journal of head trauma rehabilitation*, *18*(1), 7-32.
- Yorkston, K., Beukelman, D., & Tice, R. (1996). *Sentence Intelligibility Test (Version 1.0) [Speech analysis]*. Lincoln, NE: Tice Technology Services.
- Yorkston, K., & Beukelman, D. (1981). *Assessment of Intelligibility of Dysarthric Speech*. Portland, OR: C.C. Publications.
- Yorkston, K. M., Hakel, M., Beukelman, D. R., & Fager, S. (2007). Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: a systematic review. *Journal of Medical Speech-Language Pathology*, *15*(2), XI-XXXVI.

## APPENDIX A

### **Voice Handicap Index-10 (VHI-10)**

**Instructions:** Many people with head and neck cancer have difficulties with their voices. If you have had a laryngectomy, please answer the questions based on your experiences with the speech method you are using (e.g. electrolarynx, TEP or esophageal speech). Check the response that indicates *how frequently* you have the each experience.

0= Never; 1=Almost Never; 2=Sometimes; 3=Almost Always; 4= Always

	Never	Almost Never	Sometimes	Almost Always	Always
F1. My voice makes it difficult for people to hear me.					
P2. I run out of air when I talk.					
F3. People have difficulty understanding me in a noisy room.					
P4. The sound of my voice varies throughout the day.					
F5. My family has difficulty hearing me when I call them throughout the house.					
F6. I use the phone less often than I would like.					
E7. I'm tense when talking with others because of my voice.					
F8. I tend to avoid groups of people because of my voice.					
E9. People seem irritated with my voice.					
P10. People ask, "What's wrong with your voice?"					

## APPENDIX B

### The Communicative Participation Item Bank – General Short Form

**Instructions:**

The following questions describe a variety of situations in which you might need to speak to others. For each question, please mark how much your condition interferes with your participation in that situation. By “condition” we mean ALL issues that may affect how you communicate in these situations including speech conditions, any other health conditions, or features of the environment. If your speech varies, think about an AVERAGE day for your speech – not your best or your worst days.

	Not at all (3)	A little (2)	Quite a bit (1)	Very much (0)
1. Does your condition interfere with... ...talking with people you know?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Does your condition interfere with... ...communicating when you need to say something quickly?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Does your condition interfere with... ...talking with people you do NOT know?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Does your condition interfere with... ...communicating when you are out in your community (e.g. errands; appointments)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Does your condition interfere with... ...asking questions in a conversation?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Does your condition interfere with... ...communicating in a small group of people?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Does your condition interfere with... ...having a long conversation with someone you know about a book, movie, show or sports event?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Does your condition interfere with... ... giving someone DETAILED information?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Does your condition interfere with... ...getting your turn in a fast-moving conversation?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Does your condition interfere with... ...trying to persuade a friend or family member to see a different point of view?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>