11-11-2019 10:00 AM

Exploring the Temporal Variability of Speech Intensity, Speech Intelligibility, and Communicative Participation in Individuals with Hypophonia and Parkinson's Disease

Cynthia Mancinelli
The University of Western Ontario

Supervisor
Adams, Scott G.
The University of Western Ontario Co-Supervisor
Page, Allyson D.
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

Hypophonia, or reduced speech intensity, is frequently observed in individuals with Parkinson's disease (PD). This speech deficit can impact speech intelligibility and communicative participation. However, there is little empirical evidence exploring the day-to-day variability of speech and communicative participation in individuals with PD. The purpose of this study is to investigate the temporal variability of acoustic and perceptual speech measures and psychosocial measures in individuals with hypophonia and PD. Additionally, this study seeks to examine the relationships among measures of speech intensity, speech intelligibility, self- and proxy-rated communicative participation, demographic factors, and non-speech factors. Twenty-three participants with PD, 23 primary communication partners, and 30 control participants attended three experimental visits. At each visit, participants completed questionnaires related to speech loudness and communicative participation. Participants with PD and control participants also performed speech intensity and speech intelligibility tasks. Variability in habitual speech intensity and Lombard response slope was found for participants with PD. Differences were found in maximum speech intensity, magnitude production, speech intelligibility, and self-perceived typical speech loudness for participants with PD and control participants. The results revealed similar self- and proxy-ratings of speech loudness in participants with PD. A significant difference was found between participants with PD and control participants across self-rated communicative participation measures. Greater variability was observed for the Communicative Participation Item Bank (CPIB), six questions of the Communicative Effectiveness Survey (CES), and two subsections of the Voice Activity and Participation Profile (VAPP) in participants with PD. Self- and proxy-rated communicative participation was comparable. Significant relationships were identified between maximum speech intensity and magnitude production, between speech intelligibility measures, between the CPIB and two VAPP subsections, between VAPP subsections, between proxy-ratings of typical speech loudness and six VAPP subsections, the CPIB and select CES questions, and VAPP subsections in participants with PD. A retest analysis involving reliability and repeatability estimates for all dependent measures was also reported. These findings contribute to the understanding of hypophonia in PD

and may provide context in the interpretation of treatment outcomes in this clinical population.

Keywords

Parkinson's disease, hypophonia, speech intensity, speech intelligibility, communicative participation, retest reliability, repeatability, speech language pathology.

Summary for Lay Audience

Hypophonia, or reduced speech loudness, is frequently observed in individuals with Parkinson's disease (PD). This speech difficulty can impact an individual's speech intelligibility (how understandable they are) and their communicative participation (engaging in situations where information is shared). However, there is little empirical evidence exploring the day-to-day variability of various speech and communication measures in individuals with PD. The purpose of this study is to investigate the variability of such measures in individuals with hypophonia and PD. Additionally, this study seeks to examine the relationships among measures of speech loudness, speech intelligibility, communicative participation, and other demographic and non-speech factors. Twenty-three participants with PD, 23 communication partners of participants with PD, and 30 control participants attended three experimental visits. At each visit, participants completed questionnaires related to speech loudness and communicative participation. Participants with PD and control participants also performed speech intensity and speech intelligibility tasks. Significant variability was found in various speech tasks in participants with PD, including habitual speech loudness and Lombard response function (how one's speech loudness changes in the presence of background noise). Differences were found between participants with PD and control participants for various speech loudness, speech intelligibility, and communicative participation measures. Similar perceptions of speech loudness and communicative participation were found in participants with PD and their communication partners. Significant relationships were identified between select speech loudness, speech intelligibility, and communicative participation measures. These findings contribute to the understanding of hypophonia in PD and may facilitate the interpretation of treatment outcomes in this clinical population.

Acknowledgements

I would like to extend my appreciation and gratitude to my supervisors, Dr. Scott Adams and Dr. Allyson Page, for their support and guidance throughout my degree and the completion of this project. Their assistance and contributions have been invaluable throughout every step of the way. I would also like to thank my advisory committee, Dr. Mandar Jog and Dr. Julie Theurer, for their guidance and expertise. Many thanks to my thesis examination committee, Dr. Julie Theurer, Dr. Jeff Holmes, Dr. Mary Jenkins, and Dr. Mark Hakel, for taking the time to read my thesis and ensuring a high caliber final product.

These last six years would not have been possible without my parents, Diane Pardillo and John Mancinelli. Thank you for your unwavering encouragement and faith in me. Your unconditional love and support have allowed me to persevere and achieve my academic goals and realize that I am capable of accomplishing anything I chose to strive towards. I would also like to thank my brother and extended family for your love and encouragement throughout this journey.

Thank you to my dearest friends and lab mates, Anita Abeyesekera, Thea Knowles, and Daryn Cushnie-Sparrow, and my treasured friends Caitlin Coughler, Anjana Balakrishna, and Meg Haggitt. Thank you for your friendship, support, and motivation. Thank you for helping me gain perspective, setting and accomplishing goals, and ensuring that I have balance in my life. I could not have asked for stronger and more supportive friends.

Finally, I would like to express my deepest appreciation and thanks to all of the participants who generously donated their time to take part in this study. This project would not have been possible without them.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Acronyms

PD	Parkinson's disease
IWPD	Individuals with Parkinson's disease
dB	Decibels
SPL	Sound pressure level
LSVT	Lee Silverman Voice Treatment
SIT	Sentence Intelligibility Test
WHO	World Health Organization
ICF	International Classification of Functioning, Disability, and Health
PRO	Patient-reported outcome measures
CPIB	Communicative Participation Item Bank
VAPP	Voice Activity and Participation Profile
CES	Communicative Effectiveness Survey
UPDRS	Unified Parkinson's Disease Rating Scale
LSUS	Level of Speech Usage Scale
HL	Hearing level
Hz	Hertz
MOCA	Montreal Cognitive Assessment
GDS-15	Geriatric Depression Scale – short form
MES	Medication Effectiveness Scale
m	Meters
cm	Centimeters
kHz	Kilohertz
TLS	Typical Loudness Scale
VAS	Visual analogue scale
CETI-M	Communicative Effectiveness Index – Modified
CETI	Communicative Effectiveness Index
ANOVA	Analysis of variance
SEM	Standard error of measurement
CR	Repeatability coefficient
CR%	Percent coefficient of repeatability

ICC	Interclass correlations
SRD%	Smallest real difference percentage
MANOVA	Multivariate analysis of variance
MDD	Minimum detectable difference
VHI	Voice Handicap Index
PROMIS	Patient Reported Outcomes Measurement Information System
PDQ-8	Parkinson's Disease Questionnaire - 8

Chapter 1

1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disease that typically affects individuals between the ages of 60 and 70 years (Adams & Jog, 2009). The hallmark motor features of PD: tremor, bradykinesia, rigidity, and postural instability, typically appear once 50 – 60% and 80% of dopamine has been depleted in the substantia nigra, and striatum, respectively (Duffy, 2013; Sapir, 2014; Schneider & Obeso, 2014; Wirdefeldt, Adami, Cole, Trichopoulos, & Mandel, 2011). The degeneration of dopaminergic pathways can also result in speech difficulties (Duffy, 2013; Sapir, 2014).

Over the course of the disease, approximately 70 – 90% of individuals with PD (IWPD) will develop a variety of speech impairments, collectively referred to as hypokinetic dysarthria (Sapir, 2014). Darley, Aronson, and Brown (1969a; 1969b) coined the term hypokinetic dysarthria in order to describe the most common characteristics of Parkinsonian speech. The distinctive features of hypokinetic dysarthria include monopitch, reduced stress, monoloudness, reduced loudness, imprecise consonants, inappropriate silences, short rushes of speech, harsh voice, breathiness, low pitch, and variable rates of speech (Darley et al., 1969a; 1969b; Duffy, 2013). While it is unlikely for any one IWPD to express all of the above described speech features, there is a variety in symptom manifestation, suggesting heterogeneity in the dysarthric profiles of IWPD (Duffy, 2013).

1.1 Hypophonia and Parkinson's Disease

Hypophonia, or reduced speech intensity, is a commonly treated and highly debilitating speech feature of Parkinsonian speech, affecting an estimated 50% of IWPD (Adams, Dykstra, Abrams, Winnell, Jenkins, & Jog, 2006a; Adams, Dykstra, Jenkins, & Jog, 2008; Adams, Haralabous, Dykstra, Abrams, & Jog, 2005; Adams, Moon, Dykstra, Abrams, Jenkins, & Jog, 2006b; Dykstra, Adams, & Jog, 2012a). Additionally, hypophonia frequently emerges as an initial speech symptom in the early stages of PD (Dykstra et al., 2012a; Dykstra, Adams, & Jog, 2015). While hypokinetic dysarthria affects the vast majority of IWPD over the course of the disease, 30% of hypokinetic

speakers consider hypophonia, along with other aspects of their speech impairments, to be among the most devastating of their PD-related symptoms (Berke, Gerratt, Kreiman, & Jackson, 1999). Challenges frequently experienced by individuals with hypophonia include difficulty being heard in social settings, and numerous requests from communication partners to speak louder and to repeat themselves (Adams et al., 2006a; Adams et al., 2006b; Dykstra et al., 2012a).

Hypophonia can also have a negative impact on the acoustics of an individual's speech production (Kwan & Whitehill, 2011; Tjaden, Richards, Kuo, Wilding, & Sussman, 2013). Judgements pertaining to the severity of an individual's hypophonia may be made via acoustic measures, including but not limited to habitual speech intensity, intensity decay, and intensity variability (Adams, Winnell, & Jog, 2010; Ho, Iansek, & Bradshaw, 2001; McCaig, Adams, Dykstra, & Jog, 2016; Rosen, Kent, Delaney, & Duffy, 2006; Rosen, Kent, & Duffy, 2005;).

The habitual speech intensity of individuals with hypophonia and PD is typically reduced by 2 – 5 decibels (dB) sound pressure level (SPL) compared to healthy speakers (Adams et al., 2010). This difference can be quite dramatic, as a reduction of 4 dB SPL in speech intensity is equivalent to a 40% decrease in the perceived loudness level (McCaig et al., 2016). Intensity decay is defined as the declination of an individual's speech intensity (Ho et al., 2001; Rosen et al., 2005). Intensity decay can be calculated by examining the average and peak intensity of a syllable or prolonged vowel, as well by examining the intensity contour of a sentence or conversation (Rosen et al., 2005). Furthermore, the rate and amount of intensity declination can be observed (Ho et al., 2001; Rosen, et al., 2005). To some extent, intensity decay naturally occurs in speech tasks, such as at the end of declarative sentences (Rosen et al., 2005). However, individuals with hypophonia and PD have demonstrated a larger declination in intensity than control participants in vowel prolongation and syllable repetition, but not in conversational tasks, and with inconsistent results for sentence reading tasks (Ho et al., 2001; Rosen et al., 2005). It has been suggested that neurological differences are present in speech and non-speech motor tasks (Kent, 2004). These differences may include the complex and unique musculature associated with the speech system, the functional differences of the craniofacial, lingual,

and laryngeal muscles as compared to skeletal muscles, and the effects of medical intervention. The task-specific intensity decay reported by Rosen and colleagues (2005) may suggest that declination is the result of the interference between sub-processes required for the tasks described by Kent (2004), in addition to task demands or goals. Intensity variability is the standard deviation observed in the intensity contour of an utterance (Rosen et al., 2006). While intensity variability is typically observed during conversational speech, the proportion of variability appears to differ in IWPD and healthy speakers (Rosen et al., 2006). Rosen and colleagues (2006) found that overall intensity variability was reduced in IWPD compared to control participants during sentence and conversational tasks, at levels of 5.4 dB SPL and 6.5 dB SPL, respectively. Additionally, the speech intensity range of individuals with hypophonia may also be reduced (Adams & Dykstra, 2009; Darley et al., 1969a; 1969b; Darley, Aronson, & Brown,1975). Decreased intensity variability in conjunction with reduced speech intensity range may contribute to the perception of monoloudness in the speech of IWPD (Adams & Dykstra, 2009; Darley et al., 1969a; 1969b; 1975). While there are numerous methods to assess an individual's speech intensity, the mechanisms that regulate speech intensity in IWPD are poorly understood.

1.2 Regulation of Speech Intensity in Hypophonia and Parkinson's Disease

The progressive nature of PD often results in a variety of physiological and anatomical changes, which may manifest as various motor and non-motor symptoms observed in PD (Duffy, 2013). Hypophonia may be a manifestation of complex interactions among the neurodegenerative physiological and anatomical changes associated with PD. It is believed that the neurological system, and consequently neurological impairments, has a significant impact on the regulation of speech intensity in individuals with hypophonia and PD (Arnold, Gehrig, Gispert, Seifried, & Kell, 2014; Braak, Ghebremedhin, Rüb, Bratzke, & Del Tredici, 2004; Clark, Adams, Dykstra, Moodie, & Jog, 2014; Conte, Khan, Defazio, Rothwell, & Berardelli, 2013; Ho, Iansek, & Bradshaw, 1999b; Kempler & Van Lancker, 2002). Many functional and structural changes in PD are believed to be governed by dysfunctional basal ganglia, and these functional and structural changes are

hypothesized to be the underlying cause of speech impairments such as hypophonia in IWPD (Ho et al., 1999b).

The basal ganglia are dopamine-modulated subcortical structures that are impacted in PD (Sapir, 2014). The dopamine depletion observed in PD affects all of the structures comprising the basal ganglia, particularly the globus pallidus, subthalamic nucleus, substantia nigra, and striatum; the latter of which is composed of the caudate nucleus and putamen (Sapir, 2014). While it has been demonstrated that a reduction of dopamine within these structures results in the classic motor symptoms associated with PD, it has been suggested that basal ganglia pathology and direct cortico-striatal pathways connecting the ipsilateral caudate nucleus to the dorsolateral prefrontal cortex can be attributed to the development of hypokinetic dysarthria symptoms associated with PD, including hypophonia, (Arnold et al., 2014; Sapir, 2014). However, several studies suggest that the presentation of hypokinetic dysarthria may also be related to pathologies in non-dopaminergic structures and pathways at earlier stages of PD development (Braak et al., 2004; Sapir, 2014). Several of these key neural structures, such as the striatum, are involved in motor and sensory activities, as well as sensorimotor integration (Conte et al., 2013; Kempler & Van Lancker, 2002). These functions include the regulation of muscle tone, coordinating and stabilizing the body during voluntary movements, scaling the force, amplitude, and duration of movements, adapting movements to suit the environment, as well as play a role in learning, planning, and initiating movements (Duffy, 2013). Therefore, it has been hypothesized that the presentation of hypophonia may be causally related to a sensorimotor integration deficit in IWPD (Clark et al., 2014; Ho, Bradshaw, & Iansek, 2000).

Sensorimotor integration typically uses the available sensory information in order to monitor, plan, and execute movement, and is typically impaired in IWPD (Clark et al., 2014; Ho et al., 2000). Thus, individuals with hypophonia and PD may demonstrate evidence of sensory and somatosensory deficits in addition to sensorimotor integration deficits, as evidenced by reduced speech intensity and impaired loudness perception (Clark et al., 2014; Conte et al., 2013; Ho et al., 2000).

1.2.1 Speech intensity in hypophonia and Parkinson's disease

As discussed previously, differences in the habitual speech intensity of individuals with hypophonia compared to healthy speakers can be quite dramatic (Adams et al., 2010; McCaig et al., 2016). In addition, hypophonia may be assessed via several acoustic measures, including habitual speech intensity, intensity decay, and intensity variability (Adams et al., 2010; Ho et al., 2001; McCaig et al., 2016; Rosen et al., 2006; Rosen et al., 2005). However, anecdotal evidence suggests that IWPD and hypophonia may experience a range in their day-to-day speech intensity. Little is known about the natural variability of individuals' speech in everyday situations. While many studies have examined the day-to-day speech behaviours of teachers (Astolfi et al., 2012; Astolfi, Puglisi, Pavese, & Carullo, 2014; Bottalico, & Astolfi, 2012; Franca, 2013; Schmidt, Andrews, & McCutcheon, 1998), few studies have examined speech patterns and behaviours in individuals with speech and voice disorders (Fox & Ramig, 1997; Ramig, Sapir, Fox, & Countryman, 2001; Sapir, Spielman, Ramig, Story, & Fox, 2007; Schalling, Gustafsson, Ternström, Bulukin Wilén, & Södersten, 2013). A study by Schalling and colleagues (2013) tracked daily voice use over three weeks in six IWPD and hypophonia. Participants wore a VoxLog which recorded their speech intensity and the general noise level. A range of 11 dB was observed over the course of three weeks, which may demonstrate a potential for a substantial degree of day-to-day variation and individual differences in a population of IWPD and hypophonia.

The previously described study by Schalling and colleagues (2013) appears to be the only published study to have examined the long-term daily speech behaviours of individuals with hypophonia and PD. However, several other studies have explored the variations of speech intensity and self-perceptions of communication in IWPD and hypophonia (Fox & Ramig, 1997; Ramig et al., 2001; Sapir et al., 2007). Ramig and colleagues (2001) explored the effect of Lee Silverman Voice Treatment (LSVT) on speech intensity in IWPD and hypokinetic dysarthria. Fourteen IWPD receiving LSVT, 14 IWPD not receiving treatment, and 14 control participants were recorded while performing four speech tasks on seven different occasions: three times within two weeks prior to onset

group, twice immediately following the completion of LSVT, and twice six months following the completion of LSVT. The results demonstrated that the speech intensity of participants not receiving treatment remained relatively consistent across all speech tasks at each of the three time points, with mean intensity ranging from 69.3 – 71.9 dB, across all time points and speech tasks.

Similarly, Sapir and colleagues (2007) sought to explore the impact of LSVT on the articulation and intensity of vowels in IWPD and dysarthria. Fourteen IWPD who received LSVT, 15 IWPD who did not receive speech treatment, and 14 control participants were audio recorded while reading three different phrases during each visit in order to obtain speech intensity levels for each of the vowels /i/, /u/, and /a/. Speech recordings were collected for all participants on three different days prior to LSVT onset, as well as on two different days following the completion of LSVT. The average speech intensity for these vowels of the IWPD who received treatment ranged from 72.6 – 73.6 dB, in the pre- and post-LSVT conditions. Non-significant differences were also observed between pre- and post-LSVT values for the IWPD who did not receive treatment.

Fox and Ramig (1997) examined the speech intensity of 15 male and 15 female IWPD and dysarthria, and seven male and seven female control participants. Participants attended three visits during a four-day period, wherein they completed four speech tasks at every visit. The results demonstrated that the speech intensity differences among all four groups were statistically different from one another. It was also reported that the speech intensity of IWPD was 2 – 4 dB lower than that of control participants. Additionally, the IWPD demonstrated a smaller variability of speech intensity compared to control participants, as demonstrated by their smaller standard deviations.

It is important to note that the above-mentioned studies reported the range and standard deviations of speech intensity measures for their participants. (Fox & Ramig, 1997; Sapir et al., 2007; Schalling et al. 2013). However, measurements such as interclass correlations (ICC) and coefficients of repeatability (CR) would provide greater information. ICC are used to assess the correlation between multiple sets of measurements and incorporates the consistency of within-subject measures and the

average change of a group mean over time (Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013). The CR, also referred to as the Smallest Real Difference and the Minimal Detectable Difference (Beckerman et al., 2001; Furlan & Sterr, 2018; Steffen & Seney, 2008), is the value below which the absolute differences between two measurements would lie with 0.95 probability (Vaz, et al., 2013). In other words, difference values above the CR would reflect 95% probability of a true level of change as opposed to a difference that may simply be due to measurement error. It is also suggested that CR reflects the minimal detectable true level of change of an outcome measure. Examining the retest reliability and repeatability of speech intensity in IWPD and hypophonia via ICC and CR may provide valuable insight into the variability of the measure.

1.2.2 The effect of background noise on speech intensity in hypophonia and Parkinson's disease

The Lombard effect is described as an involuntary and reflexive increase in speech intensity in the presence of background noise (Brumm & Zollinger, 2011; Garnier & Henrich, 2014). This reflex serves to monitor and adjust one's speech intensity as needed in order to ensure appropriate speech intensity levels with communication partners in the presence of background noise (Brumm & Zollinger, 2011; Dykstra et al., 2012a). The Lombard effect is present in individuals with hypophonia and PD, as well as healthy individuals without neurological disease (Adams & Lang, 1992; Adams et al., 2005; Stathopoulos et al., 2014).

Adams and Lang (1992) explored the effect of 90 dB SPL of white background noise on the speech intensity of 10 IWPD during a reading task. All participants demonstrated a Lombard response, with observed increases in speech intensity levels ranging from 2.1 – 7.5 dB SPL. The high degree of variability of the speech intensity increase in IWPD may suggest individual differences with regard to a Lombard response (Adams & Lang, 1992). These authors suggest that there may be individual differences in the levels of background noise needed for IWPD to achieve similar increases in speech intensity.

A similar study by Stathopoulos and colleagues (2014) examined responses to the Lombard effect in 33 IWPD during a speech task. The intensity level of the background noise presented to participants increased until participants' speech intensities were an average of 3 dB SPL higher than their habitual speech intensity. The manipulation resulted in 79% of participants with PD speaking with increased speech intensity in the presence of background noise. The authors did not report any data regarding the intensity level of background noise presented. However, it is possible that the manipulation of the level of background noise by Stathopoulos and colleagues (2014) for each participant to achieve a similar increase in speech intensity is related to the individual variability in the Lombard response previously discussed by Adams and Lang (1992).

While several studies have examined the Lombard effect in individuals with hypophonia and PD (Adams et al., 2006a; Adams et al. 2005; Adams & Lang, 1992; Dykstra et al., 2012a; Ho, Bradshaw, Iansek, & Alfredson, 1999a; Stathopoulos et al., 2014), few studies have attempted to delineate potential relationships among speech intensity and different intensity levels (i.e., ranging from 50-90 dB) and types (i.e., pink, instrumental music, and multi-talker) of background noise presented (Adams et al. 2006a; Adams et al., 2005; Adams & Lang, 1992). Adams and colleagues (2005) investigated whether such a relationship between speech intensity and type of background noise existed by having 10 individuals with hypophonia and PD and 10 control participants repeat sentences in five multi-talker background noise conditions. These researchers found that a Lombard response was elicited in both experimental groups, however IWPD consistently produced speech intensity levels that were 2 – 3 dB SPL lower than control participants across all background noise conditions. In 2006a, Adams and colleagues sought to investigate the effect of different types of background noise on speech intensity. In their study, 23 IWPD and hypophonia and 15 control participants conversed for two minutes in the presence of three types of background noise (i.e., multi-talker noise, instrumental music, and pink noise), each presented at five different intensity levels. Both groups of participants demonstrated the Lombard sign, however the speech intensity levels of IWPD was on average 3 – 5 dB SPL lower than that of their healthy counterparts. Significant differences between types of background noise were only

observed with pink noise and multi-talker noise when presented at specific intensity levels. The speech intensity of IWPD was 0.66 dB SPL higher in the presence of 70 dB SPL of multi-talker noise, as compared to 70 dB SPL of pink noise. The speech intensity of control participants was 0.69 dB SPL higher in the presence of 70 dB SPL of multi-talker noise, compared to 70 dB SPL of pink noise. The speech intensity of control participants was also 0.74 dB SPL higher in the presence of 55 dB SPL of multi-talker noise, compared to 55 dB SPL of pink noise. These results support the findings of Adams and colleagues (2005) and suggest that the relationship between speech intensity and background noise for individuals with hypophonia and PD is similar, but attenuated, in comparison to the response of control participants. Additionally, the results of Adams and Lang (1992) may suggest that multi-talker background noise, compared to pink noise or instrumental music, may be more beneficial in eliciting a Lombard response in IWPD and healthy individuals without neurological impairment.

Dykstra and colleagues (2012a) investigated maximum intensity, habitual intensity, and the effect of various intensity levels of background noise (ranging from 50 – 70 dB SPL) on the conversational speech intensity of 30 individuals with hypophonia and 15 control participants. Participants with PD demonstrated reduced maximum intensity of approximately 10 dB SPL compared to control participants. Participants with PD demonstrated reduced habitual intensity of approximately 5 dB compared to control participants. Both groups of participants demonstrated a Lombard response during the conversational speech task. However, the conversational speech intensity levels of IWPD were reduced by approximately 5 dB across noise conditions. The conversational intensity response pattern in IWPD was parallel, but attenuated, compared to the response pattern of control participants. The findings from this study further support the previously identified parallel but attenuated Lombard effect (Adams et al., 2006a; Adams et al., 2005) in individuals with hypophonia and PD compared to control participants.

In sum, the literature demonstrates that while individual responses to the Lombard effect may vary, the presence of background noise naturally facilitates the ability of individuals with hypophonia and PD to increase their speech intensity. Furthermore, the performance

of IWPD parallels that of control participants regardless of the type of noise used to elicit the Lombard effect.

Additionally, noise may play multiple roles in communication in individuals with hypophonia and PD. In addition to increasing speech intensity, background noise may also influence listener perceptions of hypophonia severity, and potentially introduce communication challenges, such as reducing the speech intelligibility (Adams et al., 2008; Dykstra, Adams, & Jog, 2012b) and communicative participation (Baylor, Burns, Eadie, Britton, & Yorkston, 2011) of individuals with hypophonia.

1.3 Loudness Perception in Hypophonia and Parkinson's Disease

Individuals with hypophonia and PD may be unaware of their reduced speech loudness level (Adams & Dykstra, 2009; Clark et al., 2014; Fox & Ramig, 1997; Ho et al., 2000; Huber, Stathopoulos, Ramig, & Lancaster, 2003). Individuals with hypophonia are frequently requested by communication partners to speak louder and to repeat themselves (Dykstra et al., 2012a). A number of studies have explored loudness perception in this population (Clark et al., 2014; Fox & Ramig, 1997; Ho et al., 2000).

Ho and colleagues (2000) examined the ability of 15 individuals with hypophonia and PD and 15 control participants to perceive the loudness level of their own speech production. Participants were audio-recorded using their soft, normal, and loud voice while completing a reading task, and while engaged in conversation. Participants then adjusted a volume control knob to indicate the loudness level at which they had just spoken. These volume adjustments were completed twice: immediately following each speech production, and after replaying the recorded sample. The authors found that IWPD significantly over-estimated their spoken loudness levels during both reading and conversational tasks compared to control participants. These observed discrepancies between perceived and produced speech intensity levels may suggest impaired sensorimotor integration deficits in individuals with hypophonia and PD (Ho et al., 2000).

A recent study by Clark and colleagues (2014) sought to compare loudness perception in 17 individuals with hypophonia and PD and 25 control participants. Participants

completed a magnitude estimation task, an imitation task, and a magnitude production task, all involving a five-word target sentence. During the magnitude estimation task, participants rated the loudness of the target sentence when presented at 60, 65, 70, 75, and 80 dB SPL. Prior to beginning the task, a 70 dB SPL presentation of the target sentence was assigned a value of 100. During the imitation task, participants repeated the target sentence at the same speech intensity at which it was presented (60, 65, 70, 75, and 80 dB SPL). During the magnitude production task, participants initially read the target sentence at their habitual speaking volume, and this intensity level was designated a value of 100. Participants were then instructed to reproduce the target sentence in varying magnitudes (25, 50, 100, 200, and 400) compared to their initial performance.

Results from the magnitude estimation task indicated that IWPD rated stimuli presented at higher intensity levels (75 and 80 dB SPL) lower than did control participants, and rated stimuli presented at lower intensity levels (60 and 65 dB SPL) higher than did control participants. These findings suggest that IWPD have a flatter psychophysical loudness function and a more restricted range of intensity perception than control participants (Clark et al., 2014). Results from the imitation task demonstrated that all participants exhibited an increase in speech intensity as the intensity of the presented sentence increased, however IWPD spoke at consistently lower speech intensities than did their healthy counterparts. These findings suggest that IWPD exhibit shallower slopes in their imitation speech intensity function compared to control participants. Results from the magnitude production task revealed that IWPD made smaller adjustments to their intensity levels compared to control participants across all magnitude production conditions. These results suggest that the slope of the magnitude production function is less steep in IWPD compared to control participants.

The abnormal loudness perception results of Clark and colleagues (2014) and Ho and colleagues (2000) suggest that IWPD may have a deficit in their perception of loudness levels. Anecdotal reports that individuals with hypophonia perceive themselves as speaking too loudly when increasing their speech intensity to typical conversational intensity levels (Kwan & Whitehill, 2011) lend further support to the finding of Clark and

colleagues (2014) and Ho and colleagues (2000). It is possible that this loudness perception deficit may then influence their ability to produce speech at greater intensities.

1.4 Speech Intelligibility in Hypophonia and Parkinson's Disease

Hypophonia may influence multiple facets of one's ability to communicate in everyday life. An individual's ability to effectively and intelligibly take part in speech and communicative interactions is an essential part of one's communicative activities (De Bodt, Hernández-Díaz Huici, & Van De Heyning, 2002). A high level of speech intelligibility is critical for effective and efficient oral communication (Kent, Weismer, & Kent, 1989; Miller, 2013). It is important to assess the speech intelligibility of IWPD and hypophonia because speech intelligibility is often reduced in individuals with hypokinetic dysarthria secondary to PD (Adams et al., 2008).

Speech intelligibility is typically established by calculating the proportion of words correctly understood by a listener (Duffy, 2013; Kent et al., 1989). Furthermore, speech intelligibility is considered the "gold standard" for evaluating an individual's level of functional communication (Sussman & Tjaden, 2012). Reduced speech intelligibility is frequently characteristic of the speech of IWPD and hypokinetic dysarthria (Adams et al., 2008). The speech intelligibility of individuals with hypokinetic dysarthria and PD is typically evaluated at the single word or sentence level, with few studies examining conversational speech intelligibility (Andreetta, Adams, Dykstra, & Jog, 2016; Dykstra et al., 2012b; Tjaden & Wilding, 2011; Walsh & Smith, 2012). Some studies have suggested that the speech intelligibility of IWPD differs across speech tasks. For example, speech intelligibility is more likely to be reduced in tasks consisting of longer sentences or in spontaneous conversation than compared to single word utterances (Kempler & Van Lanker, 2002; Tjaden & Wilding, 2011; Walsh & Smith, 2012).

Sentence intelligibility is frequently used to obtain measures of overall speech intelligibility, (Miller, 2013). Once such test that is commonly used is the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Tice, 2011). The SIT is obtained via computer software that randomly generates 11 sentences of increasing word count that

range from 5 to 15 words in length. Measures of speech intelligibility are then obtained via orthographic transcription. Psychometric evaluations by Yorkston, Beukelman and Tice (1996) demonstrated that the SIT is a reliable and valid measure for assessing the speech intelligibility of individuals with dysarthria. A study by Cannito and colleagues (2012) used the SIT to examine sentence intelligibility in eight individuals with hypophonia before and after a loudness-training program. Participants were audio-recorded during three pre-treatment sessions and three post-treatment, wherein participants completed the SIT during each session. The recorded stimuli were played back to listeners at an adjusted volume of 55 dB SPL in the presence of white noise. Listeners rated sentence intelligibility via orthographic transcription. The average sentence intelligibility of participants increased from 81.11% pre-treatment, to 85.82% post-treatment. These results suggest that sentence intelligibility may increase following a loudness training treatment designed to increase the speech intensity of individuals with hypophonia and PD. However, no published studies have systematically explored any naturally occurring day-to-day variability of speech intelligibility in IWPD and hypophonia. Further investigation of the reliability and repeatability of measures of speech intelligibility using ICC and CR may lend further support to the strength of such treatment outcomes.

Additionally, the standard clinical method of assessing speech intelligibility may not fully capture the effects of reduced speech intensity on intelligibility in individuals with hypophonia (Adams et al., 2008; Dykstra et al., 2012b; Sussman & Tjaden, 2012). Therefore, obtaining conversational estimates of intelligibility and assessing intelligibility in noise may be beneficial in the management of individuals with hypophonia and PD. Few studies have explored the effect of hypophonia on conversational speech intelligibility (Adams et al., 2008; Andreetta et al., 2016; Dykstra et al., 2012b). These studies have included the effect of background noise in their study of conversational speech intelligibility in individuals with hypophonia and PD.

1.4.1 The effect of background noise on speech intelligibility in hypophonia and Parkinson's disease

Factors, such as background noise, may facilitate or limit speech intelligibility in individuals with hypophonia (Adams et al., 2008; Dykstra et al., 2012b). Adams and colleagues (2008) examined speech-to-noise levels and the speech intelligibility of individuals with hypophonia and PD in the presence of varying intensity levels of background noise. The speech intelligibility of 25 participants with hypophonia and 15 age-matched control participants was assessed during a conversational speech task in various background noise conditions, including no added background noise, 60, 65, and 70 dB SPL. Sentence intelligibility was also measured using the SIT in a quiet environment. The conversational intelligibility scores of IWPD were approximately 20 – 30% lower than those of control participants across all noise conditions, with the intelligibility scores of IWPD ranging from approximately 45 – 82% as the level of background noise increased. Additionally, the authors reported high levels of speech intelligibility for IWPD in sentences, with average SIT scores of 92%. Conversational speech intelligibility scores in noise conditions, however, were on average 5 – 10% lower than conversational intelligibility scores assessed in a quiet environment. These results suggest reduced conversational speech intelligibility is further exacerbated in increased levels of background noise in individuals with hypophonia. This reduction in conversational intelligibility appears to occur despite the relatively high levels of intelligibility of individuals with hypophonia when evaluated via the SIT.

Dykstra and colleagues (2012b) also sought to assess conversational speech intelligibility in background noise in individuals with hypophonia and PD. Two listeners rated the conversational speech intelligibility of 30 participants with hypophonia and 15 control participants. All participants completed conversation tasks in different background noise conditions, including: no added background noise, 60, 65, and 70 dB SPL. The authors found the conversational intelligibility of IWPD to be lower than that of control participants in the no added background noise condition, however; these results were not significant. The results of this study indicated that the conversational intelligibility of control participants decreased modestly as the intensity of background noise increased, with intelligibility values of 95.40%, 94.13%, and 85.03% in 60 dB SPL, 65 dB SPL, and 70 dB SPL, respectively. In contrast, the conversational intelligibility of IWPD decreased

differentially in comparison to that of control participants across all noise conditions, with intelligibility values of 77.47%, 68.98%, and 57.57% in 60 dB SPL, 65 dB SPL, and 70 dB SPL respectively. These results suggest that engaging in conversation in the presence of background noise has adverse effects on the speech intelligibility of individuals with hypophonia.

Based on the work of Adams and colleagues (2008) and Dykstra and colleagues (2012b), the conversational speech intelligibility of IWPD becomes increasingly reduced with increasing background noise. It has been established in the literature that speech intelligibility in this population deteriorates during the performance of more complex speech tasks (Kempler & Van Lancker, 2002; Tjaden & Wilding, 2011). Therefore, the findings of Adams and colleagues (2008) and Dykstra and colleagues (2012b) reveal that conversational speech intelligibility appears to capture a more ecologically valid measure of speech intelligibility than measures typically used in a clinical context with IWPD. Assessing speech intelligibility and evaluating the effectiveness of speech treatments in these ecologically valid contexts in addition to measures of sentence intelligibility can provide valuable information regarding the impact of hypophonia on communicative functioning.

1.5 Communicative Participation in Hypophonia and Parkinson's Disease

An individual's speech intelligibility may influence many factors related to one's participation in various social and communicative contexts. The construct of 'participation' is defined by the World Health Organization's (WHO) International Classification of Functioning, Disability, and Health (ICF) as "involvement in a life situation" (WHO, 2001). Participation is comprised of many components, including communication (WHO, 2001). Communication is a ubiquitous component of participation and daily life for all individuals, regardless of the presence of disease, disability, or communication disorder (Eadie et al., 2006).

Communicative participation is defined as "taking part in life situations where knowledge, information, ideas or feelings are exchanged. This may take the form of

speaking, listening, reading, writing, or nonverbal means of communication" (p. 309; Eadie et al., 2006). Communicative participation examines the communicative roles and actions in which individuals are engaged, as well as the frequency, quality, level of satisfaction, and effectiveness of their communicative endeavors (Baylor, Yorkston, Eadie, Miller, & Amtmann, 2009). It may also be used to assess the functional outcomes of impaired activities, such as speech intelligibility (Baylor et al., 2009). These measures may be used to provide a broader and more holistic perspective of an individual's experience with a disease or impairment.

IWPD and hypophonia may experience numerous changes in communicative functioning, including reduced communicative participation, reduced confidence in their speaking abilities, and feelings of embarrassment at the reactions of others (Fox & Ramig, 1997; Miller, Noble, Jones, & Burn, 2006). IWPD and their families perceive hypokinetic dysarthria as a highly disabling aspect of PD (Fox & Ramig, 1997; Ramig, Fox, & Sapir, 2008). Speech impairments, such as hypophonia, may be detrimental to the lifestyles and vocational abilities of IWPD (Jiang et al., 1999). The psychosocial impact of dysarthria may directly influence an individual's participation (Walshe & Miller, 2010). These individuals may be more likely to describe their communicative abilities as being impaired. They may also have increased concerns related to the effects of changes in speech production on their ability to communicate effectively (Kwan & Whitehill, 2011). Furthermore, IWPD believe that the impact of hypokinetic dysarthria on their communication abilities directly influences their social involvement (Miller et al., 2006). Thus, examining communicative participation in individuals with hypophonia and PD is a vital component of clinical assessment, as well as crucial for developing our understanding of the impact of hypophonia.

Patient-reported outcome measures (PROs) are particularly useful with regard to assessing an individual's communicative participation. PROs enable an individual with a communication disorder to report the subjective experiences and perspectives of their communicative interactions (Baylor et al., 2013). Three PROs that can be used to evaluate the communicative participation of individuals with motor speech disorders are the Communicative Participation Item Bank (CPIB; Baylor et al., 2013), the Voice

Activity and Participation Profile (VAPP; Ma & Yiu, 2001), and the Communicative Effectiveness Survey (CES; Donovan, Velozo, & Rosenbek, 2007). To date, few studies have explored communicative participation in IWPD using the CPIB (Baylor et al., 2014; McAuliffe, Baylor, & Yorkston, 2016) and VAPP (Simberg, Rae, Kallvik, Salo, & Martikainen, 2012). However, several studies have explored communicative effectiveness in IWPD and hypophonia (Donovan et al., 2007; Donovan, Kendall, Young, & Rosenbek, 2008; Dykstra et al., 2015). Furthermore, the retest reliability and repeatability of very few PROs exploring communicative participation have been explored in IWPD and hypophonia. A closer examination of these questionnaires may provide valuable insight into the validity of their repeated use over time in IWPD and hypophonia.

1.5.1 Communicative Participation Item Bank

The CPIB is used to evaluate the effect of an individual's speech and communication difficulties on their communicative participation. Seven hundred and one individuals with multiple sclerosis, PD, amyotrophic lateral sclerosis, or head and neck cancer completed the original 94-item CPIB in order to calibrate individual items using item response theory (Baylor et al., 2013). These analyses resulted in a 10-item questionnaire commonly referred to as the short form of the CPIB. Baylor and colleagues (2014) then explored the cross-cultural applicability of the CPIB. Two hundred eighteen IWPD in the United States and 210 IWPD in New Zealand completed the CPIB. Differential item analysis revealed no significant differences between the two groups, suggesting that the items and scoring of the CPIB is appropriate for IWPD in the United States and in New Zealand. While McAuliffe and colleagues (2016) later explored the variables associated with communicative participation in IWPD, no published studies have used the CPIB to explore communicative participation in individuals with PD and hypophonia as their primary dysarthric feature.

1.5.2 Voice Activity and Participation Profile

The VAPP was designed to assess the impact of an individual's self-perception of voice problems, activity limitations, and participation restrictions in individuals with voice disorders (Ma & Yiu, 2001). While originally validated on individuals with dysphonia and normal speakers, it has since been administered to IWPD with speech and voice

difficulties (Simberg et al., 2012). Simberg and colleagues (2012) sought to evaluate the impact of a 15-day treatment on the speech and voice in 6 IWPD. Prior to beginning the course of treatment, IWPD completed the VAPP in order to obtain self-ratings of voice function. Six months and one year following the treatment onset, individual self-ratings of voice function were evaluated using the VAPP. The authors found that participants' self-ratings of their overall VAPP scores decreased from 83 pre-treatment to 63 six months post-treatment. The post treatment self-ratings remained stable one year post-treatment, with overall VAPP self-ratings of 65. However, these changes were not significant, possibly due to increased standard deviation.

Simberg and colleagues (2012) also obtained proxy ratings of voice function for the IWPD. The spouses of the IWPD participating in the study evaluated their partner's voice impairments via the VAPP. Spouses completed the VAPP prior to their partner beginning the 15-day rehabilitation course, as well as six months and one year following the course onset. The authors reported that spousal ratings of their partner's voice functioning were less severe compared to the self-ratings of IWPD. Overall proxy-rated VAPP scores decreased from 77 pre-treatment to 44 six months post-treatment. However, overall proxy-rated VAPP scores increased to 56 one year post-treatment. Changes in proxy ratings of voice function over time were not significant. Yet, all spouses reported positive changes in the speech and voice of their partner with PD during post-treatment interviews. The authors concluded that PROs and proxy ratings provide valuable insight to the functional perspective of individuals with communication disorders.

1.5.3 Communicative Effectiveness Survey

Communicative participation is comprised of multiple components, including communicative effectiveness (Donovan et al., 2008). Communicative effectiveness can be defined as an individual's ability to successfully communicate in multiple settings in order to fulfill their various life and social roles (Donovan et al., 2008; Dykstra et al., 2015). In 2007, Donovan and colleagues developed the CES in order to evaluate an individual's perception of communicative effectiveness during various communicative interactions and contexts (Donovan et al., 2007; Donovan et al., 2008). The CES was originally validated for use in IWPD (Donovan et al., 2008).

Donovan and colleagues (2008) examined the communicative effectiveness of individuals with hypokinetic dysarthria and PD using the CES. Twenty-five IWPD and hypokinetic dysarthria, 25 participants without PD or dysarthria, and 25 primary communication partners of the IWPD used the CES to self-rate communicative effectiveness, or the communicative effectiveness of their partner with PD. Donovan and colleagues (2008) found that IWPD and dysarthria had lower CES scores than individuals without PD. Additionally, participants with dysarthria self-rated their own communicative effectiveness significantly higher than did their primary communicative partners. The authors suggest that the observed discrepancy in CES ratings between participants with PD and their primary communication partners may be related to reduced perceptual awareness on the part of IWPD with respect to their speech and communication difficulties. While this study was the first to explore communicative effectiveness in participants with hypokinetic dysarthria, the findings of this study may not be predictive of communicative effectiveness in individuals with hypophonia as their primary dysarthric feature.

In a recent study, Dykstra and colleagues (2015) sought to explore the relationship between speech intensity and self-rated communicative effectiveness in IWPD and hypophonia. Conversational intensity measures were obtained from 30 IWPD and hypophonia and 15 control participants. All participants also completed the CES to obtain a measure of self-rated communicative effectiveness. The authors found that the habitual conversational speech intensity of IWPD was approximately 5 dB lower than that of control participants.

Additionally, Dykstra and colleagues (2015) found that IWPD and hypophonia experienced reduced levels of self-perceived communicative effectiveness compared to control participants. IWPD self-reported reduced communicative effectiveness in communicative contexts and situations related to conversing over distances and conversing in background noise. These results suggest there may be a hierarchy of communicative situations that individuals with hypophonia find the most challenging.

Dykstra and colleagues (2015) also evaluated proxy ratings of communicative effectiveness made by the primary communication partners of participants with hypophonia and PD. Ratings of communicative effectiveness were similar between IWPD and their primary communication partners. These findings are in contrast to those of Donovan and colleagues (2008) who found that IWPD self-rated communicative effectiveness higher than did their primary communication partners. The difference in the findings of Donovan and colleagues (2008) and Dykstra and colleagues (2015) may be due to factors such as severity, or salient dysarthric features. For example, Donovan and colleagues (2008) examined individuals with a diagnosis of hypokinetic dysarthria. These participants likely presented with a range of dysarthric features associated with hypokinetic dysarthria (i.e., articulatory imprecision, prosodic abnormalities, impairments in speech rate; Donovan et al., 2008), whereas Dykstra and colleagues (2015) evaluated participants with hypophonia as their primary dysarthric feature.

Finally, Dykstra and colleagues (2015) identified a non-significant relationship between speech intensity measures and participants' self-ratings of communicative effectiveness. These results suggest that communicative participation may be a distinct construct that differs from perceptual or acoustic outcome measures, such as measures of speech intelligibility or speech intensity. While Dykstra and colleagues (2015) explored the relationship between speech intensity and communicative effectiveness in IWPD, other studies have explored relationships among other demographic and non-speech factors in PD.

1.6 Demographic and Non-Speech Factors in Hypophonia and Parkinson's Disease

McAuliffe and colleagues (2016) identified variables associated with communicative participation in IWPD. Three hundred seventy-eight IWPD in the United States and New Zealand completed the CPIB. Participants also provided information for possible predictors of communicative participation. These possible predictors included self-rated communication disorder severity, individual speech usage, hearing, cognition, physical activity, fatigue, pain, swallowing difficulties, and emotional problems. Backward stepwise linear regression was used to assess the relationship between communicative

participation and possible predictive factors. The authors found communicative participation to be significantly related to perceptions of speech severity, with mild speech difficulty perceptions associated with greater communicative participation. Increased level of speech usage was also associated with increased communicative participation. Increased levels of fatigue, cognitive issues, emotional difficulties, and swallowing problems were all negatively associated with communicative participation. Perceptions of emotional difficulties were obtained via one general question regarding the impact of "emotional problems such as feeling anxious, depressed or irritable" (McAuliffe et al., 2016, p.11). While no published studies have explored the relationship between communicative participation and depression in IWPD and hypophonia, Yorkston and colleagues (2008) explored the relationships between participation and personal factors, including mobility, depression, general health, fatigue, and pain, in individuals with multiple sclerosis. Additionally, older individuals reported an overall greater level of communicative participation. Gender effects were observed, with men demonstrating increased communicative participation as compared to women.

Other studies have reported mixed results regarding gender effects in IWPD (Fox & Ramig, 1997; Sapir et al., 2007). While exploring the impact of LSVT on vowel intensity, Sapir and colleagues (2007) found significant gender differences in the intensity measures of the vowels /i/, /a/, and /u/, across all treatment conditions. However, while examining the impact of LSVT, Fox and Ramig (1997) did not find gender differences in IWPD in the speech intensity across four speech tasks, or in self-ratings of nine perceptual variables.

While investigating the construct validity of the CES for IWPD and dysarthria, Donovan and colleagues (2008) found that 47% of the variability in CES scores was accounted by for the Hoehn and Yahr staging in IWPD. However, compared to the Unified Parkinson's Disease Rating Scale (UPDRS), the Hoehn and Yarh staging scales do not provide measures of motor ability. In a longitudinal study, Skodda, Flasskamp, and Schlegel (2011) explored the stability of motor speech performance in order to find possible markers of disease progression in IWPD. Fifty-eight IWPD and 35 control participants were tested and retested a minimum of twelve months later. During each visit,

participants performed a syllable repetition task, and their motor abilities were assessed via the UPDRS. The authors did not find a correlation between variability of syllable repetition and general motor impairment. The authors concluded that the underlying mechanism contributing to instability of speech measures might be independent from dopaminergic deficits. However, this study only examined the stability of syllable repetition over time. Additional studies are required to explore the stability of speech intensity and speech intelligibility over time, and to examine the relationship between speech intensity, speech intelligibility, communicative participation, and motor symptoms associated with PD in IWPD and hypophonia.

1.7 Rationale for the Current Study

It appears that few published studies have investigated the variability of habitual speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and PD. Fox and Ramig (1997) and Schalling and colleagues (2013) appear to be the only published studies to have directly explored the day-to-day variability of speech intensity in individuals with hypophonia and PD. Both of these studies reported individual variability within participants, as demonstrated by large standard deviations. Additionally, the variability (or stability) of an individual's response to background noise has received minimal attention. In addition to increasing speech intensity, background noise may also influence listener perceptions of hypophonia severity, and potentially introduce communication challenges, such as reducing the speech intelligibility (Adams et al., 2008; Dykstra et al., 2012b) and communicative participation (Baylor et al., 2011) in individuals with hypophonia. Fox and Ramig (1997) appear to be the only published study to explore the day-to-day variability in self-perceived ratings of speech intelligibility and communicative participation in IWPD. However, no published studies have examined the variability of speech intelligibility and communicative participation with PRO measures in this population.

While the temporal variability of self-perception of loudness, speech intelligibility, and communicative participation have not been directly evaluated, Fox and Ramig (1997) explored perceptual self-ratings speech and voice characteristics of the IWPD during three visits within a four-day period. These characteristics included loudness, being

understood by others, and the ability to participate in conversations. Large standard deviations were associated with these ratings. Perceived loudness ratings ranged from 51.62 – 57.80, with standard deviation values ranging from 12.89 – 16.78 over the three days. Perceived intelligibility ratings ranged from 54.71 – 56.96, with standard deviation values ranging from 13.99 – 18.21 over the three days. Perceived communicative participation ratings ranged from 57.64 – 61.50, with standard deviation values ranging from 18.65 – 20.21 over the three days. It may be possible that the large standard deviations suggest variability in individual and day-to-day responses of IWPD. While Fox and Ramig (1997) explored individuals' overall self-perception of loudness, speech intelligibility, and participation, the temporal or day-to-day variability of the speech intelligibility and communicative participation in IWPD and hypophonia has not been systematically examined.

Multi-baseline studies are needed to explore the temporal variation and fluctuations of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and PD. Such studies are needed in order to develop a deeper understanding of the stability of these measures in IWPD and hypophonia. Further understanding the fluctuations of these measures in individuals with hypophonia may assist in the development of clinical tools that can be reliably used to ascertain the severity and impact of hypophonia.

Several published studies have attempted to delineate potential relationships among acoustic and psychosocial measures, and demographic (i.e., age, gender, level of education, disease duration) and non-speech (i.e., disease severity, depression, cognition, motor symptoms associated with PD) factors in IWPD (Donovan et al., 2008; Fox & Ramig, 1997; McAuliffe et al., 2016; Sapir et al., 2007; Skodda et al., 2011). However, many of these studies explored these relationships in IWPD who did not demonstrate hypophonia as their primary dysarthric feature. Exploring whether these relationships exist in a population of IWPD and hypophonia will provide a broader and more holistic understanding of hypophonia. Additionally, it has been proposed that all facets of living with a disease or disability must be examined in order to develop a complete understanding of the particular disease or disability (Morse & Johnson, 1991). Since

hypophonia has the potential to affect all facets of an IWPD's life, it is important to explore the impact of and relationship among as many facets of an individual's life as possible. Examining hypophonia from such a perspective may allow us to better understand the mechanisms for speech intensity regulation and the impact of hypophonia on an individual's functional communicative abilities. Exploring relationships among acoustic measures, self-rated communicative participation, demographic, and non-speech factors in IWPD and hypophonia may help to guide the clinical management of speech and communication difficulties as the disease progresses.

1.8 Purpose

The purpose of this study is to systematically examine the temporal and day-to-day variability of acoustic and perceptual speech measures and psychosocial measures in individuals with hypophonia and PD. An additional purpose of this study is to examine the relationships among speech intensity measures, speech intelligibility measures, self- and proxy ratings of communicative participation, demographic factors, and non-speech factors. Speech intensity measures will include habitual speech intensity, magnitude production intensity, Lombard response intensity, and self- and proxy perceptions of typical loudness. Speech intelligibility measures will include habitual sentence intelligibility, and conversational speech intelligibility. Communicative participation measures will include self- and proxy-rated CES scores, self- and proxy-rated CPIB sores, self- and proxy-rated VAPP scores, and self- and proxy-rated Level of Speech Usage Scale (LSUS) scores. Demographic factors will include age, gender, and disease duration. Non-speech factors will include depression, disease severity, cognition, and motor symptoms associated with PD, as measures by the UPDRS. Exploring the variability within these variables, as well as defining relationships among these variables will provide a deeper clinical understanding that is crucial for the assessment and management of individuals with hypophonia and PD.

Nine objectives will be examined in this study. These objectives are:

1. To examine the temporal variability of speech intensity measures in participants with PD and control participants over three time points.

2. To evaluate the temporal variability of perceived typical speech loudness in participants with PD, their primary communication partners, and control participants over three time points.

3. To evaluate the temporal variability of speech intelligibility measures in participants with PD and control participants over three time points.

4. To evaluate the temporal variability of self-rated communicative participation in participants with PD and control participants over three time points.

5. To evaluate the temporal variability of self- and proxy-rated communicative participation in participants with PD and their primary communication partners.

6. To examine the retest reliability and repeatability of measures of speech intensity, speech intelligibility, and communicative participation in participants with PD, their primary communication partners, and control participants. This objective has been addressed and embedded within Objectives 1-5.

7. To evaluate relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation, demographic factors, and non-speech factors for participants with PD.

8. To evaluate relationships among speech intensity measures, speech intelligibility measures, proxy-rated communicative participation, demographic factors, and non-speech factors for participants with PD.

9. To evaluate relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, non-speech factors for control participants.

Six hypotheses are predicted for this study. It is hypothesized that:

1. IWPD will demonstrate greater temporal variability of speech intensity measures, speech intelligibility measures, and self-rated communicative participation measures compared to control participants.

2. IWPD will demonstrate reduced speech intensity measures, speech intelligibility measures, and self-rated communicative participation measures compared to control participants.

3. The self-rated typical speech loudness and self-rated communicative participation will be reduced compared to proxy-rated typical speech loudness and self-rated communicative participation for participants with PD.

4. Measures of speech intensity, speech intelligibility, and communicative participation will demonstrate good retest reliability and acceptable repeatability for participants with PD, their primary communication partners, and control participants.

5. Measures of self- and proxy-rated communicative participation will all be correlated with one another for participants with PD and control participants.

6. Increased disease duration, increased depression, increased UPDRS scores, decreased perceived mediation effectiveness, and reduced cognition will be correlated with measures of speech intensity, speech intelligibility, and self- and proxy-rated communicative participation in participants with PD.

This study was approved by the Health Sciences/Lawson Research Ethics Board at Western University, London, Ontario, Canada.

Chapter 2

2 Methods

2.1 Participants

2.1.1 Participants with Parkinson's disease

Thirty individuals with idiopathic PD aged 58 – 82 years, *M* = 70.2 years (*SD* = 5.46), male = 21, female = 9 were recruited to participate in this study. IWPD were referred for this study as they presented with hypokinetic dysarthria, and hypophonia as their primary distinctive speech feature as confirmed by a neurologist. However, seven IWPD were excluded from the present study as they did not meet our inclusion criteria of presenting with hypophonia. Thus, a total of 23 individuals with idiopathic PD aged 58 – 82 years, *M* = 69.48, (*SD* = 5.57), male = 16, female = 7 were included in the analysis of the present study.

Inclusion criteria for participants with PD included: having a diagnosis of idiopathic PD for a minimum of 3 years, being between 55 – 85 years of age, having a diagnosis of hypophonia, and speaking English as their first language. Exclusion criteria for participants with PD included: an inability to read and/or write English, a positive history of speech, language, or neurological impairments, except those related to PD, an inability to pass a 40 dB hearing level (HL) hearing screening at 500, 1000, and 2000 Hertz (Hz) in at least one ear, except in those individuals whose hearing was aided via hearing aids, receiving deep brain stimulation surgery as treatment for PD, currently receiving speech-language therapy, receiving a score below 21 on the Montreal Cognitive Assessment (MOCA; Nasreddine, et al., 2005), at the time of their last visit to their neurologist.

All participants with PD completed a cognitive screening using the MOCA (Nasreddine, et al., 2005), a depression screening using the Geriatric Depression Scale - short form (GDS-15; Sheikh & Yesavage, 1986), a Medication Effectiveness Scale (MES) and a motor assessment using the motor examination section (part 3) of the UPDRS (Goetz, et al., 2007). IWPD were stable on their anti-parkinsonian medication and were tested in an

"on" state. All participants attended three experimental sessions, with the exception of one participant who was unable to attend visit 3 due to scheduling difficulties. Participant visits took place at 9 am or 1 pm, and testing time remained consistent for each participant throughout the study. IWPD were recruited through the Movement Disorders Clinic at University Hospital in London, Ontario.

2.1.2 Control participants

Thirty healthy control participants aged 55 – 82 years, *M* = 69.58 years, (*SD* = 7.66), male = 9, female = 21 also took part in this study. Inclusion criteria for control participants included: not having a diagnosis of PD, being between 55 – 85 years of age, and speaking English as their first language. Exclusion criteria for control participants included: an inability to read and/or write English, a positive history of speech, language or neurological impairments, an inability to pass a 40 dB HL hearing screening at 500, 1000, and 2000 Hz in at least one ear, except in those individuals whose hearing was aided via hearing aids, and receiving a score below 26 on the MOCA.

Control participants completed a cognitive screening using the MOCA (Nasreddine, et al., 2005) and a depression screening using the GDS-15 (Sheikh & Yesavage, 1986). Control participants were also required to attend all three experimental sessions. Control participants were primarily recruited via convenience sampling through the Canadian Centre for Activity and Aging.

2.1.3 Primary communication partners of participants with Parkinson's disease

Twenty-three primary communication partners of participants with PD aged 48 – 81 years, *M* = 68.13 years (*SD* = 7.18), male = 5, female = 18 were included in this study. A primary communication partner is defined as an individual having daily or frequent contact with the participant with PD. Primary communication partners included spouses, siblings, friends, and/or caregivers of participants with PD. Inclusion criteria for primary communication partners included: being 18 years of age or older, speaking English as their first language, and having daily or frequent contact with the participant with PD. Exclusion criteria for primary communication partners included: an inability to write

and/or read English, an inability to pass a 40 dB HL hearing screening at 500, 1000, and 2000 Hz in at least one ear, except in those individuals whose hearing was aided via hearing aids, and not having daily or frequent communication with the participant with PD. Primary communication partners were also required to attend all three experimental sessions with their communication partner with PD.

2.1.4 Listener participants

Three female experienced listeners aged 22 – 28 years, *M* = 25 years (*SD* = 3.00) took part in this study. Inclusion criteria for listeners included: being 18 years of age or older and speaking English as their first language. Exclusion criteria for listeners included: an inability to read and/or write English, a positive history of speech, language, hearing, or neurological impairments, and an inability to pass a 25 dB HL bilateral hearing screening at 500, 1000, 2000, and 4000 Hz.

2.2 Materials

2.2.1 Apparatus

Participants with PD and control participants were seated in a quiet laboratory environment in Elborn College at Western University for all three experimental sessions. Participants were seated 1.5 metres (m) away from the experimenter throughout each session. Participants' speech was audio recorded throughout each session via a headset microphone (AKG C520) and a second microphone mounted 1 m off the ground on a desk and tripod, at a distance of 2 m from participants. The headset microphone was located six centimetres (cm) from each participants' mouth. The microphone placement was verified prior to calibration, speech intensity tasks, and speech intelligibility tasks. The headset microphone was calibrated using a sound level meter placed 15 cm from participants' mouths. During calibration, participants produced a steady and prolonged 'ah'. The intensity of a steady segment was noted in dBA SPL, as indicated on the sound level meter. Background noise was presented to participants over a loudspeaker (M Audio AV40) located 2 m away from participants. Speech samples were presented to listeners over loudspeakers (M Audio AV40), located 60 cm away from participants during separate sessions.

2.2.2 Noise source

A recording of multi-talker noise (Audiotech – 4 talker noise) was presented to participants during a background noise speech task. The multi-talker noise was presented at four different intensity increments ranging from 60 to 75 dB SPL. The presentation order of background noise intensity level was randomized across all participants and sessions. The presentation order was determined using www.randomizer.org.

2.2.3 Audio recordings

The speech of participants with PD and control participants was recorded using Praat (Boersma & Weenink, 2018), with a recording sampling rate of 22.05 kilohertz (kHz). The audio recordings were stored in an uncompressed (.wav) file format. The software Praat (Beorsma & Weenink, 2018) was used to analyze the recorded audio files and generate audio clips later used to assess speech intelligibility.

2.2.4 Speech intensity measures

The sentence "she saw Patty buy two poppies" was used during habitual speech intensity, maximum speech intensity, Lombard response function, and magnitude production tasks. Praat (Boersma & Weenink, 2018) was used to determine the speech intensity produced by participants during each of these tasks. The first and last syllables of the sentence "she saw Patty buy two poppies" were excluded from speech intensity calculations. During the Lombard response function task, participants produced the target sentence using their habitual speech intensity in 60, 65, 70, and 75 dB SPL of multi-talker background noise. The Lombard response function was then determined by calculating the slope of speech intensity when participants repeated the target sentence in 60, 65, 70, and 75 dB SPL of multi-talker background noise. Participants also repeated the target sentence at their habitual speech intensity, two times louder than their habitual speech intensity, four times louder than their habitual speech intensity, and their maximal speech intensity. The magnitude production function was determined by calculating the slope of speech intensity when participants repeated the target sentence at their habitual speech intensity, two time louder than their habitual speech intensity, and four times louder than their habitual speech intensity.

2.2.5 Speech intelligibility measures

2.2.5.1 Sentence Intelligibility Test

The SIT (Yorkston et al., 2011) was used to obtain speech intelligibility measures. The SIT consists of 11 randomly generated sentences of increasing word count. The length of these sentences ranges from 5 to 15 words. Each participant read a unique list of 11 randomly generated sentences, presented on an 8 ½ by 11inch piece of white paper in 18 point Times New Roman font. However, only those SIT sentences comprising of 13, 14, and 15 words were evaluated by listeners.

2.2.5.2 Conversational intelligibility

Conversational intelligibility samples were obtained by asking participants to discuss a familiar topic for two minutes. Participants were prompted with the following discussion points:

1. Tell me about the jobs you've had.

2. Tell me about the hobbies you enjoy.

3. Tell me about your favourite vacation.

Secondary questions were asked for each topic if prompting for additional conversation was needed.

Conversation samples were then pared down to include recorded conversational utterances, with excerpts ranging in length from 15 – 20 seconds. These conversational excerpts were then presented to listeners in order to obtain ratings of conversational intelligibility.

2.2.6 Questionnaires

2.2.6.1 Intake questions

Participants completed intake questions in order to collect demographic information. The demographic information obtained for IWPD included partial date of birth, gender,

occupation, time since onset of PD, time since PD symptom onset, general health, and history of speech therapy. See Table 1 for a complete description of participants with PD. The demographic information obtained for control participants included partial date of birth, gender, occupation, general health, and history of speech therapy. See Table 2 for a description of control participants. The demographic information obtained for primary communication partners of participants with PD included partial date of birth, gender, and relationship to their communication partner with PD. See Table 3 for a description of primary communication partners.

Table 1: Description of Demographic and Non-Speech Factors for Participants with Parkinson's Disease

Participant ID	Age (years)	Gender	Disease duration (years	GDS-15 score	MES score	MOCA score	UPDRS score
1	72	Male	10	3	6	22	24
2	75	Male	5	1	3	20	N/A*
3	72	Female	9	2	7	29	27
4	62	Male	10	2	6	25	18
5	74	Male	3	2	3	28	51
6	58	Female	9	15	4	23	40
7	66	Female	8	0	5	29	37
8	70	Male	5	3	2	24	31
9	67	Female	11	2	5	27	56
10	73	Male	10	4	5	22	66
11	72	Male	9	3	6	21	34
12	63	Male	15	3	5	20	36
13	67	Female	6	2	5	28	33
14	66	Male	10	3	4	25	40
15	75	Male	22	4	5	16**	56
16	66	Male	29	0	5	25	42
17	69	Male	5	0	2	28	35
18	71	Male	7	3	5	24	43
19	64	Female	15	1	5	28	24
20	69	Male	13	1	4	27	22
21	79	Male	6	0	4	26	30
22	66	Male	13	3	4	23	43
23	82	Female	14	0	6	28	40

Note. This table describes the demographic and non-speech factors of participants with PD. *Participant 2 was not able to attend visit 3 due to scheduling conflicts. As a result, the UPDRS was not completed for this participant. **Participant 15 did not wish to complete the delayed recall and fluency sections of the MOCA. As a result, his MOCA score was calculated out of a total of 24 points instead of 30 points. GDS-15 scores between 0 – 4 fall within the normal range, scores between 5 – 8 are indicative of mild depression, scores between 9 – 11 are indicative of moderate depression, scores between 12 – 15 are indicative of severe depression (Sheikh & Yesavage, 1986). The UPDRS reflects the severity of PD motor symptoms with a total possible score of 108.

Table 2: Description of Demographic and Non-Speech Factors for Control Participants

Participant ID	Age (years)	Gender	GDS-15 score	MOCA score
1	55	Female	3	30
2	66	Female	0	30
3	55	Female	0	26
4	65	Male	0	29
5	69	Female	4	28
6	74	Female	0	26
7	79	Male	1	27
8	68	Female	0	27
9	68	Female	1	30
10	66	Female	0	28
11	72	Male	1	29
12	74	Female	0	30
13	75	Male	1	26
14	71	Female	0	29
15	63	Female	0	30
16	58	Female	0	26
17	78	Male	0	26
18	77	Female	1	28
19	72	Female	0	29
20	82	Female	2	29
21	72	Female	0	27
22	60	Female	1	29
23	62	Female	2	28
24	81	Male	0	26
25	66	Female	0	28
26	61	Female	0	29
27	63	Male	0	26
28	70	Male	1	27
29	79	Female	0	30
30	80	Male	0	30

Note. This table describes the demographic and non-speech factors of control participants. GDS-15 scores between 0 – 4 fall within the normal range, scores between 5 – 8 are indicative of mild depression, scores between 9 – 11 are indicative of moderate depression, scores between 12 – 15 are indicative of severe depression (Sheikh & Yesavage, 1986).

Table 3: Description of Demographic and Non-Speech Factors for Primary Communication Partners of Participants with Parkinson's Disease

Participant ID	Age (years)	Gender	Relationship to participant with PD
1	66	Female	Spouse
2	71	Female	Sibling
3	62	Male	Spouse
4	62	Female	Spouse
5	75	Female	Spouse
6	48	Male	Friend
7	67	Male	Spouse
8	69	Female	Spouse
9	63	Male	Sibling
10	71	Female	Spouse
11	67	Female	Spouse
12	61	Female	Spouse
13	73	Male	Spouse
14	70	Female	Spouse
15	72	Female	Spouse
16	64	Female	Spouse
17	65	Female	Spouse
18	71	Female	Spouse
19	67	Female	Friend
20	81	Female	Spouse
21	81	Female	Spouse
22	64	Female	Spouse
23	77	Female	Friend

Note. This table describes the age, gender, and relationship of primary communication partners to the participants with PD.

2.2.6.2 Montreal Cognitive Assessment

The MOCA is a 10-minute screening tool used to detect cognitive impairment (Dalrymple-Alford, et al., 2010; Nasreddine et al., 2005). The MOCA assesses individuals on eight cognitive domains, including visuospatial/executive, naming, memory, attention, language, abstraction, delayed recall, and orientation. Additionally, several alternate forms of the MOCA have been developed and found to be valid for serial assessments of cognitive impairment in order to avoid practice effects (Costa et al., 2012). The MOCA appears to be a valid and reliable tool for detecting cognitive impairment in IWPD (Dalrymple-Alford, et al., 2010; Gill, Freshman, Blender, & Ravina, 2008; Hoops et al., 2009;).

2.2.6.3 Geriatric Depression Scale

The GDS-15 is a 15-item self-report measure used to assess an individual's non-somatic symptoms of depression, such as the psychological features and the social repercussions associated with depression (Sheikh & Yesavage, 1986). Each item is evaluated via a 'yes' or 'no' response (Sheikh & Yesavage, 1986). The GDS-15 has been found to be a valid and reliable tool to screen for depression in IWPD (Schrag et al., 2007; Weintraub, Oehlberg, Katz, & Stern, 2006; Weintraub, Saboe, & Stern, 2007a; Weintraub, Xie, Karlawish, & Siderowf, 2007b). See Appendix A to view the GDS-15.

2.2.6.4 Medication Effectiveness Scale

Participants with PD evaluated their medication effectiveness on the MES. Participants indicated on a 7-point scale the effectiveness of their medication at managing their symptoms of PD. The anchor of "not at all effective" was associated with a score of 1. The anchor of "very effective" was associated with a score of 7. See Appendix B to view the Medication Effectiveness Scale.

2.2.6.4 Typical Loudness Scale

Participants evaluated typical loudness on a Typical Loudness Scale (TLS) using visual analogue scale (VAS) estimation. A VAS measuring 10 cm was presented to participants with the anchors "very quiet" and "normal loudness". See Appendix C to view the Typical Loudness Scale. Participants rated typical loudness by indicating their rating with an X or a dash on the VAS.

2.2.6.5 Communicative Effectiveness Survey

The CES is an eight-item questionnaire using a four-point equal appearing interval scale (Donovan et al., 2007). The CES assesses an individual's perception of their own communicative effectiveness in different communicative interactions and contexts (Donovan et al., 2007; Donovan et al., 2008). The CES was adapted from the Communicative Effectiveness Index - Modified (CETI-M; Ball, Beukelman, & Pattee, 2004). The CETI-M was modified from the Communicative Effectiveness Index (CETI), which was originally designed to assess changes in an individual's communicative effectiveness over time (Lomas et al., 1989). The CETI was designed as a proxy measure

to assess changes in the communicative effectiveness of individuals with aphasia (Lomas et al., 1989). The CETI was found to have acceptable test-retest reliability (Lomas et al., 1989). Since the CES was adapted from a modified CETI, it is also used to assess changes in an individual's communicative effectiveness over time (Donovan et al., 2007; Donovan et al., 2008). Thus, repeated administration of the CES to participants in this study is considered to be an appropriate procedure. While the CETI was originally validated for use in adults with aphasia (Lomas et al., 1989), the CES was originally validated for use in IWPD (Donovan et al., 2007). See Appendix D to view the CES.

2.2.6.6 Voice Activity and Participation Profile

The VAPP is a 28-item, self-report questionnaire (Ma & Yiu, 2001). The VAPP evaluates an individual's perception of their voice problem, their activity limitations, and restrictions to their participation, in accordance with the definitions put forth by the WHO (1997; Ma & Yiu, 2001). The VAPP is used to explore the effects of an individual's voice impairment on their job, daily communication, social communication, and emotions, as well as self-perceptions of their voice problem. Originally validated in individuals with dysphonia, the VAPP has since been validated for use with a variety of individuals with speech and voice concerns (Bassi et al., 2011; Dragone, 2011; Duarte De Almeida, Santos, Bassi, Teixeira, & Côrtes Gama, 2013; Dykstra, Adams, & Jog, 2007; Kleemola, Helminen, Rorarius, Sihvo, & Isotalo, 2011b; Martinello, Lauris, & Brasolotto, 2011; Piwowarczyk, Oliveira, Loureno, & Behlau, 2012; Simberg et al., 2012), as well as cross-culturally (Behlau, Oliveira, dos Santos, & Ricarte, 2009; Fava, Paolillo, Oliveira, & Behlau, 2014; Kleemola, Helminen, Rorarius, Isotalo, & Sihvo, 2011a; Ricarte, Oliveira, & Behlau, 2013; Sukanen et al., 2007;). Additionally, the original VAPP and its cross-cultural adaptations demonstrate good test-retest reliability (Kleemola, et al., 2011a; Ma & Yiu, 2001; Ricarte et al., 2013; Sukanen, et al., 2007;). See Appendix E to view the VAPP.

2.2.6.7 Communication Participation Item Bank

The CPIB short form is a 10-item questionnaire using a four-point equal appearing interval scale and is used to assess the impact of an individual's speech and communication difficulties on their communicative participation (Baylor et al., 2013;

Baylor et al., 2009). The CPIB was developed according to the principles of Item Response Theory and validated for use across individuals with a wide variety of communication disorders (Baylor et al., 2013; Baylor et al., 2009). The CPIB has also been validated cross-culturally in IWPD (Baylor et al., 2014). However, no published study has assessed the test-retest reliability of the CPIB. See Appendix F to view the CPIB.

2.2.6.8 Level of Speech Usage Scale

The LSUS is a self-report categorical scale designed for use with adults with any number of communication disorders (Baylor, Yorkston, Eadie, Miller, & Amtmann, 2008). While completing this questionnaire, participants were instructed to select the everyday degree of speech usage from five different categories: undemanding, intermittent, routine, extensive, and extraordinary (Baylor et al., 2008). However, no published study has assessed the test-retest reliability of the LSUS. See Appendix G to view the LSUS.

2.3 Procedures

All participants attended a total of three experimental visits, with each visit ranging from 45 – 90 minutes in duration. Visit 1 denoted the start of the study, visit 2 took place two to five day following visit 1, and visit 3 took place four weeks following visit 2.

2.3.1 Visit 1

All participants read a detailed letter of information and provided their written consent prior to beginning the study. The Letter of Information provided to participants with PD is found in Appendix H. The Consent Form signed by participants with PD is provided in Appendix I. The Letter of Information provided to control participants is found in Appendix J. The Consent Form signed by control participants is provided in Appendix K. The Letter of Information provided to primary communication partners is found in Appendix L. The Consent Form signed by primary communication partners is provided in Appendix M. A hearing screening was conducted on all participants. Participants with PD and control participants completed the GDS-15 and the MOCA. All participants then completed five questionnaires (TLS, VAPP, CPIB, LSUS, and CES). Proxy-measures for the participants with PD for each questionnaire were collected from their primary

communication partners. The presentation order of the scales was randomized across participants and visits. The presentation order was determined using www.randomizer.org.

Participants with PD and control participants then completed various speech tasks. Participants completed maximum sustained vowel phonation and diadochokinetic tasks. Praat was used to generate voice reports indicating jitter, shimmer, and signal-to-noise ratio for the maximum sustained vowel phonation task. The rate of syllable production (syllables per second) for /p∧/, /t∧/, and /k∧/ was manually counted using Praat. Table 4 provides diadochokinetic task rates for participants with PD and Table 5 provides description of voice report summaries for participants with PD. Table 6 provides diadochokinetic task rates for control participants and Table 7 provides a description of voice report summaries for control participants. Various speech intensity measures were then collected. The sentence "she saw Patty buy two poppies" was used during habitual and maximum speech intensity tasks. Participants also produced this sentence using their habitual speech intensity in 60, 65, 70, and 75 dB SPL of multi-talker background noise. Participants also completed a magnitude production task, wherein they repeated the sentence at their habitual speech intensity, two times louder than their habitual speech intensity, four times louder than their habitual speech intensity, and their maximal speech intensity. Participants then completed the SIT and engaged in two minutes of conversation with the experimenter.

Table 4: Description of Diadochokinetic Rates (syllables/second) for Participants with Parkinson's Disease Across All Three Visits

Participant ID	Visit 1			Visit 2			Visit 3		
	/p∧/	/t∧/	/k∧/	/p∧/	/t∧/	/k∧/	/p∧/	/t∧/	/k∧/
1	5.64	5.79	5.20	6.14	6.43	5.62	5.80	5.81	5.36
2	5.01	4.09	4.64	5.02	4.89	5.30	N/A*	N/A*	N/A*
3	6.74	5.50	5.47	6.39	5.67	5.71	6.51	5.97	5.88
4	6.34	5.95	4.96	5.49	5.31	4.64	6.47	5.69	4.98
5	7.27	7.67	5.99	7.05	6.00	6.75	7.18	7.28	6.00
6	5.70	5.98	5.61	5.96	5.70	5.26	6.03	5.79	5.16
7	5.82	4.67	4.39	4.74	4.62	4.29	5.65	5.07	4.45
8	6.44	5.49	3.95	6.00	5.86	5.52	6.29	5.07	5.09
9	5.77	5.51	5.18	5.59	5.32	4.91	5.59	5.38	4.97
10	5.50	5.03	4.86	5.96	4.81	4.84	5.92	5.06	5.09
11	5.67	5.30	5.26	4.99	5.43	5.48	5.93	5.26	5.27
12	6.23	6.07	5.65	6.31	5.78	5.44	6.69	6.45	5.95
13	6.39	5.94	5.73	6.49	5.92	5.62	6.34	6.02	5.67
14	5.19	5.67	5.38	5.75	5.55	5.40	5.59	5.10	5.15
15	5.18	4.27	4.47	5.35	4.95	4.61	5.36	4.77	4.73
16	6.88	5.78	3.20	7.00	5.49	4.08	6.42	5.72	4.88
17	6.61	4.64	3.39	5.87	3.51	3.40	6.16	3.81	3.45
18	6.87	7.14	7.68	6.43	7.05	6.89	6.65	6.31	6.17
19	6.35	6.18	5.41	6.24	5.84	5.02	6.09	6.10	5.06
20	4.75	4.94	4.11	4.25	4.95	3.93	4.08	3.94	3.56
21	5.58	5.24	4.86	5.40	5.19	5.01	5.55	5.17	5.05
22	6.81	7.02	6.73	6.26	6.19	6.13	6.54	6.46	5.76
23	6.39	5.00	4.19	6.42	4.98	4.46	7.03	4.47	4.81

Note. This table describes the diadochokinetic rate (syllables/second) data collected from participants with PD across all 3 visits. *Participant 2 was not able to attend visit 3 due to scheduling conflicts.

Table 5: Description of Voice Quality for Participants with Parkinson's Disease Across All Three Visits

Participant ID	Visit 1			Visit 2			Visit 3		
	Jitter	Shimmer	Signal-to-noise	Jitter	Shimmer	Signal-to-noise	Jitter	Shimmer	Signal-to-noise
1	0.55	1.72	22.91	0.39	1.53	24.21	0.43	2.21	22.99
2	0.77	3.72	18.58	0.60	2.55	21.57	N/A*	N/A*	N/A*
3	0.14	1.36	30.60	0.23	1.62	28.04	0.15	1.37	29.84
4	0.46	2.22	22.46	0.55	3.69	20.74	0.42	2.53	22.48
5	0.37	1.95	26.03	0.41	2.20	22.30	0.55	2.29	23.54
6	0.79	4.01	21.85	0.55	3.18	21.41	0.71	4.81	19.66
7	0.48	3.70	26.13	0.61	4.81	23.42	0.38	3.19	26.52
8	0.98	4.77	17.32	1.20	6.87	16.53	0.91	6.28	17.56
9	0.32	1.52	25.46	0.29	1.96	25.46	0.27	1.21	28.39
10	0.75	4.98	20.18	1.17	3.80	20.61	0.92	4.81	18.70
11	1.07	8.32	14.55	1.00	9.73	14.63	1.05	8.49	13.23
12	0.64	3.06	22.54	1.26	4.20	18.80	0.47	2.19	25.20
13	0.53	2.77	21.10	0.27	0.94	24.78	0.37	2.22	21.66
14	0.75	7.83	19.80	0.50	3.81	21.42	0.59	3.37	20.22
15	2.06	5.70	18.21	2.71	6.96	17.09	0.87	2.93	20.80
16	0.89	3.17	21.93	0.55	3.33	21.18	1.03	2.94	17.28
17	0.34	2.45	23.65	0.23	1.03	27.07	0.49	3.28	21.94
18	0.56	4.02	23.33	0.35	2.43	27.71	0.37	1.86	27.80
19	0.20	1.23	33.48	0.19	1.18	32.80	0.22	1.23	32.67
20	0.42	1.47	28.59	0.24	1.08	31.96	0.38	2.01	29.24
21	0.54	5.55	19.42	0.54	3.63	20.60	0.39	2.54	23.30
22	0.49	3.17	24.15	1.00	4.52	19.88	0.85	5.21	20.05
23	3.91	8.76	8.38	1.67	5.90	14.46	0.64	4.64	21.76

Note. This table describes the voice quality data collected from participants with PD across all 3 visits. *Participant 2 was not able to attend visit 3 due to scheduling conflicts.

Table 6: Description of Diadochokinetic Rates (syllables/second) for Control Participants Across all Three Visits

Participants ID	Visit 1			Visit 2			Visit 3		
	/p∧/	/t∧/	/k∧/	/p∧/	/t∧/	/k∧/	/p∧/	/t∧/	/k∧/
1	6.87	6.22	5.98	7.20	6.75	6.41	6.76	6.41	6.03
2	5.56	5.09	4.81	5.49	4.88	4.73	5.68	4.80	4.59
3	7.43	7.49	6.79	7.24	7.31	6.64	7.36	7.21	6.56
4	5.78	5.28	4.59	6.01	5.47	5.17	5.97	5.70	5.27
5	6.01	6.20	6.54	7.10	6.54	6.68	6.56	6.36	6.25
6	5.98	5.79	5.70	6.00	5.65	5.71	6.00	5.93	5.69
7	5.81	5.30	4.76	5.74	5.55	4.99	5.61	5.36	4.84
8	7.56	6.37	6.38	7.33	6.28	6.17	7.48	6.13	6.27
9	5.55	5.38	5.28	6.00	5.78	5.53	5.90	5.56	5.58
10	6.60	6.53	5.82	6.63	6.52	5.93	6.52	6.35	5.93
11	5.76	5.31	4.68	5.13	4.80	4.41	5.35	4.80	4.42
12	7.23	7.51	6.53	7.22	7.31	6.08	7.33	7.59	6.34
13	5.77	5.16	4.96	5.88	5.33	5.11	6.04	5.34	5.11
14	6.05	6.81	6.17	6.00	6.67	6.16	6.37	6.90	6.31
15	6.32	5.74	5.27	6.34	5.88	5.56	6.16	5.54	5.23
16	6.63	6.25	6.06	6.76	6.77	6.35	6.57	6.24	5.95
17	6.13	5.77	4.71	6.26	5.48	4.86	6.02	5.66	4.90
18	6.47	6.02	5.21	6.52	6.00	5.20	6.39	6.18	5.11
19	6.53	5.87	5.52	6.18	5.68	5.54	6.42	5.97	5.72
20	6.44	5.56	5.16	6.53	5.36	5.02	6.36	5.29	5.21
21	5.99	6.59	5.49	6.67	6.74	5.20	6.75	6.82	5.43
22	6.75	7.05	6.22	6.17	6.53	6.30	6.34	6.80	6.58
23	6.58	5.95	5.58	6.38	5.75	5.44	6.67	6.01	5.69
24	5.79	5.28	4.63	5.86	5.28	5.09	5.79	5.24	4.72
25	5.17	4.82	4.58	6.51	5.72	4.77	5.32	5.64	5.00
26	7.00	6.68	6.13	6.82	6.49	5.95	6.68	6.27	7.84
27	6.64	6.50	5.78	6.69	6.16	5.64	6.81	6.49	5.78
28	5.80	5.56	5.20	5.68	5.35	5.19	5.81	5.33	4.94
29	6.51	5.96	5.53	6.61	6.09	5.65	6.25	5.55	5.44
30	6.39	6.02	5.78	6.61	6.30	5.91	6.16	5.98	5.63

Note. This table describes the diadochokinetic rate (syllables/second) data collected from control participants across all 3 visits.

Table 7: Description of Voice Quality for Control Participants Across All Three Visits

Participant ID	Visit 1			Visit 2			Visit 3		
	Jitter	Shimmer	Signal-to-noise	Jitter	Shimmer	Signal-to-noise	Jitter	Shimmer	Signal-to-noise
1	0.46	3.13	27.30	0.41	3.28	27.33	0.31	1.81	30.12
2	0.29	1.68	29.41	0.34	1.55	26.45	0.29	1.55	28.78
3	0.22	1.50	28.88	0.17	1.22	31.28	0.21	1.58	28.79
4	1.96	5.15	19.14	1.70	5.13	20.15	0.78	6.08	20.79
5	0.19	1.06	30.64	0.33	1.87	26.33	0.35	3.13	25.24
6	1.80	3.21	24.64	1.66	2.69	21.17	1.08	2.68	20.24
7	0.40	1.18	24.16	0.35	1.90	25.71	0.25	0.78	27.78
8	0.25	1.74	32.61	0.24	2.21	31.19	0.18	1.43	35.16
9	0.44	2.33	26.28	0.61	2.13	26.41	0.35	1.50	28.11
10	0.38	2.04	25.91	0.35	2.31	24.43	0.44	2.20	24.13
11	0.25	1.23	28.04	0.23	1.23	27.61	0.29	1.28	28.12
12	0.45	2.43	23.99	0.41	1.97	24.76	0.28	2.10	26.55
13	0.43	2.10	26.75	0.34	1.65	28.72	0.41	2.61	25.72
14	0.33	1.54	24.87	0.39	2.03	23.45	0.38	2.29	22.15
15	0.25	1.17	26.49	0.17	0.92	30.44	0.21	1.09	26.66
16	0.36	2.38	25.31	0.32	1.42	27.60	0.37	1.99	26.82
17	0.40	3.05	20.82	0.52	3.29	19.62	0.39	3.17	20.77
18	0.54	3.72	22.46	0.49	3.00	22.66	0.45	4.74	23.03
19	0.29	2.42	25.74	0.40	1.89	25.83	0.23	1.57	26.10
20	0.42	2.10	26.75	0.48	2.36	24.63	0.28	1.66	28.16
21	0.48	2.49	22.55	0.34	1.87	24.35	0.50	2.38	22.44
22	0.34	2.57	25.14	0.49	2.01	24.58	0.53	1.48	24.64
23	0.41	1.77	26.05	0.33	1.62	26.58	0.24	1.51	27.99
24	1.67	3.82	16.45	2.03	3.97	16.47	1.06	2.93	19.28
25	0.15	1.03	30.88	0.14	2.12	30.91	0.27	4.57	24.34
26	0.16	1.05	31.75	0.26	1.19	27.69	0.25	1.47	28.77
27	1.57	3.39	21.42	0.58	2.14	24.51	0.44	1.97	22.85
28	0.43	1.81	24.36	0.56	2.14	22.20	0.80	2.53	22.82
29	1.56	3.54	21.72	2.09	4.92	22.47	0.60	3.52	24.95
30	0.51	1.57	23.61	0.45	1.99	22.47	0.85	1.97	23.94

Note. This table describes the voice quality data collected from control participants across all 3 visits.

2.3.2 Visit 2

The motor symptoms associated with PD were assessed in participants with PD via the UPDRS. Participants with PD also indicated their self-perceived level of medication effectiveness via the MES. All participants complete five questionnaires (TLS, VAPP, CPIB, LSUS, and CES). Proxy measures for the participants with PD for each

questionnaire were collected from their primary communication partners. The presentation order of the scales was randomized across participants and visits. The presentation order was determined using www.randomizer.org.

Participants with PD and control participants then completed various speech tasks. Participants completed maximum sustained vowel phonation and diadochokinetic tasks. Table 4 provides diadochokinetic task rates for participants with PD and Table 5 provides description of voice report summaries for participants with PD. Table 6 provides diadochokinetic task rates for control participants and Table 7 provides a description of voice report summaries for control participants. Various speech intensity measures were then collected. The sentence "she saw Patty buy two poppies" was used during a magnitude production task, where participants repeated the sentence at their habitual speech intensity, two times louder than their habitual speech intensity, four times louder than their habitual speech intensity, and their maximal speech intensity. The magnitude production function was determined by calculating the slope of speech intensity when participants repeated the target sentence at their habitual speech intensity, two time louder than their habitual speech intensity, and four times louder than their habitual speech intensity. Participants also produced this sentence using their habitual speech intensity in 60, 65, 70, and 75 dB SPL of multi-talker background noise. The Lombard response function was then determined by calculating the slope of speech intensity when participants repeated the target sentence in 60, 65, 70, and 75 dB SPL of multi-talker background noise. Participants then completed the SIT and engaged in two minutes of conversation with the experimenter.

2.3.3 Visit 3

All participants completed five questionnaires (TLS, VAPP, CPIB, LSUS, and CES). Proxy measures for the participants with PD for each questionnaire were collected from their primary communication partners. The presentation order of the scales was randomized across participants and visits. The presentation order was determined using www.randomizer.org.

Participants with PD and control participants then completed various speech tasks. Participants completed maximum sustained vowel phonation and diadochokinetic tasks. Table 4 provides diadochokinetic task rates for participants with PD and Table 5 provides description of voice report summaries for participants with PD. Table 6 provides diadochokinetic task rates for control participants and Table 7 provides a description of voice report summaries for control participants. Various speech intensity measures were then collected. The sentence "she saw Patty buy two poppies" was used during a magnitude production task, where participants repeated the sentence at their habitual speech intensity, two times louder than their habitual speech intensity, four times louder than their habitual speech intensity, and their maximal speech intensity. The magnitude production function was determined by calculating the slope of speech intensity when participants repeated the target sentence at their habitual speech intensity, two time louder than their habitual speech intensity, and four times louder than their habitual speech intensity. Participants also produced this sentence using their habitual speech intensity in 60, 65, 70, and 75 dB SPL of multi-talker background noise. The Lombard response function was then determined by calculating the slope of speech intensity when participants repeated the target sentence in 60, 65, 70, and 75 dB SPL of multi-talker background noise. Finally, participants completed the SIT and engaged in two minutes of conversation with the experimenter.

2.3.4 Listener perceptual evaluation

All listeners read a detailed letter of information and provided their written consent prior to beginning the study. The Letter of Information presented to listeners is found in Appendix N. The Consent Form signed by listeners is provided in Appendix O. Listeners took part in three sessions, each lasting two to three hours in duration. All listening sessions took place in a quiet laboratory environment. Auditory stimuli were presented to listeners via free-field presentations with M-Audio speakers (AV 40) placed 60 cm away from listeners. During these sessions, listeners evaluated the sentence and conversational speech intelligibility of IWPD and control participants. Listeners were blinded to group and visit throughout their evaluation of speech intelligibility.

2.3.4.1 Sentence Intelligibility Test

Only 13-, 14-, and 15-word sentences of the SIT were presented to and orthographically transcribed by listeners using a custom Praat script on a 2016 13-inch MacBook Pro laptop. Each sentence was presented to listeners twice. Sentence intelligibility was determined by calculating the percentage of words correctly transcribed.

Following the presentation and transcription of each set of 13-, 14-, and 15-word sentences of the SIT, listeners provided an overall rating of speech intelligibility for the speaker. Speech intelligibility measures for the presented SIT sentences were obtained using VAS estimation. A VAS with the anchors "0% intelligible" and "100% intelligible" was presented to listeners using a custom Praat script on a 2016 13-inch MacBook Pro laptop following each set of 13-, 14-, and 15-word sentences of the SIT. Listeners rated speech intelligibility by indicating their rating with their cursor on the VAS. See Appendix P for an example of the VAS used for sentence intelligibility. Speech intelligibility was then automatically translated to a percentage for each rating. Stimuli presentation was randomized for across participants and visits using a custom Praat script.

2.3.4.2 Conversational intelligibility

Conversational intelligibility was evaluated using VAS estimation. From the two minutes of conversation recorded, each participant's recorded utterances were edited into a single spontaneous conversational excerpt ranging in length from 15 – 20 seconds in durations. Each conversational speech intelligibility sample was presented once to listeners prior to evaluating speaker intelligibility. A VAS with the anchors "0% intelligible" and "100% intelligible" was presented to listeners on a 2016 13-inch MacBook Pro laptop following each conversational speech intelligibility sample. Listeners rated speech intelligibility by indicating their rating with their cursor on the VAS. See Appendix P for an example of the VAS used for conversational speech intelligibility. Speech intelligibility was converted to a percentage for each rating. Stimuli presentation was randomized for across participants and visits using a custom Praat script.

2.4 Data Analysis

Eight objectives were established in this study. The first objective addressed the variability of speech intensity measures of the participants with PD and control participants. The second objective focused on the perceptions of typical speech loudness between the participants with PD, their primary communication partner, and control participants. The third objective addressed the variability of sentence and conversational speech intelligibility of the participants with PD and control participants. The fourth and fifth objectives evaluated the variability of self-rated communicative participation of the participants with PD and control participants, and the participants with PD and their primary communication partner respectively.

The next three objectives focused on relationships across speech intensity, speech intelligibility, communicative participation, demographic factors, and non-speech factors. The first of these objectives focused on the relationships among these variables for the participants with PD. The second of these objectives focused on the relationships among proxy-rated variables for participants with PD. The third of these objectives focused on the relationships between these variables for control participants. The statistical procedures are outlined below.

2.4.1 Statistical analysis for objective 1: Speech intensity

This analysis examined the temporal variability of speech intensity measures of participants with PD and control participants while repeating the phrase “She saw Pattie buy two poppies”. The measures of speech intensity under examination in this objective were habitual speech intensity, maximal speech intensity, Lombard response, and magnitude productions. For each of the four dependent measures related to speech intensity, a separate two-factor repeated measures analysis of variance (ANOVA) were performed to evaluate differences in speech intensity between participants with PD and control participants over time. The following factors were used in this analysis: one between-group independent factor with two levels [PD, control], one within-group independent factor with three levels [visit 1, visit 2, visit 3].

A secondary retest analysis exploring the retest reliability and repeatability of each dependent measure of speech intensity was performed for both participants with PD and

control participants. This retest analysis of the intensity measures used 1) correlations between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3.

ICC were used to assess the correlation between multiple sets of measurements and incorporates the consistency of within-subject measures and the average change of a group mean over time (Vaz et al., 2013). The mean difference is a measure of the difference between the mean values of two groups of a single measurement. The mean difference was calculated by subtracting mean values of one visit from the mean values of another visit. The mean difference was evaluated statistically using a paired t-test. The SEM is an estimate of how much variation (error) there would be across repeated measures of a single participant's values or scores. The SEM is the within-subject standard deviation of a measurement and is a measure of the degree to which scores are spread around a true score. The SEM reflects the precision of a measure. Thus, a lower SEM reflects a more precise measure. The SEM was calculated by multiplying the pooled standard deviation, obtained for the pair of retest visits, by the square root of one minus the ICC. The CR was calculated by multiplying the SEM by 2.77. CR is calculated in the same units as the dependent measure under study, as a result, it is a useful parameter for estimating the probable limits of measurement error when interpreting the effects of clinical treatments (Vaz, et al., 2013). The CR is the value below which the absolute differences between two measurements would lie with 0.95 probability (Vaz, et al., 2013). In other words, difference values above the CR would reflect 95% probability of a true level of change as opposed to a difference that may simply be due to measurement error. It is also suggested that CR reflects the minimal detectable true level of change of an outcome measure. The CR% is also referred to as the smallest real difference percentage (SRD%). The CR% was calculated by dividing the CR by the mean of the dependent variable and multiplying the resulting value by 100. Previous studies have proposed that CR% values less than or equal to 10% are indicative of good repeatability (Lu, Chen, Huan, & Hsieh, 2007; Smidt, et al., 2002). The current study also proposes

CR% values between 11 – 20 % to be indicative of marginal repeatability, and CR% values greater than 20% to be indicative of unacceptable repeatability.

These analyses were used to address objective 1 and the associated research question: *Do speech intensity measures differ between and within participants with PD and control participants over time?* In addition, these analyses were used to address objective 6 and the associated research question: *Do speech intensity measures demonstrate good retest reliability and acceptable repeatability for participants with PD and control participants?*

2.4.2 Statistical analysis for objective 2: Typical speech loudness

This analysis on evaluating whether perceptions of typical speech loudness differ between participants with PD, their primary communication partner, and control participants using their responses on the Typical Loudness Scale. A multi-factor repeated measures ANOVA was performed to evaluate differences in perceptions of typical speech loudness among participants with PD, their primary communication partners (PCP) and control participants over time. The following factors were used in this analysis: one between-group independent factor with three levels [PD, communication partners, control], one within-group independent factor with three levels [visit 1, visit 2, visit 3], one dependent factor [perceptions of typical loudness].

A secondary retest analysis exploring the retest reliability and repeatability of perceived typical speech loudness was performed for participants with PD, their primary communication partners, and control participants. This retest analysis of perceived loudness measures used 1) correlations between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3.

These analyses were used to address objective 2 and the associated research question: *Do participants with PD perceive their typical speech loudness differently from their primary communicative partners and control participants over time?* In addition, these analyses

were used to address objective 6 and the associated research question: *Do ratings of perceived typical speech loudness demonstrate good retest reliability and acceptable repeatability for participants with PD, their primary communication partners, and control participants?*

2.4.3 Statistical analysis for objective 3: Speech intelligibility

This analysis explored the temporal variability of measures of speech intelligibility of participants with PD and control participants. The measures of speech intelligibility under examination in this objective was sentence intelligibility as measured via SIT transcription and VAS scores, and conversational speech intelligibility. For each of the three dependent measures related to speech intelligibility, a separate two-factor repeated-measures ANOVA was performed to compare the speech intelligibility of participants with PD and control participants over time. The following factors were used in this analysis: one between-group independent factor with two levels [PD, control], one within-group independent factors with 3 levels [visit 1, visit 2, visit 3].

A secondary retest analysis exploring the retest reliability and repeatability of each dependent measure of speech intelligibility was performed for both participants with PD and control participants. This retest analysis of speech intelligibility measures used 1) correlations between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3.

These analyses were used to address objective 3 and the associated research question: *Do speech intelligibility measures differ between and within participants with PD and control participants over time?* In addition, these analyses were used to address objective 6 and the associated research question: *Do speech intelligibility measures demonstrate good retest reliability and acceptable repeatability for participants with PD and control participants?*

2.4.4 Statistical analysis for objective 4: Self-rated communicative participation

This analysis explored the temporal variability of self-rated communicative participation of the participants with PD and control participants over time. The measures of communicative participation under examination in this objective was the individual CES question scores, the individual VAPP subsection scores, the standardized CPIB scores, and the LSUS scores. Note that the section exploring the effects of voice impairment on an individual's job on the VAPP was excluded from calculations as the majority of participants were retired at the time of this study and were unable to complete this section of the questionnaire. For the individual CES question scores and the individual VAPP subsection scores, a separate repeated measure multivariate analysis of variance (MANOVA) was performed to analyze communicative participation between participants with PD and control participants. For CPIB and LSUS scores, a separate two-factor repeated measures ANOVA was performed to analyze communicative participation between participants with PD and control participants. The following factors were used in each of these analyses was: one between-group independent factor with two levels [PD, control], one within-group independent variable with three levels [visit 1, visit 2, visit 3.

A secondary retest analysis exploring the retest reliability and repeatability of each dependent measure of communicative participation was performed for both participants with PD and control participants. This retest analysis of communicative participation measures used 1) correlations between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3.

These analyses were used to address objective 4 and the associated research question: *Do self-ratings of communicative participation differ between and within participants with PD and control participants over time?* In addition, these analyses were used to address objective 6 and the associated research question: *Do self-ratings of communicative participation demonstrate good retest reliability and acceptable repeatability for participants with PD and control participants?*

2.4.5 Statistical analysis for objective 5: Self- and proxy-rated communicative participation

This analysis explored differences in ratings of communicative participation between participants with PD and their primary communication partner. Self- and proxy-measures of communicative participation under examination in this objective was the individual CES question scores, the individual VAPP subsection scores, the standardized CPIB scores, and the LSUS scores. Note that the section exploring the effects of voice impairment on an individual's job on the VAPP was excluded from calculations as the majority of participants were retired at the time of this study and were unable to complete this section of the questionnaire. For the individual CES question scores and the individual VAPP subsection scores, a separate repeated measures MANOVA was performed to analyze communicative participation between participants with PD and their primary communication partners. For CPIB and LSUS scores, a separate two-factor repeated measures ANOVA was performed to analyze communicative participation between participants with PD and their primary communication partners. The following factors were used in these analyses: one between-group independent factor with two levels [PD, communication partners], one within-group independent variable with three levels [visit 1, visit 2, visit 3].

A secondary retest analysis exploring the retest reliability and repeatability of each dependent measure of communicative participation was performed for both participants with PD and their primary communication partners. This retest analysis of communicative participation measures used 1) correlations between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3.

These analyses were used to address objective 5 and the associated research question: *Do ratings of communicative participation differ between and within participants with PD and their primary communication partners over time?* In addition, these analyses were used to address objective 6 and the associated research question #6: *Do ratings of*

communicative participation demonstrate good retest reliability and acceptable repeatability for participants with PD and their primary communication partners?

2.4.6 Statistical analysis for objective 7: Inter-relationships among variables in participants with Parkinson's disease

This exploratory analysis evaluated inter-relationships among in speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD across each of the three experimental visits. A matrix of inter-correlations was obtained via three series of Pearson correlation procedures ($p<.05$) applied to all possible pairwise combination of the experimental variables, demographic factors, and non-speech factors among participants with PD. The criteria for interpreting the strength of the correlations was as follows: 0 – 0.25 = little to no correlation; 0.25 – 0.50 = a fair correlation; 0.50 – 0.75 = a good to moderate correlation; 0.75 and above = a good to excellent correlation (Portney & Watkins, 2000). The resulting correlation matrix was used to examine potential relationships among the various experimental variables for each of the three visits.

These analyses were used to address objective 7 and the associated research question: *Are measures of speech intensity, speech intelligibility, communicative participation, demographic factors, and non-speech factors related to one another in participants with PD?*

2.4.7 Statistical analysis for objective 8: Inter-relationships among proxy-measures in participants with Parkinson's disease

This exploratory analysis evaluated inter-relationships among in speech intensity measures, speech intelligibility measures, proxy-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD across each of the three experimental visits. A matrix of inter-correlations was obtained via a series of Pearson correlation procedures ($p<.05$) applied to all possible pairwise combination of the experimental variables, demographic factors, and non-speech factors

among participants with PD. The criteria for interpreting the strength of the correlations was as follows: 0 – 0.25 = little to no correlation; 0.25 – 0.50 = a fair correlation; 0.50 – 0.75 = a good to moderate correlation; 0.75 and above = a good to excellent correlation (Portney & Watkins, 2000). The resulting correlation matrix was used to examine potential relationships among the various experimental variables for each of the three visits.

These analyses were used to address objective 8 and the associated research question: *Are measures of speech intensity, speech intelligibility, proxy ratings of communicative participation, demographic factors, and non-speech factors related to one another in participants with PD?*

2.4.8 Statistical analysis for objective 9: Inter-relationships among variables in control participants

This exploratory analysis evaluated inter-relationships among in speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for control participants across each of the three experimental visits. A matrix of inter-correlations was obtained via a series of Pearson correlation procedures ($p<.05$) applied to all possible pairwise combination of the experimental variables, demographic factors, and non-speech factors among participants with PD and control participants. The criteria for interpreting the strength of the correlations was as follows: 0 – 0.25 = little to no correlation; 0.25 – 0.50 = a fair correlation; 0.50 – 0.75 = a good to moderate correlation; 0.75 and above = a good to excellent correlation (Portney & Watkins, 2000). The resulting correlation matrix was used to examine potential relationships among the various experimental variables.

These analyses were used to address objective 9 and the associated research question: *Are measures of speech intensity, speech intelligibility, communicative participation, demographic factors, and non-speech factors related to one another in control participants?*

Chapter 3

3 Results

This study examined the temporal variability of acoustic and perceptual speech measures, and psychosocial measures in individuals with hypophonia and PD. Additionally, this study explored the relationships among speech intensity measures, speech intelligibility measures, self- and proxy-rated communicative participation measures, demographic factors, and non-speech factors. In order to provide a comprehensive representation of the different variables of interest and their relationship to one another, each of the nine objectives in this study were analyzed separately, with the exception of objective 6 which is embedded within objectives 1-5. The results for objectives 1 through 5 consist of analyses that will evaluate variability over time. Objective 6 describes the retest analyses of each variable and, therefore, objective 6 is embedded with the first 5 objectives. The results for objectives 7 through 9 consist of analyses that will evaluate relationships existing among variables.

3.1 Statistical Power

Statistical power reflects the prospect of identifying differences resulting from a treatment and probability of the successful replication of a study (Keppel, 1991). Statistical power is established based on the interaction and relationship between sample size, variance within data, effect size, and statistical significance (Portney & Watkins, 2000). G*Power v3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) was used to perform a post-hoc power analysis. The power calculations were based on the findings of previous studies exploring speech intensity (Adams et al., 2006b; Dykstra et al., 2012a) and communicative participation (Dykstra et al., 2015). The Adams and colleagues (2006b) study was used to estimate differences between mean habitual conversational speech intensity for participants with PD and control participants to be approximately 4 dB with a standard deviation of approximately 3 dB. The resulting effect size was 1.66. A post-hoc power analysis using a sample size of 53 (PD = 23; controls = 30), an effect size of 1.33, and an alpha level of .05. This post-hoc analysis provided a power estimate of .99.

The Dykstra and colleagues (2012a) study was used to estimate differences between mean habitual conversational speech intensity for participants with PD and control participants to be approximately 5 dB with a standard deviation of approximately 2.99 dB. The resulting effect size was 1.66. A post-hoc power analysis using a sample size of 53 (PD = 23; controls = 30), an effect size of 1.66, and an alpha level of .05. This post-hoc analysis provided a power estimate of .99. The work of Dykstra and colleagues (2012a) was used to estimate differences between maximum speech intensity for participants with PD and control participants to be approximately 10.50 dB with a standard deviation of approximately 4.25 dB. The resulting effect size was 2.47. A post-hoc power analysis using a sample size of 53 (PD = 23; controls = 30), an effect size of 2.47, and an alpha level of .05. This post-hoc analysis provided a power estimate of 1.00. The Dykstra and colleagues (2012a) study was also used to estimate differences between speech intensity during 65 dB of background noise for participants with PD and control participants to be approximately 5.60 dB with a standard deviation of approximately 2.89 dB. The resulting effect size was 1.93. A post-hoc power analysis using a sample size of 53 (PD = 23; controls = 30), an effect size of 1.93, and an alpha level of .05. This post-hoc analysis provided a power estimate of .99. A study by Dykstra and colleagues (2015) study was used to estimate differences between ratings of perceived communication effectiveness for participants with PD and control participants to be approximately 2.0 with a standard deviation of approximately 1.07. The resulting effect size was 1.86. A post-hoc power analysis using a sample size of 53 (PD = 23; controls = 30), an effect size of 1.66, and an alpha level of .05. This post-hoc analysis provided a power estimate of .99. Based on these five estimates of power, it appears that the current study demonstrates of power estimate of .99. These results suggest that statistical power is satisfactory for the present study.

3.2 Reliability

Inter-rater reliability was calculated for all three measures of speech intelligibility (SIT transcription scores, SIT VAS scores, and conversational intelligibility VAS scores). The ICC values related to the measures of speech intelligibility was found to be the following: SIT transcription scores of $ICC = .85$, $p < .001$ for visit 1, $ICC = .80$, $p < .001$ for visit 2,

and *ICC* = .74, $p < .001$ for visit 3; SIT VAS scores of *ICC* = .84, $p < .001$ for visit 1, *ICC* = .82, $p < .001$ for visit 2, and *ICC* = .84, $p < .001$ for visit 3; and conversational intelligibility VAS scores of *ICC* = .72, $p < .001$ for visit 1, *ICC* = .63, $p < .001$ for visit 2, and *ICC* = .78, $p < .001$ for visit 3. See Table 8 for a complete description of inter-rater reliability. These results suggest good reliability between listeners for all three measures of speech intelligibility.

Table 8: Inter-Rater Reliability for Listeners Across Measures of Speech Intelligibility

	SIT transcription scores		SIT VAS scores		Conversational intelligibility VAS scores	
	ICC[95% CI]	*p* value	ICC[95% CI]	*p* value	ICC[95% CI]	*p* value
Visit 1	. 85[.76-.91]	<.001	.84[.69-.91]	<.001	.72[.52-.84]	<.001
Visit 2	.80[.69-.88]	<.001	.82[.71-.89]	<.001	.63[.38-.78]	<.001
Visit 3	.74[.59-.85]	<.001	.84[.69-.91]	<.001	.78[.53-.89]	<.001

Note. This table illustrates the interclass correlations, 95% confidence intervals, and statistical significance as a measure of inter-rater reliability for listener ratings for the different speech intelligibility measures.

Scores generated by each listener for each of the three speech intelligibility tasks were measured against each other in order to ascertain each listener's intra-rater reliability. In order to determine intra-rater reliability, each listener re-measured 10% of the intelligibility data. The Cronbach's alphas related to the measures of speech intelligibility ranged from .58 - .98 for listener 1, .56 - .97 for listener 2, and .19 - .95 for listener 3. See Table 9 for a complete description of intra-rater reliability. These results suggest moderate to excellent reliability for listeners on eight of nine measures of speech intelligibility.

Table 9: Intra-Rater Reliability for Listeners Across Measures of Speech Intelligibility

	SIT transcription scores Cronbach's alpha	SIT VAS scores Cronbach's alpha	Conversational intelligibility VAS scores Cronbach's alpha
Listener 1	.98	.95	.58
Listener 2	.56	.91	.97
Listener 3	.19	.78	.95

Note. This table illustrates Cronbach's alpha as a measure of intra-rater reliability for each listener ratings for the different speech intelligibility measures.

3.3 Statistical Analysis for Objective 1: Speech Intensity

In order to answer the question '*Do speech intensity measures differ over time between and within participants with PD and control participants?',* a separate two-factor repeated measures ANOVA was performed on each of the four dependent speech intensity measures (i.e., habitual speech intensity, maximum speech intensity, Lombard response function, magnitude production). Each of the two-factor repeated measures ANOVAs involved a "Group" factor with two separate levels (participants with PD and control participants) and a "Visit" factor with three levels (visit 1, visit 2, and visit 3). Whenever the ANOVA resulted in a significant main effect for the factor "Visit", a post-hoc analysis was performed using pairwise comparisons.

3.3.1 Habitual speech intensity

Measures of habitual speech intensity were obtained from audio recordings of participants with PD and control participants reading the sentence "She saw Patty buy two poppies" using their habitual speaking loudness. Descriptive statistics for habitual speech intensity are shown in Table 10. The results of the two-way repeated measures ANOVA for the dependent measure of habitual speech intensity showed that there was no significant main effect of "Group" $F(1,49) = 1.60$, $p = .211$ with participants with PD having a similar marginal mean ($M = 67.71$, $SD = 3.61$) to that of control participants ($M = 69.01$, $SD = 3.61$). In contrast, there was a significant main effect of "Visit" on habitual speech intensity $F(2,98) = 5.54$, $p = .005$. A closer look at the factor "Visit", using pairwise post-hoc analyses indicated that the marginal mean habitual speech intensity at visit 1 ($M = 67.35$, $SD = 4.28$) was significantly lower than the mean habitual speech intensity at visit 2 ($M = 69.24$, $SD = 4.57$) and visit 3 ($M = 68.50$, $SD = 4.21$). It is important to note that this main effect of "Visit" needs to be qualified because of the finding of a significant "Group" by "Visit" interaction $F(2,98) = 3.50$, $p = .034$ for habitual speech intensity. This significant interaction is illustrated in Figure 1. It appears that at visit 1 the group difference between the PD and control participants' habitual speech intensity is greater than the group differences that were found at visit 2 and visit 3.

Results of a post-hoc analyses involving comparisons of these group differences at each visit provided additional information about this significant "Group" by "Visit" interaction. For the post-hoc comparison related to visit 1, the participants with PD had a mean habitual speech intensity (M = 66.09, SD = 4.79) that was significantly lower (-3.06 dB SPL) than the mean habitual speech intensity of control participants (M = 69.15, SD = 4.12), $t(2,51) = -2.50$, $p = .016$. In contrast, the post-hoc comparisons related to group differences were not significant at visit 2 $t(2,50) = -0.27$, $p = .786$ or visit 3 $t(2,50) = -0.69$, $p = .494$. Thus, it appears that although the habitual speech intensity of the IWPD was significantly lower than that of the controls at visit 1, it increased to a level that was not significantly different from the controls at visit 2 and visit 3.

Table 10: Descriptive Statistics of Speech Intensity Measures for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	Habitual speech intensity mean(*SD*)	Maximum speech intensity mean(*SD*)	Lombard response function mean(*SD*)	Magnitude production mean(*SD*)
Participants with PD				
Visit 1	66.09*(4.79)*	77.55*(6.61)*	0.34*(0.17)*	2.34*(1.50)*
Visit 2	69.09*(4.39)*	79.32*(6.50)*	0.31*(0.12)*	2.01*(1.06)*
Visit 3	68.27*(4.02)*	78.68*(6.11)*	0.28*(0.14)*	2.02*(0.97)*
Control participants				
Visit 1	69.15*(4.12)*	83.75*(6.36)*	0.32*(0.13)*	3.13*(1.31)*
Visit 2	69.43*(4.57)*	82.40*(5.75)*	0.26*(0.16)*	2.60*(1.09)*
Visit 3	69.13*(4.70)*	83.35*(7.08)*	0.25*(0.17)*	2.81*(1.16)*
Marginal means				
Participants with PD	67.71*(3.61)*	78.20*(5.68)*	0.31*(0.14)*	2.09*(1.08)*
Control participants	69.01*(3.61)*	82.88*(5.65)*	0.27*(0.11)*	2.86*(1.08)*
Visit 1	67.35*(4.28)*	80.22*(6.76)*	0.32*(0.14)*	2.75*(1.43)*
Visit 2	69.24*(4.57)*	80.63*(6.45)*	0.29*(0.14)*	2.24*(1.00)*
Visit 3	68.50*(4.21)*	80.77*(6.99)*	*0.26(0.14)*	2.42*(1.14)*

Note. This table illustrates the means and standard deviations for the different speech intensity measures for participants with PD and control participants across visits.

Figure 1: Means of Habitual Speech Intensity for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

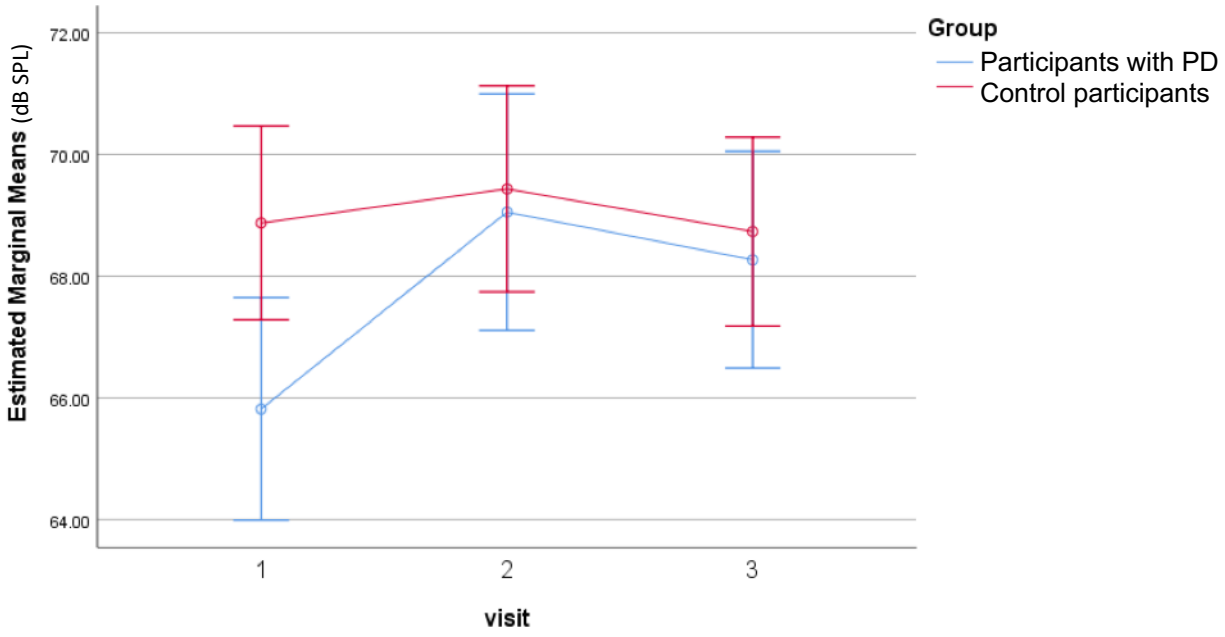

Figure 1. This figure demonstrates the changes in habitual speech intensity scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.3.1.1 Retest analyses

Following the primary variability analysis described above, secondary retest analyses related to the evaluation of the retest reliability and repeatability of habitual speech intensity were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. The purpose of these analyses was to 1) Evaluate the strength of the retest reliability values using the previously recommended correlation of *ICC* > .75. This ICC value was used as a point of reference since an ICC of .75 or higher is indicative of good to excellent reliability (Koo & Li, 2016). 2) Provide an estimate of the smallest value that would be considered to represent the measurement error. 3) Indicate the difference between the mean values of two groups of a single measurement. 4) Provide an estimate

of a significant change between two or more time points of a measure. The CR is also referred to as the Smallest Real Difference (SRD; Beckerman, Roebroeck, Lankhorst, Becher, Bezemer, & Verbeek, 2001) and the Minimal Detectable Difference (MDD; Furlan & Sterr, 2018; Steffen & Seney, 2008). 5) Evaluating the percentage of variation present in a measure. Based on previous studies, CR% values less than or equal to 10% are used to indicate good repeatability in the present study (Lu et al., 2007; Smidt, et al., 2002). The current study also uses CR% values between 11 – 20 % to indicate marginal repeatability, and CR% values greater than 20% to indicate unacceptable repeatability. The retest analysis, involving the ICC, mean difference t-test, SEM, CR and CR%, were performed on the participants with PD and control participants. Results for the retest analysis of habitual speech intensity are summarized in Table 11.

Table 11: Retest Analyses of Habitual Speech Intensity for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.64[.10-.85]	-3.00	4.30	-3.35	.003	4.59	2.76	7.64	11.55
Visits 1 – 3	.64[.15-.85]	-2.45	4.22	-2.73	.013	4.42	2.65	7.35	10.64
Visits 2 – 3	.79[.50-.91]	0.78	3.55	1.03	.313	4.21	1.93	5.34	7.83
Control participants									
Visits 1 – 2	.77[.52-.89]	-0.56	3.67	-0.83	.416	4.35	2.09	5.78	8.36
Visits 1 – 3	.73[.43-.87]	0.02	4.12	0.03	.978	4.42	2.30	6.36	9.16
Visits 2 – 3	.66[.28-.84]	0.70	4.45	0.85	.402	4.64	2.70	7.49	10.83

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of habitual speech intensity for participants with PD and control participants.

3.3.1.1.1 Retest analysis of habitual speech intensity for participants with Parkinson's disease

The ICC values related to the repeated measurement of habitual speech intensity across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .64$, $p = .002$, visit 1 versus visit 3 $ICC = .64$, $p = .004$, and visit 2 versus visit 3 $ICC = .79$, $p < .001$. These results suggest that habitual speech intensity as measured in the present study

did not demonstrate good retest reliability in IWPD because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for habitual speech intensity was -3.00 dB SPL (*SD* = 4.30) and this difference was significant $t(22) = -3.35$, $p = .003$. The mean difference between visit 1 and visit 3 for habitual speech intensity was -2.45 dB SPL (*SD* = 4.22) and this difference was significant $t(21) = -2.73$, $p = .013$. The mean difference between visit 2 and visit 3 for habitual speech intensity was 0.78 dB SPL (*SD* = 3.55) and this difference was not significant $t(21) = 1.034$, $p = .313$. These mean differences in retest values ranged from -6.57 – 9.36 dB SPL. Additionally, the following CR values and CR percentages were obtained for the following visit comparisons: for visit 1 vs visit 2 *CR* = 7.64 and *CR%* = 11.55%, for visit 1 vs visit 3 *CR* = 7.35 and *CR%* = 10.64%, and for visit 2 vs visit 3, *CR* = 5.34 and *CR%* = 7.83%. Based on the CR, an observed change in the habitual speech intensity of at least 5.34– 7.64 dB SPL would suggest an acceptable amount of measurement error ranging from 7.83 – 11.55% variation in the habitual speech intensity of IWPD. These results suggest that the measure of habitual speech intensity demonstrates fairly good repeatability for IWPD.

3.3.1.1.2 Retest analysis of habitual speech intensity for control participants

The ICC values related to the repeated measurement of habitual speech intensity across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .77, $p < .001$, visit 1 versus visit 3 *ICC* = .73, $p < .001$, and visit 2 versus visit 3 *ICC* = .66, $p = .003$. These results suggest that habitual speech intensity as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for habitual speech intensity was -0.56 dB SPL (*SD* = 3.67) and this difference was not significant $t(28) = -0.83$, $p = .416$. The mean differences in retest values ranged from -6.80 – 6.95 dB SPL. The mean difference between visit 1 and visit 3 for habitual speech intensity was 0.02 dB SPL (*SD* = 4.12) and this difference was not significant $t(29) = 0.03$, $p = .978$. The mean differences in retest

values ranged from -6.08 – 10.05 dB SPL. The mean difference between visit 2 and visit 3 for habitual speech intensity was 0.70 dB SPL (*SD* = 4.45) and this difference was not significant $t(28) = 0.85$, $p = .402$. The mean differences in retest values ranged from -6.79 – 14.91 dB SPL. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 5.78 and *CR%* = 8.36%, for visit 1 vs visit 3, *CR* = 6.36 and *CR%* = 9.16%, and for visit 2 vs visit 3, *CR* = 7.49 and *CR%* = 10.83%. Based on the CR, an observed change in the habitual speech intensity of at least 5.78 – 7.49 dB SPL would suggest an acceptable amount of measurement error ranging from 8.36 – 10.83% variation in the habitual speech intensity of control participants. These results suggest that the measure of habitual speech intensity demonstrates good repeatability for control participants.

3.3.2 Maximum speech intensity

Measures of maximum speech intensity were obtained from audio recordings of participants with PD and control participants reading the sentence "She saw Patty buy two poppies" using as loud a voice as possible. Descriptive statistics for maximum speech intensity are shown in Table 10. The results of the two-way repeated measures ANOVA for the dependent measure of maximum speech intensity showed that there was significant main effect of "Group" $F(1,49) = 8.49$, $p = .005$ with the participants with PD having a statistically significantly lower marginal mean (*M* = 78.20, *SD* = 5.68) to that of the control participants (*M* = 82.88, *SD* = 5.65). In contrast, there was no significant main effect of "Visit" on maximum speech intensity $F(2,98) = 0.42$, $p = .661$. A closer look at the factor "Visit", using pairwise post-hoc analyses indicated that the marginal mean of maximum speech intensity at visit 1 (*M* = 80.23, *SD* = 6.76) was similar to the marginal mean of maximum speech intensity at visit 2 (*M* = 80.63, *SD* = 6.45) and visit 3 (*M* = 80.77, *SD* = 6.99) It is important to note that this main effect of "Visit" needs to be qualified because of the finding of a "Group" by "Visit" interaction that trended towards significance $F(2,98) = 2.76$, $p = .069$ for maximum speech intensity. This significant interaction is illustrated in Figure 2. It appears that at visit 1 the group difference between the PD and control participants' maximum speech intensity is greater than the group differences that were found at visit 2 and visit 3. Results of a post-hoc analyses involving

comparisons of these group differences at each visit provided additional information about this significant "Group" by "Visit" interaction. For the post-hoc comparison related to visit 1, the participants with PD had a mean maximum speech intensity (*M* = 77.07, *SD* = 6.34) that was significantly lower (-6.31 dB) than the mean maximum speech intensity of control participants (*M* = 83.38, *SD* = 6.13), t(2,51) = -3.46, p=.001. For the post-hoc comparison related to visit 3, the participants with PD had a mean maximum speech intensity (*M* = 78.68, *SD* = 6.11) that was significantly lower (-4.67 dB) than the mean maximum speech intensity of control participants (*M* = 83.35, *SD* = 7.08), *t*(2,50) = -2.48, *p* = .016. In contrast, the post-hoc comparisons related to group differences was not significant at visit 3 *t*(2,50) = -1.81, *p* = .077. Thus, it appears that although the maximum speech intensity of the participants with PD was significantly lower than that of control participants at visit 1 and visit 3, it increased to a level that was not significantly different from control participants at visit 2.

Figure 2: Means of Maximum Speech Intensity for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

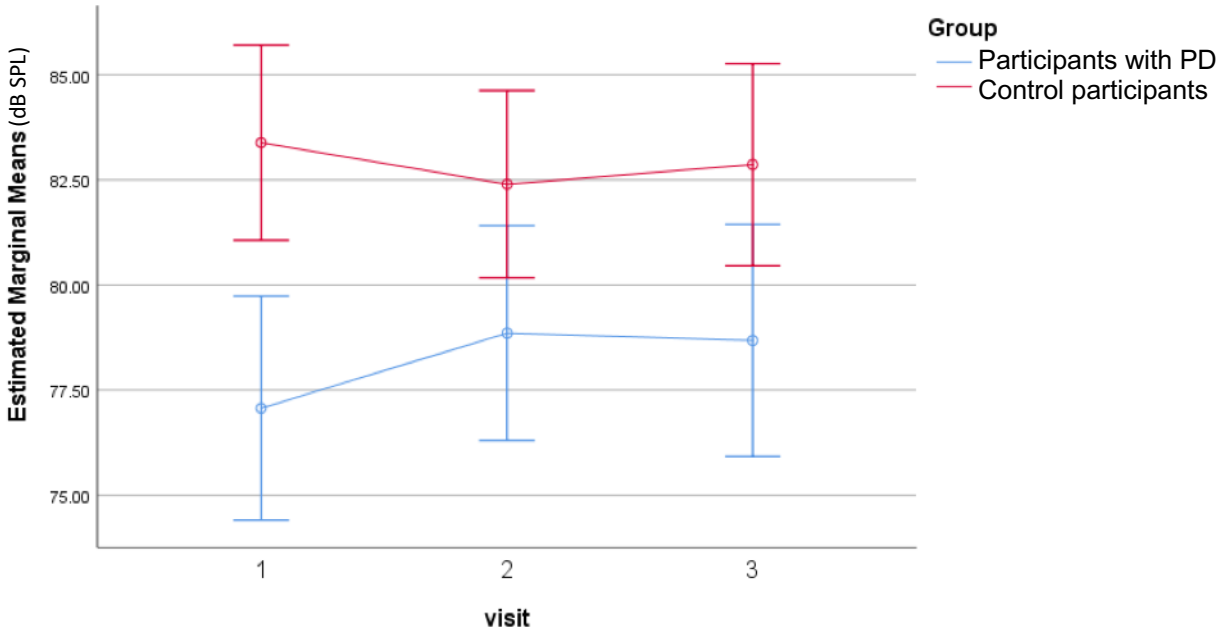

Figure 2. This figure demonstrates the changes in maximum speech intensity scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.3.2.1 Retest analyses

An analysis of the retest reliability and repeatability of maximum speech intensity were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants. Results for retest analyses of maximum speech intensity are summarized in Table 12.

Table 12: Retest Analyses of Maximum Speech Intensity for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.86[.67-.94]	-1.77	4.39	-1.94	.066	6.56	2.45	6.79	8.76
Visits 1 – 3	.80[.53-.92]	-1.61	4.96	-1.54	.142	6.36	2.85	7.88	9.94
Visits 2 – 3	.93[.81-.97]	0.17	3.37	0.24	.812	6.31	1.67	4.62	5.88
Control participants									
Visits 1 – 2	.87[.74-.94]	0.99	3.93	1.35	.187	6.06	2.19	6.06	7.23
Visits 1 – 3	.89[.76-.95]	0.41	4.33	0.51	.609	6.73	2.23	6.18	7.50
Visits 2 – 3	.82[.62-.92]	-0.46	4.88	-0.51	.612	6.45	2.74	7.58	9.09

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of maximum speech intensity for participants with PD and control participants.

3.3.2.1.1 Retest analysis of maximum speech intensity for participants with Parkinson's disease

The ICC values related to the repeated measurement of maximum speech intensity across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .86, $p < .001$, visit 1 versus visit 3 *ICC* = .80, $p < .001$, and visit 2 versus visit 3 *ICC* = .93, $p < .001$. These results suggest that maximum speech intensity as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for maximum speech intensity was -1.77 dB SPL (*SD* = 4.39) and this difference approached significance $t(22) = -1.94$, $p = .066$. The mean differences in retest values ranged from -11.17 – 5.21 dB SPL. The mean difference between visit 1 and visit 3 for maximum speech intensity was -1.61 dB SPL (*SD* = 4.96) and this difference was not significant $t(21) = -1.53$, $p = .142$. The mean differences in retest values ranged from -12.12 – 6.34 dB SPL. The mean difference between visit 2 and visit 3 for maximum speech intensity was 0.17 dB SPL (*SD* = 3.37) and this difference was not significant $t(21) = 0.24$, $p = .812$. The mean differences in retest values ranged from -6.34 – 6.94 dB SPL. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 6.79 and

CR% = 8.76%, for visit 1 vs visit 3, *CR* = 7.88 and *CR%* = 9.94%, and for visit 2 vs visit 3, *CR* = 4.62 and *CR%* = 5.88%. Based on the CR, an observed change in the maximum speech intensity of at least 4.62 – 7.88 dB SPL would suggest an acceptable amount of measurement error ranging from 5.88 – 9.94% variation in the maximum speech intensity of IWPD. These results suggest that the measure of maximum speech intensity demonstrates good repeatability for IWPD.

3.3.2.1.2 Retest analysis of maximum speech intensity for control participants

The ICC values related to the repeated measurement of maximum speech intensity across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .87, $p < .001$, visit 1 versus visit 3 *ICC* = .89, $p < .001$, and visit 2 versus visit 3 *ICC* = .82, $p < .001$. These results suggest that maximum speech intensity as measured in the present study demonstrated good retest reliability in healthy speakers because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for maximum speech intensity was 0.99 dB SPL (*SD* = 3.93) and this difference was not significant $t(28) = 1.35$, $p = .187$. The mean differences in retest values ranged from -6.35 – 8.96 dB SPL. The mean difference between visit 1 and visit 3 for maximum speech intensity was 0.41 dB SPL (*SD* = 4.33) and this difference was not significant $t(29) = 0.52$, $p = .609$. The mean differences in retest values ranged from -10.74 – 10.84 dB SPL. The mean difference between visit 2 and visit 3 for maximum speech intensity was -0.46 dB SPL (*SD* = 4.88) and this difference was not significant $t(28) = -0.51$, $p = .612$. The mean differences in retest values ranged from -14.13 – 11.78 dB SPL. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 6.06 and *CR%* = 7.23%, for visit 1 vs visit 3, *CR* = 6.18 and *CR%* = 7.50%, and for visit 2 vs visit 3, *CR* = 7.58 and *CR%* = 9.09%. Based on the CR, an observed change in the maximum speech intensity of at least 6.06 – 7.58 dB would suggest an acceptable amount of measurement error ranging from 7.23 – 9.09% variation in the maximum speech intensity of control participants. These results suggest that the measure of maximum speech intensity demonstrates good repeatability for IWPD.

3.3.3 Lombard response function

Lombard response function was determined from audio recordings of participants with PD and control participants reading the sentence “She saw Patty buy two poppies” in 60, 65, 70, and 75 dB SPL of multi-talker background noise. The Lombard response function was calculated by determining the slope of speech intensity across each of these four background noise conditions. Descriptive statistics for Lombard response function can be found in Table 10. The results of the two-way repeated measures ANOVA for the dependent measure of Lombard response function showed that there was no significant main effect of “Group” $F(1,49) = 1.13$, $p = .292$ with participants with PD having a similar marginal mean ($M = 0.31$, $SD = 0.14$) to that of the control participants ($M = 0.27$, $SD = 0.11$). In contrast, there was a significant main effect of “Visit” of the Lombard response function $F(2,98) = 3.95$, $p = .022$. A closer look at the factor “Visit”, using pairwise post-hoc analyses indicated that the marginal mean Lombard response function at visit 1 ($M = 0.32$, $SD = 0.14$) was significantly greater than the marginal mean of the Lombard response function at visit 2 ($M = 0.29$, $SD = 0.14$) and visit 3 ($M = 0.26$, $SD = 0.14$). Additionally, no significant “Group” by “Visit” interaction was found $F(2,98) = 0.40$, $p = .674$ for Lombard response function. This non-significant interaction is illustrated in Figure 3. Thus, it appears that although the Lombard response function of the IWPD was not significantly different from control participants, the Lombard response function at visit 1 was significantly different compared to the Lombard response function at visit 2 and visit 3.

Figure 3: Means of Lombard Response Function for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

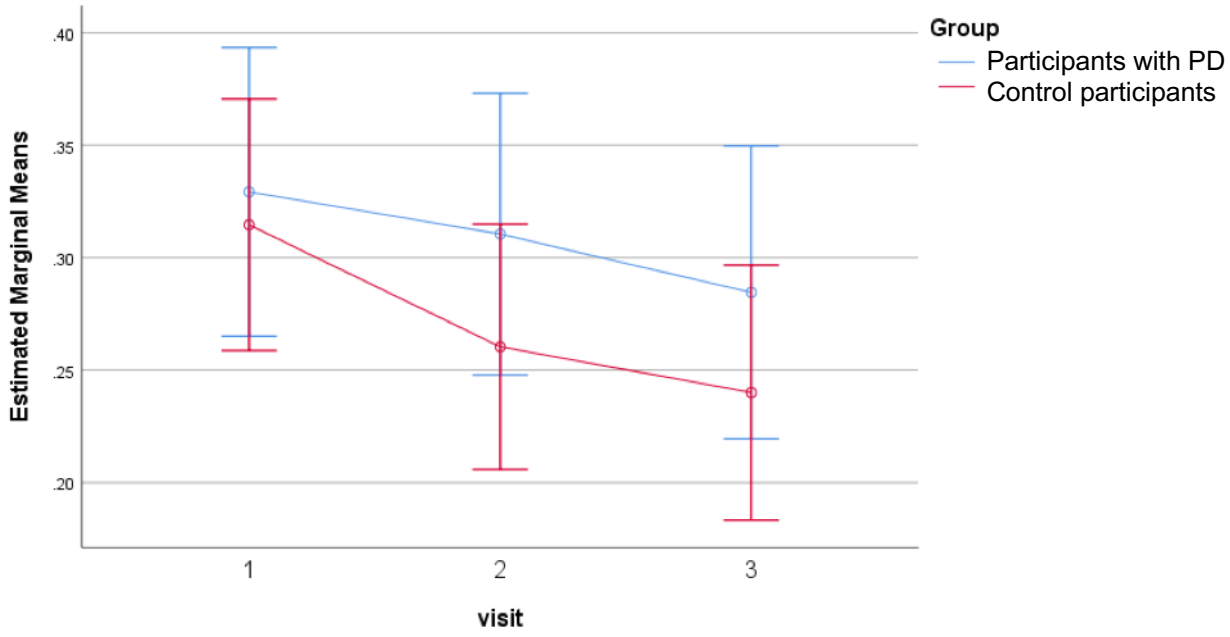

Figure 3. This figure demonstrates the changes in Lombard response function scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.3.3.1 Retest analyses

Following the primary variability analysis described above, secondary retest analyses related to the evaluation of the retest reliability and repeatability of Lombard response function were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants. Results for retest analyses of Lombard response function are summarized in Table 13.

Table 13: Retest Analyses of Lombard Response Function for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.15[-1.02-.64]	0.035	0.20	0.83	.415	0.15	0.14	0.38	110.52
Visits 1 – 3	.49[-.20-.79]	0.046	0.18	1.19	.246	0.16	0.11	0.31	99.37
Visits 2 – 3	.72[.32-.88]	0.027	0.12	1.02	.318	0.13	0.07	0.19	68.25
Control participants									
Visits 1 – 2	.73[.43-.88]	0.053	0.13	2.21	.036	0.15	0.08	0.21	65.57
Visits 1 – 3	.66[.30-.84]	0.066	0.15	2.44	.021	0.15	0.09	0.24	94.01
Visits 2 – 3	.77[.52-.89]	0.021	0.14	0.81	.422	0.17	0.08	0.22	87.72

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of Lombard response function for participants with PD and control participants.

3.3.3.1.1 Retest analysis of Lombard response function for participants with Parkinson's disease

The ICC values related to the repeated measurement of Lombard response function across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .15, *p* = .351, visit 1 versus visit 3 *ICC* = .49, *p* = .063, and visit 2 versus visit 3 *ICC* = .72, *p* = .003. These results suggest that Lombard response function as measured in the present study did not demonstrate good retest reliability in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for Lombard response function was 0.04 (*SD* = 0.20) and this difference was not significant *t*(22) = 0.83, *p* = .415 . The mean differences in retest values ranged from -0.33 – 0.40. The mean difference between visit 1 and visit 3 for Lombard response function was 0.05 (*SD* = 0.18) and this difference was not significant *t*(21) = 1.19, *p* = .246. The mean differences in retest values ranged from -0.35 – 0.35. The mean difference between visit 2 and visit 3 for Lombard response function was 0.03 (*SD* = 0.12) and this difference was not significant *t*(21) = 1.02, *p* = .318. The mean differences in retest values ranged from -0.26 – 0.31. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.38 and *CR%* = 110.52%, for visit 1 vs visit 3, *CR* = 0.31 and *CR%* = 99.37%,

and for visit 2 vs visit 3, *CR* = 0.19 and *CR%* = 68.25%. Based on the CR, an observed change in the Lombard response function of at least 0.19 – 0.38 would suggest a large amount of measurement error ranging from 68.25 – 110.52% variation in the Lombard response function of IWPD. These results suggest that the measure of Lombard response function demonstrates unacceptable repeatability for IWPD.

3.3.3.1.2 Retest analysis of Lombard response function for control participants

The ICC values related to the repeated measurement of Lombard response function across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .73, $p < .001$, visit 1 versus visit 3 *ICC* = .66, $p = .001$, and visit 2 versus visit 3 *ICC* = .77, $p < .001$. These results suggest that Lombard response function as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for Lombard response function was 0.05 (*SD* = 0.13) and this difference was significant $t(28) = 2.20$ $p = .036$. The mean difference between visit 1 and visit 3 for Lombard response function was 0.07 (*SD* = 0.15) and this difference was significant $t(29) = 2.44$, $p = .021$. The mean difference between visit 2 and visit 3 for Lombard response function was 0.02 (*SD* = 0.14) and this difference was not significant $t(28) = 0.81$, $p = .422$. The mean differences in retest values ranged from -0.18 – 0.34. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.21 and *CR%* = 65.57%, for visit 1 vs visit 3, *CR* = 0.24 and *CR%* = 94.01%, and for visit 2 vs visit 3, *CR* = 0.22 and *CR%* = 87.72%. Based on the CR, an observed change in the Lombard response function of at least 0.21 – 0.24 would suggest a large amount of measurement error ranging from 65.57 – 94.01% variation in the Lombard response function of control participants. These results suggest that the measure of Lombard response function demonstrates unacceptable repeatability for control participants.

3.3.4 Magnitude production

Measures of magnitude production were obtained from audio recordings of participants with PD and control participants reading the sentence "She saw Patty buy two poppies" using their typical speech loudness, speaking two times louder than normal, and speaking four times louder than normal. Descriptive statistics for magnitude production can be found in Table 10. The results of the two-way repeated measures ANOVA for the dependent measure of magnitude production revealed a significant main effect of "Group" $F(1,49) = 6.47$, $p = .014$, with participants with PD having a significantly lower (-0.77) marginal mean ($M = 2.09$, $SD = 1.08$) compared to that of the control participants ($M = 2.86$, $SD = 1.08$). A significant main effect of "Visit" was also found, $F(2,98) = 8.95$, $p < .001$. A closer look at the factor "Visit", using pairwise post-hoc analyses, indicated that the marginal mean of magnitude production at visit 1 ($M = 2.75$, $SD = 1.43$) was significantly greater (+0.51) than the marginal mean of magnitude production at visit 2 ($M = 2.24$, $SD = 1.00$) and (+0.33) visit 3 ($M = 2.42$, $SD = 1.14$). For the post-hoc comparison related to visit 1, the participants with PD had a mean magnitude production ($M = 2.35$, $SD = 1.53$) that was significantly lower (-0.81) than the mean magnitude production of control participants ($M = 3.16$, $SD = 1.32$), $t(2,51) = -2.02$, $p = .048$. For the post-hoc comparison related to visit 2, the participants with PD had a mean magnitude production ($M = 2.01$, $SD = 1.06$) that was lower (-0.59) than the mean magnitude production of control participants ($M = 2.60$, $SD = 1.09$). This comparison trended towards significance, $t(2,51) = -1.95$, $p = .056$. For the post-hoc comparison related to visit 3, the participants with PD had a mean magnitude production ($M = 2.02$, $SD = 0.97$) that was significantly lower (-0.79) than the mean magnitude production of control participants ($M = 2.81$, $SD = 1.16$), $t(2,50) = -2.58$, $p = .013$. Additionally, no significant "Group" by "Visit" interaction was found $F(2,98) = 0.08$, $p = .928$ for magnitude production. This non-significant interaction is illustrated in Figure 4. It seems that the magnitude production of participants with PD was significantly lower, or the difference approached significance, than that of control participants across all visits. Furthermore, it appears that the group difference between the magnitude production for participants with PD and control participants is fairly consistent across all three visits.

Figure 4: Means of Magnitude Production for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

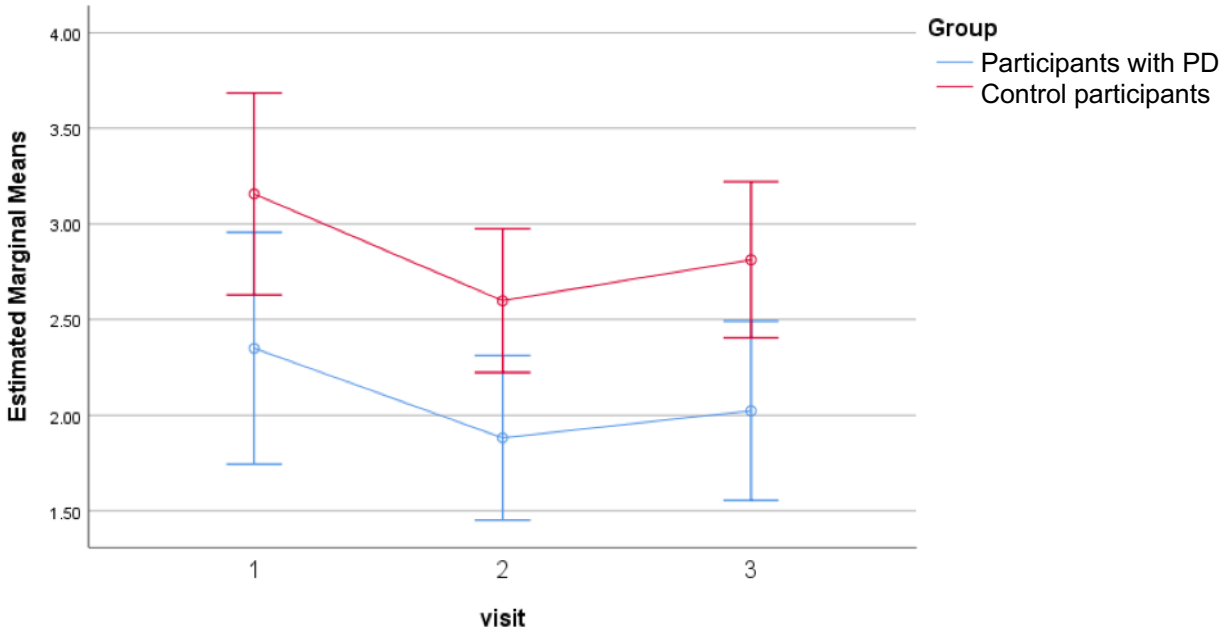

Figure 4. This figure demonstrates the changes in magnitude production scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.3.4.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of magnitude production were performed via 1) ICC between visits 1-2, 1-3 and 2-3, 2) pairwise t-tests between visits 1-2, 1-3 and 2-3, 3) calculation of the SEM and 4) determination of the CR related to visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants. Results for retest analyses of magnitude production are summarized in Table 14.

Table 14: Retest Analyses of Magnitude Production for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.77[.46-.90]	0.33	1.11	1.44	.16	1.30	0.62	1.73	73.73
Visits 1 – 3	.81[.55-.92]	0.33	1.01	1.52	.14	1.26	0.55	1.53	75.88
Visits 2 – 3	.91[.78-.96]	-0.14	0.54	-1.23	.23	1.02	0.30	0.84	41.80
Control participants									
Visits 1 – 2	.75[.41-.89]	0.56	1.00	3.01	.01	1.21	0.60	1.67	53.32
Visits 1 – 3	.85[.68-.93]	0.31	0.87	2.00	.06	1.24	0.48	1.33	51.05
Visits 2 – 3	.86[.70-.93]	-0.21	0.78	-1.48	.15	1.13	0.42	1.17	41.51

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of magnitude production for participants with PD and control participants.

3.3.4.1.1 Retest analysis of magnitude production for participants with Parkinson's disease

The ICC values related to the repeated measurement of magnitude production across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .77$, $p < .001$, visit 1 versus visit 3 $ICC = .81$, $p < .001$, and visit 2 versus visit 3 $ICC = .91$, $p < .001$. These results suggest that magnitude production as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for magnitude production was 0.33 ($SD = 1.11$) and this difference was not significant $t(22) = 1.44$, $p = .164$. The mean differences in retest values ranged from -2.62 – 2.69. The mean difference between visit 1 and visit 3 for magnitude production was 0.33 ($SD = 1.01$) and this difference was not significant $t(21) = 1.52$, $p = .143$. The mean differences in retest values ranged from -1.37 – 2.49. The mean difference between visit 2 and visit 3 for magnitude production was -0.14 ($SD = 0.54$) and this difference was not significant $t(21) = -1.23$, $p = .232$. The mean differences in retest values ranged from -1.02 – 0.93. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 1.73$ and $CR\% = 73.73\%$, for visit 1 vs visit 2, $CR = 1.53$ and $CR\% = 75.88\%$, and for visit 2

vs visit 3, *CR* = 0.84 and *CR%* = 41.80%. Based on the CR, an observed change in the magnitude production of at least 0.84 – 1.73 would suggest a fairly large amount of measurement error ranging from 41.80 – 73.73% variation in the magnitude production of IWPD. These results suggest that the measure of magnitude production demonstrates unacceptable repeatability for IWPD.

3.3.4.1.2 Retest analysis of magnitude production for control participants

The ICC values related to the repeated measurement of magnitude production across the three pairwise visits were found to be the following: visit 1 versus visit 2 *ICC* = .75, $p < .001$, visit 1 versus visit 3 *ICC* = .85, $p < .001$, and visit 2 versus visit 3 *ICC* = .86, $p < .001$. These results suggest that magnitude production as measured in the present study demonstrated good retest reliability in healthy speakers because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for magnitude production was 0.56 (*SD* = 1.00) and this difference was significant $t(28) = 3.01$, $p = .006$. The mean difference between visit 1 and visit 3 for magnitude production was 0.32 (*SD* = 0.87) and this difference trended towards significance $t(29) = 2.00$, $p = .055$. The mean differences in retest values ranged from -1.65 – 1.94. The mean difference between visit 2 and visit 3 for magnitude production was -0.21 (*SD* = 0.78) and this difference was not significant $t(28) = -1.48$, $p = 0.15$. The mean differences in retest values ranged from -1.79 – 1.49. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.67 and *CR%* = 53.32%, for visit 1 vs visit 3, *CR* = 1.33 and *CR%* = 51.05%, and for visit 2 vs visit 3, *CR* = 1.17 and *CR%* = 41.51%. Based on the CR, an observed change in the magnitude production of at least 1.17 – 1.67 would suggest a fairly large amount of measurement error ranging from 41.51 – 53.32% variation in the magnitude production of control participants. These results suggest that the measure of magnitude production demonstrates unacceptable repeatability for control participants.

3.4 Statistical Analysis for Objective 2: Typical Speech Loudness

In order to answer the question '*Do participants with PD perceive their typical speech loudness differently over time from their primary communicative partners and as compared to control participants?',* a multi-factor repeated measures ANOVA was performed to evaluate: 1) differences in perception of typical speech loudness over time among participants with PD, 2) how primary communication partners rate the loudness of his/her partner with PD, and 3) how control participants rate their typical speech loudness. The multi-factor repeated measures ANOVA involved a "Group" factor with three separate levels [PD, control, communication partners] and a "Visit" factor with three levels [visit 1, visit 2, and visit 3]. Whenever the ANOVA resulted in a significant main effect, a post-hoc analysis was performed using pairwise comparisons.

Measures of perceived typical speech loudness were obtained from the Typical Speech Loudness scale, wherein participant's VAS responses were measured and assigned a percentage value. Descriptive statistics for perceptions of typical speech loudness scores can be found in Table 15. The results of the multi-factor repeated measures ANOVA for the perceived typical speech loudness revealed a significant main effect of "Group" $F(2,71) = 29.73$, $p < .001$ with control participants having a greater (+31.90 %) marginal mean ($M = 83.22$, $SD = 17.69$) compared to participants with PD ($M = 51.32$, $SD = 17.68$) and a greater (+32.66%) marginal mean compared to the primary communication partners of participants with PD ($M = 50.56$, $SD = 17.68$). In contrast, there was no significant main effect of "Visit" on perceived typical speech loudness $F(2,142) = 2.48$, $p = .088$ with a marginal mean at visit 1 ($M = 58.31$, $SD = 21.68$) that was similar to visit 2 ($M = 64.08$, $SD = 21.51$) and visit 3 ($M = 62.71$, $SD = 23.91$). Additionally, there was no significant "Group" by "Visit" interaction $F(4,142) = 0.52$, $p = .723$ for perception of typical speech loudness. This interaction is illustrated in Figure 5.

Table 15: Descriptive Statistics of Self- and Proxy-Rated Typical Speech Loudness for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	Typical speech loudness mean(*SD*)
Participants with PD	
Visit 1	51.96*(23.78)*
Visit 2	52.35*(25.43)*
Visit 3	51.95*(26.05)*
Control participants	
Visit 1	78.60*(19.04)*
Visit 2	86.07*(15.02)*
Visit 3	85.00*(18.84)*
Primary communication partners	
Visit 1	46.22*(21.84)*
Visit 2	56.39*(24.36)*
Visit 3	51.18*(26.93)*
Marginal means	
Participants with PD	51.32*(17.68)*
Control participants	83.22*(17.69)*
Primary communication partners	50.56*(17.68)*
Visit 1	58.31*(21.68)*
Visit 2	64.08*(21.51)*
Visit 3	62.71*(23.91)*

Note. This table illustrates the means and standard deviations for perceptions of typical speech loudness for participants with PD, their primary communication partners, and control participants across visits.

Figure 5: Means of Typical Speech Loudness Perceptions by Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

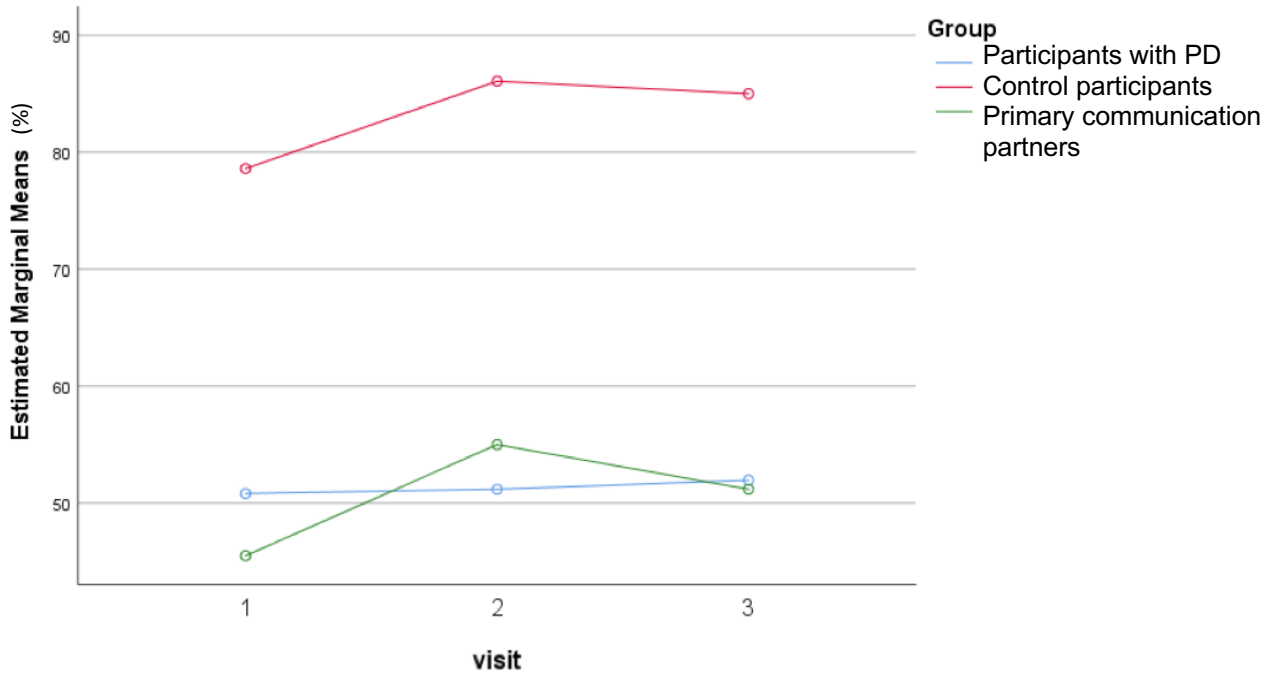

Figure 5. This figure demonstrates the changes in typical speech loudness perception scores for participants with PD, their primary communication partners, and control participants across visits. Error bars represent standard deviations.

3.4.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of typical speech loudness ratings. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD, control participants, and primary communication partners of participants with PD. Results for retest analyses of typical speech loudness are summarized in Table 16.

Table 16: Retest Analyses of Self- and Proxy-Rated Typical Speech Loudness for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.59[.01-.83]	-0.39	26.74	-0.07	.945	24.62	15.76	43.67	84.04
Visits 1 – 3	.50[-.24-.80]	-1.14	28.94	-0.18	.856	24.94	17.64	48.85	93.32
Visits 2 – 3	.85[.63-.94]	-0.77	18.96	-0.19	.850	25.74	9.97	27.62	53.16
Control participants									
Visits 1 – 2	.53[.06-.77]	7.47	18.95	-2.16	.039	17.15	11.76	32.56	41.43
Visits 1 – 3	.25[-.52-.64]	-6.40	24.72	-1.42	.167	18.94	16.40	45.44	52.79
Visits 2 – 3	.30[-.51-.67	1.07	21.95	0.267	.792	17.04	14.25	39.48	46.45
Primary communication partners									
Visits 1 – 2	.70[.30-.87]	-10.17	21.28	-2.29	.032	23.13	12.67	35.10	75.94
Visits 1 – 3	.77[.47-.91]	-5.68	20.99	-1.27	.218	24.52	11.76	32.57	57.76
Visits 2 – 3	.74[.36-.89]	3.82	23.46	0.76	.454	25.68	13.09	36.27	70.86

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of perceived typical speech loudness for participants with PD, their primary communication partners, and control participants.

3.4.1.1 Retest analysis of self-rated typical speech loudness for participants with Parkinson's disease

The ICC values related to the repeated measurement of self-rated typical speech loudness across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .59, $p = .023$, visit 1 versus visit 3 *ICC* = .50, $p = .066$, and visit 2 versus visit 3 *ICC* = .85, $p < .001$. These results suggest that self-rated typical speech loudness as measured in the present study did not demonstrate good retest reliability in IWPD because ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for self-rated typical speech loudness scores was -0.39 % (*SD* = 26.74) and this difference was not significant $t(22) = -0.07$, $p = .945$. The mean differences in retest values ranged from -55 – 69 %. The mean difference between visit 1 and visit 3 for self-ratings of typical speech loudness scores was -1.14 % (*SD* = 28.94) and this difference was not significant $t(21) = -0.18$, $p = .856$. The mean

differences in retest values ranged from -45 – 73 %. The mean difference between visit 2 and visit 3 for self-ratings of typical speech loudness scores was -0.77 % (*SD* = 18.96) and this difference was not significant $t(21) = -0.19$, $p = .850$. The mean differences in retest values ranged from -47 – 28 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 43.67 and *CR%* = 84.04%, for visit 1 vs visit 3, *CR* = 48.85 and *CR%* = 93.32%, and for visit 2 vs visit 3, *CR* = 27.62 and *CR%* = 53.16%. Based on the CR, an observed change in self-rated typical speech loudness on at least 27.62 – 48.85 % would suggest a fairly large amount of measurement error ranging from 53.16 – 93.32% variation in the self-ratings of typical speech loudness of IWPD. These results suggest that the measure of self-rated typical speech loudness demonstrates unacceptable repeatability for IWPD.

3.4.1.2 Retest analysis of self-rated typical speech loudness for control participants

The ICC values related to the repeated measurement of self-rated typical speech loudness across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .53, $p = .015$, visit 1 versus visit 3 *ICC* = .25 $p = .213$, and visit 2 versus visit 3 *ICC* = .30, $p = .180$. These results suggest that self-rated typical speech loudness as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for self-rated typical speech loudness was -7.47 % (*SD* = 18.95) and this difference was significant $t(29) = -2.16$, $p = .039$. The mean difference between visit 1 and visit 3 for self-rated typical speech loudness was -6.40 % (*SD* = 24.72) and this difference was not significant $t(29) = -1.42$, $p = .167$. The mean differences in retest values ranged from -53 – 79 %. The mean difference between visit 2 and visit 3 for self-rated typical speech loudness was 1.07 % (*SD* = 21.95) and this difference was not significant $t(29) = 0.27$, $p = .792$. The mean differences in retest values ranged from -52 – 75 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 32.56 and *CR%* = 41.43%, for visit 1 vs visit 3, *CR* = 45.44 and *CR%* = 52.79%, and for visit 2 vs visit 3, *CR* = 39.48 and *CR%* = 46.45%. Based on the CR, an observed change in the self-rated typical

speech loudness of at least 32.56 – 45.44 % would suggest a fairly large amount of measurement error ranging from 41.43 – 52.79% variation in the self-ratings of typical speech loudness of control participants. These results suggest that the measure of self-rated typical speech loudness demonstrates unacceptable repeatability for control participants.

3.4.1.3 Retest analysis of proxy-rated typical speech loudness by primary communication partners

The ICC values related to the repeated measurement of proxy-rated typical speech loudness across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .70 *p* = .002, visit 1 versus visit 3 *ICC* = .77, *p* = .001, and visit 2 versus visit 3 *ICC* = .74, *p* = .002. These results suggest that proxy-rated typical speech loudness as measured in the present study did not demonstrate good retest reliability by primary communication partners of IWPD because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for proxy-rated typical speech loudness scores was -10.17 % (*SD* = 21.28) and this difference was significant *t*(22) = -2.29, *p* = .032. The mean difference between visit 1 and visit 3 for proxy-ratings of typical speech loudness scores was -5.68 % (*SD* = 20.99) and this difference was not significant *t*(21) = -1.27, *p* = .218. The mean differences in retest values ranged from -47 – 40 %. The mean difference between visit 2 and visit 3 for proxy-rated typical speech loudness scores was 3.82 % (*SD* = 23.46) and this difference was not significant *t*(21) = 0.76, *p* = .454. The mean differences in retest values ranged from -53 – 46 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 35.10 and *CR%* = 75.94%, for visit 1 vs visit 3, *CR* = 32.57 and *CR%* = 57.76%, and for visit 2 vs visit 3, *CR* = 36.27 and *CR%* = 70.86%. Based on the CR, an observed change in the proxy-rated typical speech loudness of at least 32.57 – 36.27 % would suggest a large amount of measurement error ranging from 57.76 – 75.94% variation in the proxy-ratings of typical speech loudness made by primary communication partners of IWPD. These results suggest that the measure of proxy-rated typical speech loudness demonstrates unacceptable repeatability for primary communication partners.

3.5 Statistical Analysis for Objective 3: Speech Intelligibility

In order to answer the question '*Do speech intelligibility measures differ over time between and within participants with PD and control participants?',* a separate two-factor repeated measures ANOVA was performed on each of the three dependent intelligibility measures (i.e., SIT transcription scores, SIT VAS scores, conversational intelligibility scores). Each of the two-factor repeated measures ANOVAs involved a "Group" factor with two separate levels [PD, control] and a "Visit" factor with three levels [visit 1, visit 2, and visit 3]. Whenever the ANOVA resulted in a significant main effect for the factor "Visit", a post-hoc analysis was performed using pairwise comparisons.

3.5.1 SIT transcription scores

SIT transcription scores were obtained by calculating the percentage of correct words transcribed by listeners of SIT sentences 13, 14, and 15. Descriptive statistics for SIT transcription scores can be found in Table 17. The results of the two-way repeated measures ANOVA for the dependent measure of SIT transcription scores revealed a significant main effect of "Group" $F(1,49) = 12.50$, $p = .001$ with participants with PD having a significantly lower (-3.92 %) marginal mean ($M = 94.20$, $SD = 3.94$) to that of control participants ($M = 98.12$, $SD = 3.93$). In contrast, there was no significant main effect of "Visit" on SIT transcription scores $F(2,98) = 1.96$, $p = .146$ indicating that the marginal mean of the SIT transcription score at visit 1 ($M = 96.42$, $SD = 4.21$) was similar than the mean SIT transcription score at visit 2 ($M = 95.48$, $SD = 5.36$) and visit 3 ($M = 96.59$, $SD = 4.36$). Additionally, no significant "Group" by "Visit" interaction was found $F(2,98) = 1.08$, $p = .342$ for SIT transcription scores. This non-significant interaction is illustrated in Figure 6.

Table 17: Descriptive Statistics of Speech Intelligibility Measures for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	SIT transcription scores mean*(SD)*	SIT VAS scores mean*(SD)*	Conversational intelligibility VAS scores mean*(SD)*
Participants with PD			
Visit 1	94.27*(6.09)*	79.16*(11.26)*	79.95*(14.93)*
Visit 2	92.20*(8.53)*	78.10*(15.01)*	80.55*(14.03)*
Visit 3	95.06*(6.29)*	79.38*(10.06)*	80.75*(15.58)*
Control participants			
Visit 1	98.41*(1.33)*	91.12*(2.91)*	93.79*(3.00)*
Visit 2	97.89*(2.40)*	90.99*(2.88)*	92.15*(5.25)*
Visit 3	98.15*(1.75)*	90.81*(4.25)*	92.74*(3.39)*
Marginal means			
Participants with PD	94.20*(3.94)*	79.82*(6.89)*	80.74*(9.38)*
Control participants	98.12*(3.93)*	90.91*(6.89)*	92.88*(9.37)*
Visit 1	96.42*(4.21)*	85.41*(7.71)*	86.96*(10.36)*
Visit 2	95.48*(5.36)*	85.64*(7.57)*	86.74*(10.00)*
Visit 3	96.59*(4.36)*	85.03*(7.43)*	86.74*(10.64)*

Note. This table illustrates the means and standard deviations for the different speech intelligibility measures for participants with PD and control participants across visits.

Figure 6: Means of Speech Intelligibility Test Transcription Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

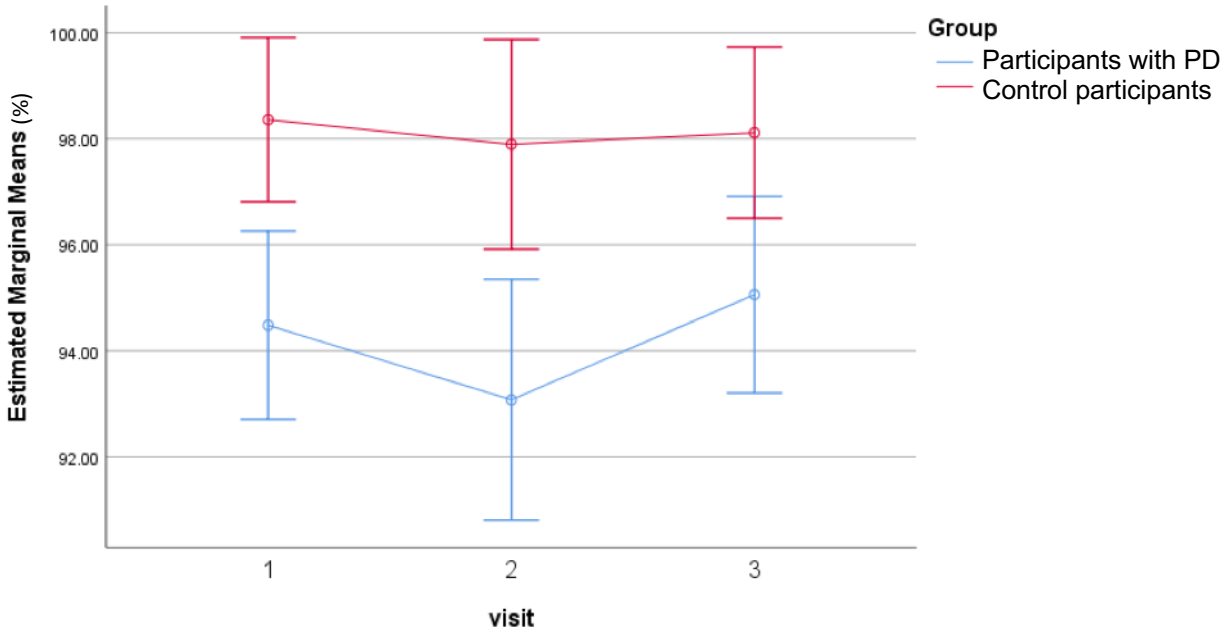

Figure 6. This figure demonstrates the changes in SIT transcription scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.5.1.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of SIT transcription scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants. Results for retest analyses of SIT transcription scores are summarized in Table 18.

Table 18: Retest Analyses of Sentence Intelligibility Transcription Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.74[.40-.89]	2.07	6.62	1.50	.148	7.41	3.38	10.47	11.10
Visits 1 – 3	.77[.44-.90]	-.578	5.47	-0.50	.625	6.19	2.97	8.22	8.92
Visits 2 – 3	.76[.43-.90]	-1.98	6.08	-1.53	.141	7.49	3.67	10.17	10.70
Control participants									
Visits 1 – 2	.28[-.54-.66]	0.46	2.51	1.00	.328	1.94	1.65	4.56	4.63
Visits 1 – 3	.05[-1.03-.55]	0.27	2.17	0.67	.509	1.55	1.51	4.20	4.29
Visits 2 – 3	.42[-.25-.73]	-0.22	2.56	-0.46	.651	2.10	1.60	4.43	4.51

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of SIT transcription scores for participants with PD and control participants.

3.5.1.1.1 Retest analysis of SIT transcription scores of participants with Parkinson's disease

The ICC values related to the repeated measurement of SIT transcription scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .74, *p* = .001, visit 1 versus visit 3 *ICC* = .77, *p* = .001, and visit 2 versus visit 3 *ICC* = .76, *p* = .001. These results suggest that SIT transcription scores as measured in the present study demonstrated good retest reliability in IWPD because the ICC values across all comparisons were above or approached our criterion of .75.

The mean difference between visit 1 and visit 2 for SIT transcription scores was 2.07 % (*SD* = 6.62) and this difference was not significant *t*(22) = 1.50, *p* = .148. The mean differences in retest values ranged from -7.14 – 18.25 %. The mean difference between visit 1 and visit 3 for SIT transcription scores was -0.58 % (*SD* = 5.47) and this difference was not significant *t*(21) = -0.50, *p* = .625. The mean differences in retest values ranged from -15.87 – 9.52 %. The mean difference between visit 2 and visit 3 for SIT transcription scores was -1.98% (*SD* = 6.08) and this difference was not significant *t*(21) = -1.53, *p* = .141. The mean differences in retest values ranged from -19.05 – 7.14 %. Additionally, the following CR values were obtained for the following visit comparisons:

for visit 1 vs visit 2, *CR* = 10.47 and *CR%* = 11.10%, for visit 1 vs visit 3, *CR* = 8.22 and *CR%* = 8.92%, and for visit 2 vs visit 3, *CR* = 10.17 and *CR%* = 10.70%. Based on the CR, an observed change in the SIT transcription scores of at least 8.22 – 10.47% would suggest an acceptable amount of measurement error ranging from 8.92 – 11.10% variation in the SIT transcription scores of IWPD. These results suggest that the measure of SIT transcription scores demonstrates fairly good repeatability for IWPD.

3.5.1.1.2 Retest analysis of SIT transcription scores for control participants

The ICC values related to the repeated measurement of SIT transcription scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .28, *p* = .200, visit 1 versus visit 3 *ICC* = .05, *p* = .443, and visit 2 versus visit 3 *ICC* = .42, *p* = .081. These results suggest that SIT transcription scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for SIT transcription scores was 0.46 % (*SD* = 2.51) and this difference was not significant *t*(28) = 1.00, *p* = .328. The mean differences in retest values ranged from -3.17 – 4.76 %. The mean difference between visit 1 and visit 3 for SIT transcription scores was 0.27 % (*SD* = 2.17) and this difference was not significant *t*(29) = 0.67, *p* = .509. The mean differences in retest values ranged from -3.17 – 5.56 %. The mean difference between visit 2 and visit 3 for SIT transcription scores was -0.22 % (*SD* = 2.56) and this difference was not significant *t*(28) = -0.46, *p* = .651. The mean differences in retest values ranged from -4.76 – 4.76 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 4.56 and *CR%* = 4.63%, for visit 1 vs visit 3, *CR* = 4.20 and *CR%* = 4.63%, and for visit 2 vs visit 3, *CR* = 4.43 and *CR%* = 4.51%. Based on the CR, an observed change in the SIT transcription scores of at least 4.20 – 4.56 % would suggest an acceptable amount of measurement error ranging from 4.29 – 4.63% variation in the SIT transcription scores of control participants. These results suggest that the measure of SIT transcription scores demonstrates good repeatability for control participants.

3.5.2 SIT VAS scores

SIT VAS scores were calculated for each speaker as percent intelligibility based on listeners ratings of SIT sentences 13, 14, and 15. Descriptive statistics for SIT VAS scores can be found in Table 17. The results of the two-way repeated measures ANOVA for the dependent measure of SIT VAS scores revealed a significant main effect of "Group" $F(1,49) = 32.58$, $p < .001$ with participants with PD having a significantly lower (-11.09 %) marginal mean ($M = 79.82$, $SD = 6.89$) to that of control participants ($M = 90.91$, $SD = 6.89$). In contrast, there was no significant main effect of "Visit" on SIT VAS scores $F(2,98) = 0.372$, $p = .690$ indicating that the marginal mean of the SIT VAS score at visit 1 ($M = 85.41$, $SD = 7.71$) was similar than the mean SIT VAS score at visit 2 ($M = 85.64$, $SD = 7.57$) and visit 3 ($M = 85.03$, $SD = 7.43$). Additionally, no significant "Group" by "Visit" interaction was found $F(2,98) = 0.11$, $p = .896$ for SIT VAS scores. This non-significant interaction is illustrated in Figure 7.

Figure 7: Means of Speech Intelligibility Test Visual Analogues Scale Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

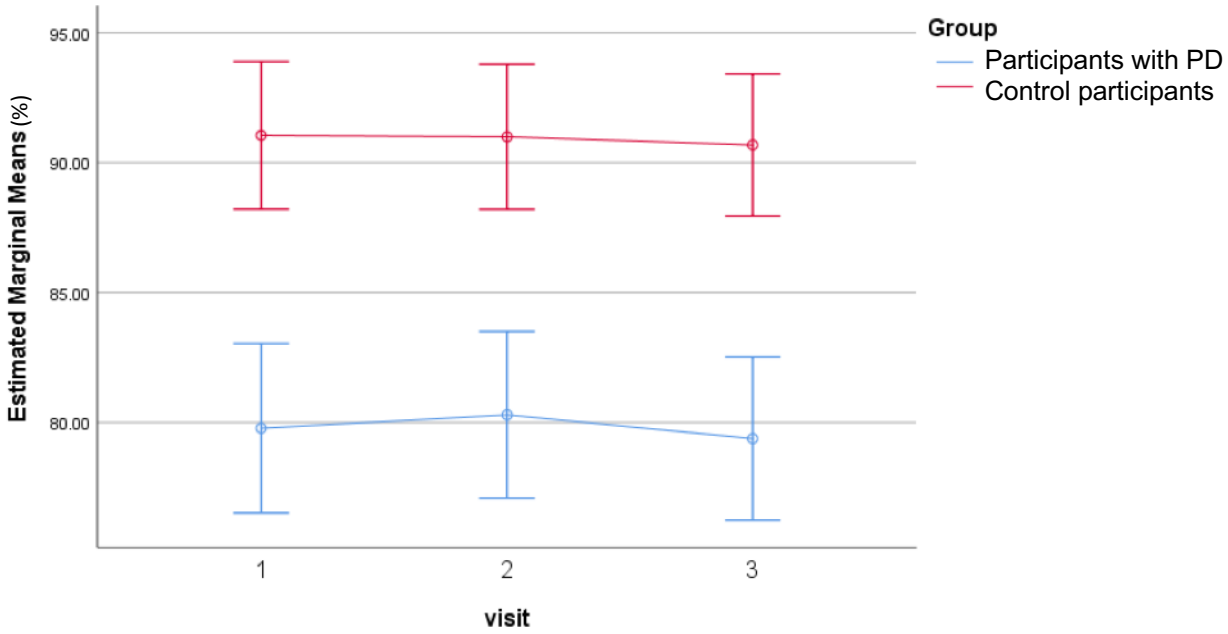

Figure 7. This figure demonstrates the changes in SIT VAS scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.5.2.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of SIT VAS scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants. Results for retest analyses of SIT VAS scores are summarized in Table 19.

Table 19: Retest Analyses of Speech Intelligibility Test Visual Analogue Scale Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.84[.61-.93]	1.06	10.08	0.51	.618	13.27	5.31	14.70	18.57
Visits 1 – 3	.89[.73-.95]	0.40	6.89	0.27	.787	10.68	3.54	9.81	12.56
Visits 2 – 3	.87[.69-.95]	0.92	7.15	0.60	.554	12.78	4.61	12.76	16.08
Control participants									
Visits 1 – 2	.81[.59-.91]	0.05	2.37	0.12	.903	2.90	1.26	3.50	3.84
Visits 1 – 3	.71[.39-.86]	0.31	3.47	0.49	.630	3.64	1.96	5.43	5.97
Visits 2 – 3	.77[.51-.89]	0.32	3.17	0.54	.596	3.63	1.74	4.82	5.31

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of SIT VAS scores for participants with PD and control participants.

3.5.2.1.1 Retest analysis of SIT VAS scores for participants with Parkinson's disease

The ICC values related to the repeated measurement of SIT VAS scores across the three pairwise visits were found to be the following: visit 1 versus visit 2 $ICC = .84$, $p < .001$, visit 1 versus visit 3 $ICC = .89$, $p < .001$, and visit 2 versus visit 3 $ICC = .87$, $p < .001$. These results suggest that SIT VAS scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for SIT VAS scores was 1.06 % (SD = 10.08) and this difference was not significant $t(22) = 0.51$, $p = .618$. The mean differences in retest values ranged from -18.97 – 35.18 %. The mean difference between visit 1 and visit 3 for SIT VAS scores was 0.40 % (SD = 6.89) and this difference was not significant $t(21) = 0.27$, $p = .787$. The mean differences in retest values ranged from -10.84 – 16.13 %. The mean difference between visit 2 and visit 3 for SIT VAS scores was 0.92 % (SD = 7.15) and this difference was not significant $t(21) = 0.60$, $p = .554$. The mean differences in retest values ranged from -15.97 – 17.53 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit

2, *CR* = 14.70 and *CR%* = 18.57%, for visit 1 vs visit 3, *CR* = 9.81 and *CR%* = 12.56%, and for visit 2 vs visit 3, *CR* = 12.76 and *CR%* = 16.08%. Based on the CR, an observed change in the SIT VAS scores of at least 9.81 – 14.70% would suggest the possibility of an acceptable amount of measurement error ranging from 12.56 – 18.57% variation in the SIT VAS scores of IWPD. These results suggest that the measure of SIT VAS scores demonstrates marginal repeatability for IWPD.

3.5.2.1.2 Retest analysis of SIT VAS scores for control participants

The ICC values related to the repeated measurement of SIT VAS scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .81, $p < .001$, visit 1 versus visit 3 *ICC* = .71, $p = .001$, and visit 2 versus visit 3 *ICC* = .77, $p < .001$. These results suggest that SIT VAS scores as measured in the present study demonstrated good retest reliability in healthy speakers because the ICC values two 2 of three comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for SIT VAS scores was 0.05 % (*SD* = 2.37) and this difference was not significant $t(28) = 0.12$, $p = .903$. The mean differences in retest values ranged from -4.43 – 4.56 %. The mean difference between visit 1 and visit 3 for SIT VAS scores was 0.31 % (*SD* = 3.47) and this difference was not significant $t(29) = 0.49$, $p = .630$. The mean differences in retest values ranged from -7.22 – 9.03 %. The mean difference between visit 2 and visit 3 for SIT VAS scores was 0.32 % (*SD* = 3.17) and this difference was not significant $t(28) = 0.54$, $p = .596$. The mean differences in retest values ranged from -2.83 – 8.57 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 3.50 and *CR%* = 3.84%, for visit 1 vs visit 3, *CR* = 5.43 and *CR%* = 5.97%, and for visit 2 vs visit 3, *CR* = 4.82 and *CR%* = 5.31%. Based on the CR, an observed change in the SIT VAS scores of at least 3.50 – 5.43% would suggest an acceptable amount of measurement error ranging from 3.84 – 5.97% variation in the SIT VAS scores of control participants. These results suggest that the measure of SIT VAS scores demonstrates good repeatability for control participants.

3.5.3 Conversational intelligibility VAS scores

Conversational intelligibility VAS scores were obtained by listeners indicating the percent intelligibility of a speaker for a conversational speech sample. Descriptive statistics for conversational intelligibility VAS scores can be found in Table 17. The results of the two-way repeated measures ANOVA for the dependent measure of conversational intelligibility VAS scores revealed a significant main effect of "Group" $F(1,49) = 21.09$, $p < .001$ with participants with PD having a significantly lower (-12.14 %) marginal mean ($M = 80.74$, $SD = 9.38$) to that of control participants ($M = 92.88$, $SD = 9.37$). In contrast, there was no significant main effect of "Visit" on conversational intelligibility VAS scores $F(2,98) = 0.031$, $p = .970$ indicating that the marginal mean of the conversational intelligibility VAS score at visit 1 ($M = 86.96$, $SD = 10.36$) was similar to the marginal mean of the conversation VAS score at visit 2 ($M = 86.74$, $SD = 10.00$) and at visit 3 ($M = 86.74$, $SD = 10.64$). Additionally, no significant "Group" by "Visit" interaction was found $F(2,98) = 1.00$, $p = .371$ for conversational intelligibility VAS scores. This non-significant interaction is illustrated in Figure 8.

Figure 8: Means of Conversational Intelligibility Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

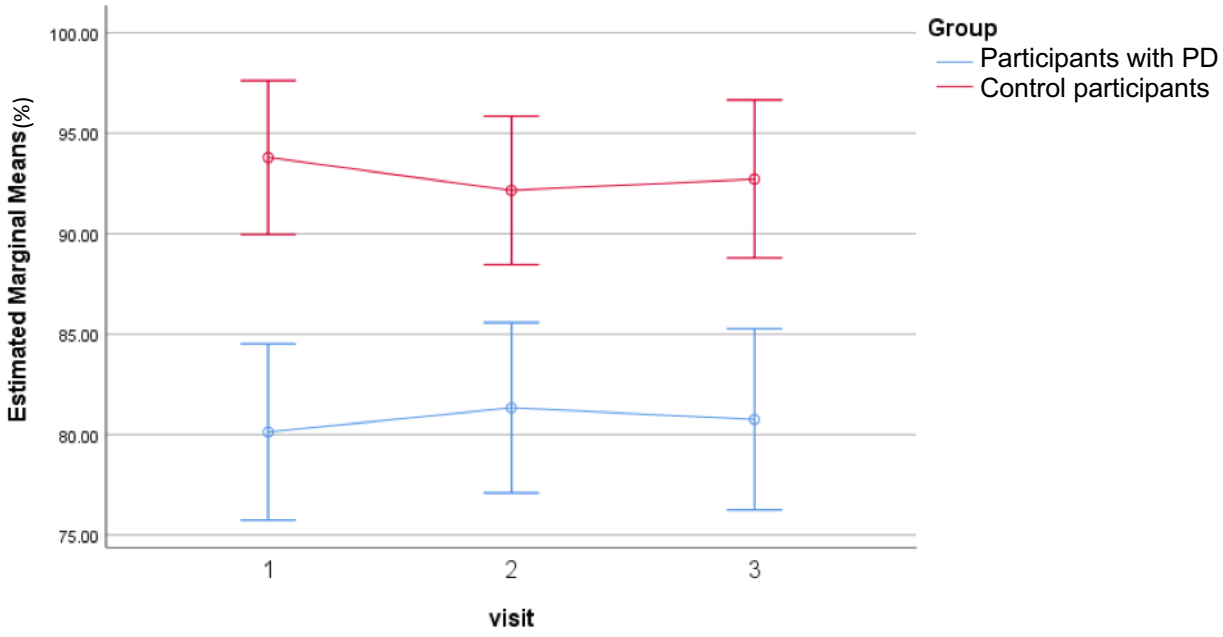

Figure 8. This figure demonstrates the changes in conversational intelligibility scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.5.3.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of conversational intelligibility VAS scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants. Results for retest analyses of conversational intelligibility VAS scores are summarized in Table 20.

Table 20: Retest Analyses of Conversational Intelligibility Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.83[.60-.93]	-0.59	11.20	-0.25	.802	14.49	1.76	4.88	5.21
Visits 1 – 3	.93[.83-.97]	-0.63	8.17	-0.36	.721	15.26	0.85	2.35	2.55
Visits 2 – 3	.87[.68-.95]	0.58	10.21	0.27	.794	14.83	1.59	4.41	4.76
Control participants									
Visits 1 – 2	.68[.33-.85]	1.63	4.10	2.14	.041	4.28	2.42	6.70	7.14
Visits 1 – 3	.78[.54-.90]	1.05	2.60	2.22	.034	3.20	1.50	4.16	4.51
Visits 2 – 3	.56[.06-.80]	-0.56	4.92	-0.62	.543	4.42	2.93	8.12	8.76

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of conversational intelligibility scores for participants with PD and control participants.

3.5.3.1.1 Retest analysis of VAS conversational intelligibility scores for participants with Parkinson's disease

The ICC values related to the repeated measurement of conversational intelligibility VAS scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .83$, $p < .001$, visit 1 versus visit 3 $ICC = .93$, $p < .001$, and visit 2 versus visit 3 $ICC = .87$, $p < .001$. These results suggest that conversational intelligibility as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for conversational intelligibility VAS scores was -0.59 % ($SD = 11.20$) and this difference was not significant $t(22) = -0.25$, $p = .802$. The mean differences in retest values ranged from -30.25 – 21.22 %. The mean difference between visit 1 and visit 3 for conversational intelligibility VAS scores was -0.63 % ($SD = 8.17$) and this difference was not significant $t(21) = -0.36$, $p = .721$. The mean differences in retest values ranged from -21.22 – 16.26 %. The mean difference between visit 2 and visit 3 for conversational intelligibility VAS scores was 0.58 % ($SD = 10.21$) and this difference was not significant $t(21) = 0.27$, $p = .794$. The mean differences in retest values ranged from -27.59 – 19.52 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 4.88$

and *CR%* = 5.21%, for visit 1 vs visit 3, *CR* = 2.35 and *CR%* = 2.55%, and for visit 2 vs visit 3, *CR* = 4.41 and *CR%* = 4.76%. Based on the CR, an observed change in the conversational intelligibility VAS scores of at least 2.35 – 4.88% would suggest an acceptable amount of measurement error ranging from 2.55 – 5.21% variation in the conversation VAS scores of IWPD. These results suggest that the measure of conversational VAS scores demonstrates good repeatability for IWPD.

3.5.3.1.2 Retest analysis of VAS conversational intelligibility scores for control participants

The ICC values related to the repeated measurement of conversational intelligibility VAS scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .68, $p = .001$, visit 1 versus visit 3 *ICC* = .78, $p < .001$, and visit 2 versus visit 3 *ICC* = .56, $p = .018$. These results suggest that conversational intelligibility as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for conversational intelligibility VAS scores was 1.63 % (*SD* = 4.10) and this difference was significant $t(28) = 2.14$, $p = .041$. The mean difference between visit 1 and visit 3 for conversational intelligibility VAS scores was 1.05 % (*SD* = 2.60) and this difference was significant $t(29) = 2.22$, $p = .034$. The mean difference between visit 2 and visit 3 for conversational intelligibility VAS scores was -0.56 % (*SD* = 4.92) and this difference was not significant $t(28) = -0.62$, $p = .543$. The mean differences in retest values ranged from -12.89 – 10.47 %. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 6.70 and *CR%* = 7.14%, for visit 1 vs visit 3, *CR* = 4.16 and *CR%* = 4.51%, and for visit 2 vs visit 3, *CR* = 8.12 and *CR%* = 8.76%. Based on the CR, an observed change in the conversational intelligibility VAS scores of at least 4.16 – 8.12% would suggest an acceptable amount of measurement error ranging from 4.51 – 8.76% variation in the conversational intelligibility VAS scores of control participants. These results suggest that the measure of conversational intelligibility VAS scores demonstrates good repeatability for control participants.

3.6 Statistical Analysis for Objective 4: Self-Rated Communicative Participation

In order to answer the question '*Does self-rated communicative participation differ over time between and within participants with PD and control participants?',* a repeated measures MANOVA was performed to analyze the variability of self-rated communicative participation of the participants with PD and control participants over three time points for two of the four dependent measures related to communicative participation (CES question scores, VAPP subtest scores). The following factors were used in this analysis: one between-group independent factor with two levels "Group" [PD, control], one within-group independent factor with three levels "Visit" [visit 1, visit 2, visit 3]. Additionally, a repeated measures ANOVA was performed to analyze the variability of self-rated communicative participation of the participants with PD and control participants over three time points for the remaining two of four dependent measures related to communicative participation (CPIB scores, LSUS scores). The following factors were used in this analysis: one between-group independent factor with two levels "Group" [PD, control], one within-group independent factor with three levels "Visit" [visit 1, visit 2, visit 3].

3.6.1 CES question scores

Scores for each of the eight individual CES questions were obtained based on participants' responses of 1, 2, 3, or 4 for each question. Descriptive statistics for individual CES question scores can be found in Table 21. The results of the repeated measures MANOVA for the dependent variables of the individual CES questions showed that there was a significant multivariate effect of "Group" $F(8,41) = 14.54$, $p < .001$. The results of this repeated measures MANOVA can be found in Table 22. The results of subsequent univariate testing for each of the individual CES questions revealed the following results: 1) there was a significant univariate effect of "Group" $F(1,48) = 67.63$, $p < .001$ for CES question 1 (*Having a conversation with a family member or friends at home*) with participants with PD having a lower (-0.97) marginal mean ($M = 2.92$, $SD = 0.41$) compared to control participants ($M = 3.89$, $SD = 0.43$); 2) there was a significant univariate effect of "Group" $F(1,48) = 89.69$, $p < .001$ for CES question 2 (*Participating*

in conversation with strangers in a quiet place) with participants with PD having a lower (-1.14) marginal mean (M = 2.71, SD = 0.41) compared to control participants (M = 3.85, SD = 0.43); 3) there was a significant univariate effect of "Group" $F(1,48) = 82.13$, $p < .001$ for CES question 3 (*Conversing with a familiar person over the telephone*) with participants with PD having a lower (-1.03) marginal mean (M = 2.86, SD = 0.41) compared to control participants (M = 3.89, SD = 0.38); 4) there was a significant univariate effect of "Group" $F(1,48) = 91.67$, $p < .001$ for CES question 4 (*Conversing with a stranger over the telephone*) with participants with PD having a lower (-1.26) marginal mean (M = 2.51, SD = 0.46) compared to control participants (M = 3.77, SD = 0.48); 5) there was a significant univariate effect of "Group" $F(1,48) = 78.29$, $p < .001$ for CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) with participants with PD having a lower (-1.33) marginal mean (M = 2.19, SD = 0.50) compared to control participants (M = 3.52, SD = 0.54); 6) there was a significant univariate effect of "Group" $F(1,48) = 89.40$, $p < .001$ for CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) with participants with PD having a lower (-1.26) marginal mean (M = 2.37, SD = 0.46) compared to control participants (M = 3.63, SD = 0.48); 7) there was a significant univariate effect of "Group" $F(1,48) = 98.29$, $p < .001$ for CES question 7 (*Having a conversation while traveling in a car*) with participants with PD having a lower (-1.19) marginal mean (M = 2.59, SD = 0.41) compared to control participants (M = 3.78, SD = 0.43); and 8) there was a significant univariate effect of "Group" $F(1,48) = 69.19$, $p < .001$ for CES question 8 (*Having a conversation with someone at a distance (across a room)*) with participants with PD having a lower (-1.29) marginal mean (M = 2.29, SD = 0.55) compared to control participants (M = 3.58, SD = 0.54).

Table 21: Descriptive Statistics of Self- and Proxy-Rated Individual Questions of the Communication Effectiveness Survey for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	CES Q1 mean*(SD)*	CES Q2 mean*(SD)*	CES Q3 mean*(SD)*	CES Q4 mean*(SD)*	CES Q5 mean*(SD)*	CES Q6 mean*(SD)*	CES Q7 mean*(SD)*	CES Q8 mean*(SD)*
Participants with PD								
Visit 1	3.00*(0.60)*	2.83*(0.72)*	3.05*(0.65)*	2.65*(0.78)*	2.30*(0.76)*	2.57*(0.90)*	2.70*(0.70)*	2.30*(0.63)*
Visit 2	2.87*(0.81)*	2.78*(0.60)*	2.87*(0.69)*	2.39*(0.58)*	2.04*(0.56)*	2.30*(0.63)*	2.61*(0.66)*	2.35*(0.65)*
Visit 3	2.82*(0.59)*	2.50*(0.67)*	2.73*(0.77)*	2.45*(0.74)*	2.09*(0.87)*	2.23*(0.81)*	2.45*(0.67)*	2.18*(0.91)*
Control participants								
Visit 1	3.77*(0.43)*	3.67*(0.55)*	3.80*(0.41)*	3.60*(0.56)*	3.27*(0.69)*	3.45*(0.57)*	3.73*(0.45)*	3.27*(0.74)*
Visit 2	3.87*(0.35)*	3.90*(0.31)*	3.90*(0.31)*	3.87*(0.35)*	3.53*(0.57)*	3.70*(0.53)*	3.73*(0.45)*	3.57*(0.63)*
Visit 3	3.97*(0.18)*	3.97*(0.18)*	3.93*(0.25)*	3.83*(0.38)*	3.70*(0.53)*	3.77*(0.43)*	3.80*(0.48)*	3.77*(0.43)*
Primary communication partners								
Visit 1	3.30*(0.76)*	3.30*(0.56)*	3.17*(0.72)*	2.83*(0.83)*	2.65*(0.88)*	3.00*(0.87)*	3.04*(0.64)*	2.57*(0.90)*
Visit 2	3.30*(0.82)*	3.00*(0.67)*	3.13*(0.81)*	2.52*(0.85)*	2.43*(0.84)*	2.55*(0.86)*	2.91*(1.00)*	2.39*(0.84)*
Visit 3	3.36*(0.73)*	3.00*(0.76)*	3.18*(0.66)*	2.73*(0.88)*	2.36*(0.90)*	2.67*(0.80)*	3.09*(0.68)*	2.32*(0.95)*
Marginal means								
Participants with PD	2.92*(0.41)*	2.71*(0.41)*	2.86*(0.41)*	2.51*(0.46)*	2.19*(0.50)*	2.37*(0.46)*	2.59*(0.41)*	2.29*(0.55)*
Control participants	3.89*(0.43)*	3.85*(0.43)*	3.89*(0.38)*	3.77*(0.48)*	3.52*(0.54)*	3.63*(0.48)*	3.78*(0.43)*	3.58*(0.54)*
Communication partners	3.35*(0.60)*	3.13*(0.55)*	3.18*(0.55)*	2.71*(0.64)*	2.51*(0.69)*	2.73*(0.64)*	3.03*(0.55)*	2.46*(0.69)*
PD and control visit 1	3.42*(0.49)*	3.27*(0.64)*	3.41*(0.49)*	3.14*(0.71)*	2.85*(0.71)*	3.03*(0.71)*	3.24*(0.57)*	2.82*(0.71)*
PD and control visit 2	3.40*(0.57)*	3.33*(0.42)*	3.38*(0.57)*	3.12*(0.49)*	2.82*(0.57)*	2.99*(0.57)*	3.17*(0.57)*	3.00*(0.64)*
PD and control visit 3	3.39*(0.42)*	3.25*(0.49)*	3.32*(0.57)*	3.15*(0.57)*	2.89*(0.71)*	2.98*(0.64)*	3.15*(0.57)*	2.97*(0.71)*
PD and partners visit 1	3.19*(0.71)*	3.12*(0.65)*	3.10*(0.71)*	2.76*(0.84)*	2.57*(0.78)*	2.81*(0.91)*	2.88*(0.65)*	2.48*(0.78)*
PD and partners visit 2	3.12*(0.84)*	2.88*(0.65)*	3.00*(0.78)*	2.45*(0.78)*	2.26*(0.71)*	2.41*(0.78)*	2.76*(0.84)*	2.41*(0.78)*
PD and partners visit 3	3.10*(0.65)*	2.76*(0.71)*	2.95*(0.71)*	2.62*(0.84)*	2.21*(0.91)*	2.43*(0.78)*	2.79*(0.71)*	2.24*(0.97)*

Note. This table illustrates the means and standard deviations for the individual questions of the Communicative Effectiveness Survey for participants with PD, their primary communication partners, and control participants across visits.

Table 22: Multivariate Testing Analysis of Self-Rated Individual Questions of the Communicative Effectiveness Survey for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p* value
"Group"	8	41	14.54	<.001
"Visit"	16	180	0.68	.810
"Group"*"Visit"	16	180	2.90	<.001

Note. This table illustrates the main effects of "Group" and "Visit", and the "Group" by "Visit" interaction found from multivariate analyses of the individual questions of the Communicative Effectiveness Survey for participants with PD and control participants.

The results of the repeated measures MANOVA for the dependent variables of the individual CES questions showed that there was no significant multivariate effect of "Visit" $F(16,180) = 0.68$, $p = .810$. In contrast, there was a significant "Group" by "Visit" interaction on the individual CES questions $F(16,180) = 2.90$, $p < .001$. The results of this repeated measures MANOVA can be found in Table 22. A closer look at the "Visit" factor via subsequent univariate testing for each of the individual CES questions revealed the following results: 1) There was no significant univariate effect of "Visit" $F(2,96) = 0.10$, $p = .901$, but there was a significant "Group" by "Visit" interaction $F(2,96) = 4.06$, $p = .020$ for CES question 1 (*Having a conversation with a family member or friends at home*). This significant interaction is illustrated in Figure 9. These results indicated that the marginal mean of the CES question 1 score at visit 1 ($M = 3.42$, $SD = 0.49$) was similar to the marginal mean of the CES question 1 score at visit 2 ($M = 3.40$, $SD = 0.57$) and visit 3 ($M = 3.39$, $SD = 0.42$). It appears that at visit 1 the group difference between the ratings of participants with PD and control participants on CES question 1 is smaller than the group differences that were found at visit 2 and visit 3. 2) There was no significant univariate effect of "Visit" $F(2,96) = 0.71$, $p = .496$, but there was a significant "Group" by "Visit" interaction $F(2,96) = 8.87$, $p < .001$ for CES question 2 (*Participating in conversation with strangers in a quiet place*). This significant interaction is illustrated in Figure 10. These results indicated that the marginal mean of CES question 2 at visit 1 ($M = 3.27$, $SD = 0.64$) was similar to the marginal mean the CES question 2 score at visit 2 ($M = 3.33$, $SD = 0.42$) and visit 3 ($M = 3.25$, $SD = 0.49$). It appears that at visit 1 the group difference between the ratings of participants with PD

and control participants on CES question 2 is smaller than the group differences that were found at visit 2 and visit 3. 3) There was no significant univariate effect of "Visit" $F(2,96) = 0.58$, $p = .563$ or "Group" by "Visit" interaction $F(2,96) = 2.61$, $p = .078$ for CES question 3 (*Conversing with a familiar person over the telephone*). This interaction is illustrated in Figure 11. 4) There was no significant univariate effect of "Visit" $F(2,96) = 0.07$, $p = .934$, but there was a significant "Group" by "Visit" interaction $F(2,96) = 5.25$, $p = .007$ for CES question 4 (*Conversing with a stranger over the telephone*). This significant interaction is illustrated in Figure 12. These results indicated that the marginal mean of CES question 4 at visit 1 ($M = 3.14$, $SD = 0.71$) was similar to the marginal mean of the CES question 4 score at visit 2 ($M = 3.12$, $SD = 0.49$) and visit 3 ($M = 3.15$, $SD = 0.57$). 5) There was no significant univariate effect of "Visit" $F(2,96) = 0.22$, $p = .807$, but there was a significant "Group" by "Visit" interaction $F(2,96) = 10.17$, $p < .001$ for CES question 5 (*Being part of a conversation in a noisy environment (social gathering*). This significant interaction is illustrated in Figure 13. These results indicated that the marginal mean of the CES question 5 score at visit 1 ($M = 2.85$, $SD = 0.71$) was similar to the marginal mean of the CES question 5 score at visit 2 ($M = 2.82$, $SD = 0.57$) and visit 3 ($M = 2.89$, $SD = 0.71$). 6) There was no significant univariate effect of "Visit" $F(2,96) = 0.16$, $p = .850$, but there was a significant "Group" by "Visit" interaction $F(2,96) = 6.36$, $p = .003$ for CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*). This significant interaction is illustrated in Figure 14. These results indicated that the marginal mean of the CES question 6 score at visit 1 ($M = 3.03$, $SD = 0.71$) was similar to the marginal mean of the CES question 6 score at visit 2 ($M = 2.99$, $SD = 0.57$) and visit 3 ($M = 2.98$, $SD = 0.64$). 7) There was no significant univariate effect of "Visit" $F(2,96) = 0.49$, $p = .617$ or "Group" by "Visit" interaction $F(2,96) = 1.38$, $p = .256$ for CES question 7 (*Having a conversation while traveling in a car*). This interaction is illustrated in Figure 15. 8) There was no significant univariate effect of "Visit" $F(2,96) = 2.14$, $p = .123$, but there was a significant "Group" by "Visit" interaction $F(2,96) = 6.78$, $p = .002$ for CES question 8 (*Having a conversation with someone at a distance (across a room)*). This significant interaction is illustrated in Figure 16. These results indicated that the marginal mean of the CES question 8 score at visit 1 ($M = 2.82$, $SD = 0.71$) was similar to the marginal mean of the

CES question 8 score at visit 2 (*M* = 3.00, *SD* = 0.64) and visit 3 (*M* = 2.97, *SD* = 0.71). The results of univariate testing for the individual CES question scores are provided in Table 23.

Figure 9: Means of Communicative Effectiveness Survey Question 1 (*Having a Conversation with a Family Member or Friends at Home*) Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

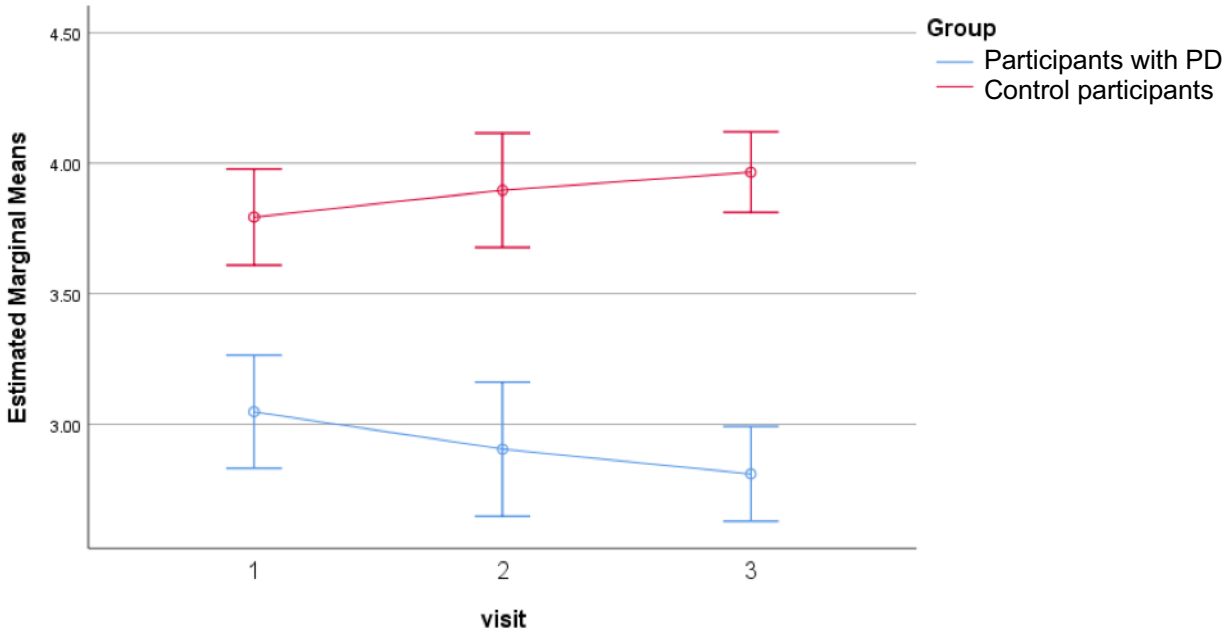

Figure 9. This figure demonstrates the changes in mean CES question 1 *(Having a conversation with a family member or friends at home)* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 10: Means of Communicative Effectiveness Survey Question 2 *(Participating in Conversation with Strangers in a Quiet Place)* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

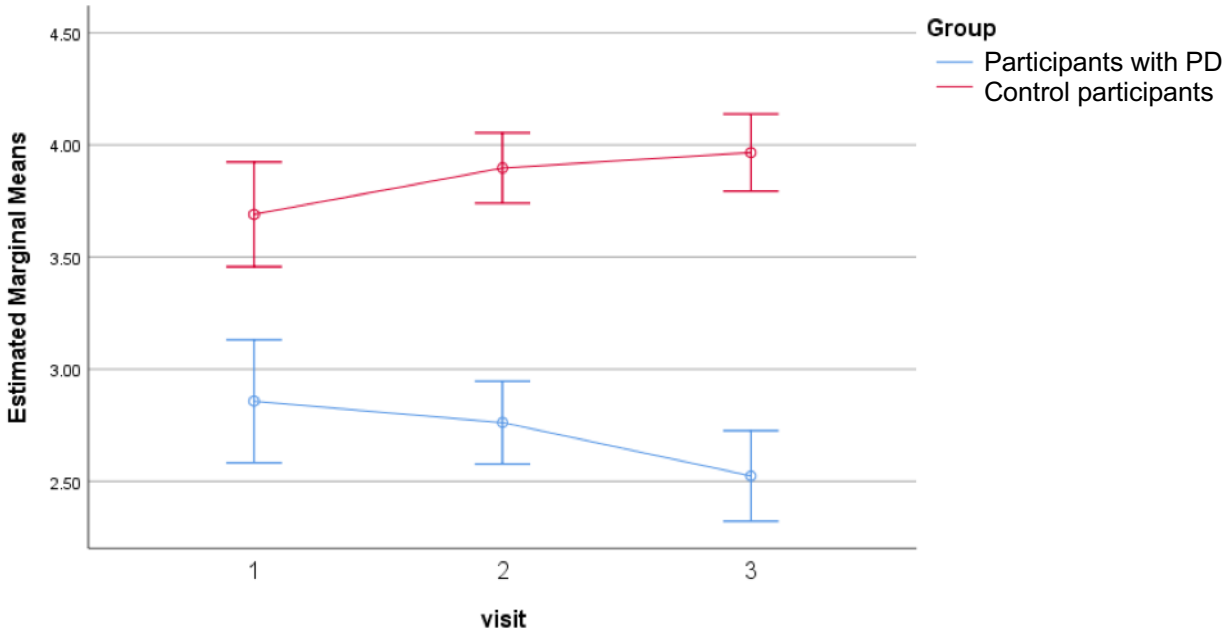

Figure 10. This figure demonstrates the changes in mean CES question 2 *(Participating in conversation with strangers in a quiet place)* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 11: Means of Communicative Effectiveness Survey Question 3 (*Conversing with a Familiar Person over the Telephone)* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

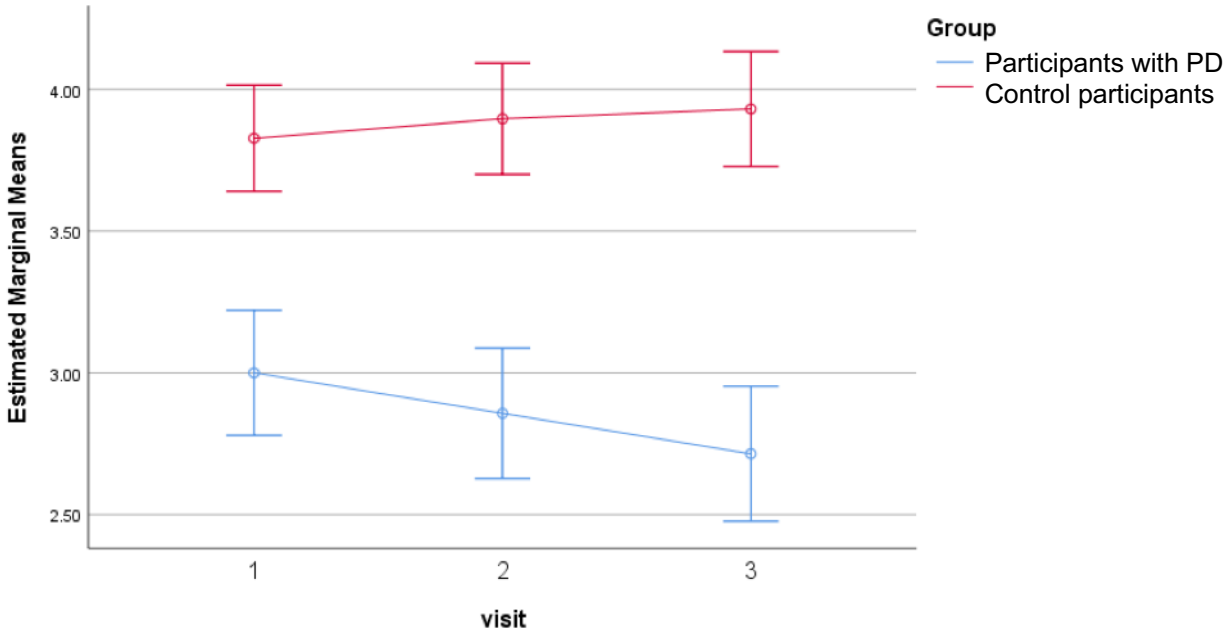

Figure 11. This figure demonstrates the changes in mean CES question 3 *(Conversing with a familiar person over the telephone)* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 12: Means of Communicative Effectiveness Survey Question 4 *(Conversing with a Stranger over the Telephone)* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

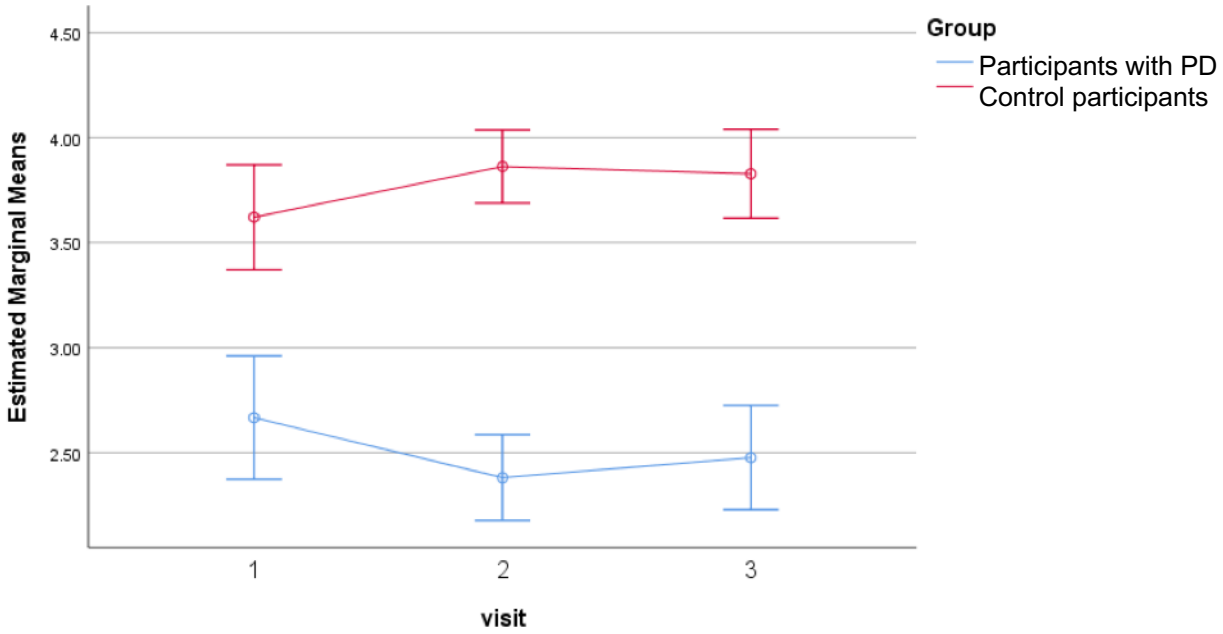

Figure 12. This figure demonstrates the changes in mean CES question 4 *(Conversing with a stranger over the telephone)* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 13: Means of Communicative Effectiveness Survey Question 5 *(Being Part of a Conversation in a Noisy Environment (Social Gathering))* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

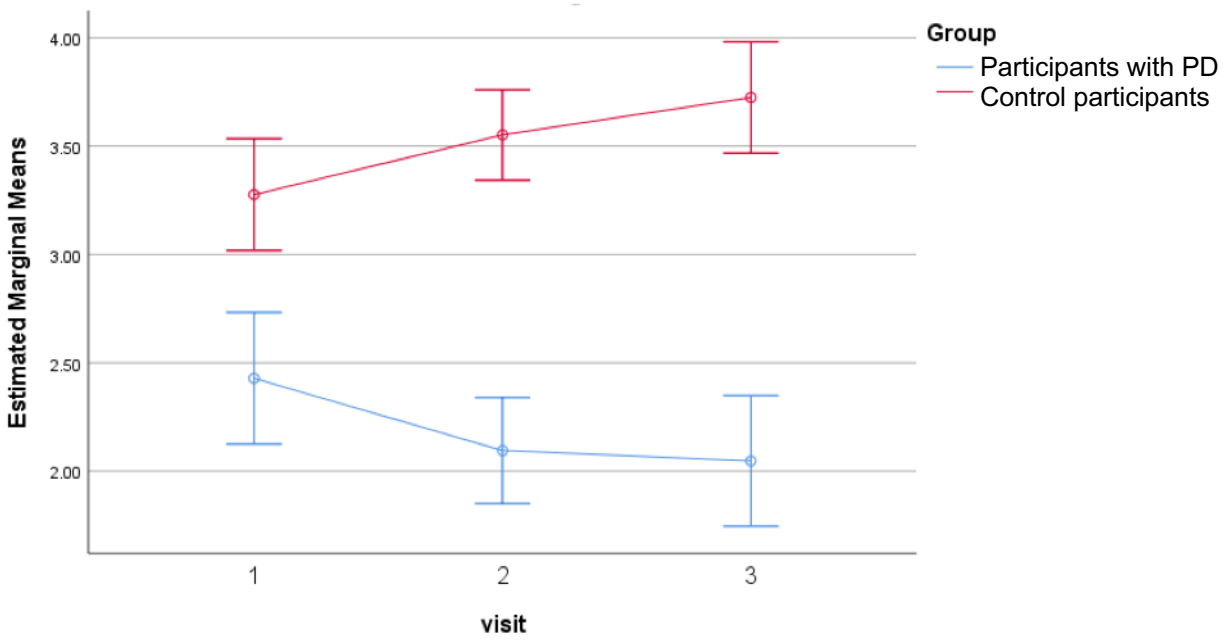

Figure 13. This figure demonstrates the changes in mean CES question 5 *(Being part of a conversation in a noisy environment (social gathering))* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 14: Means of Communicative Effectiveness Survey Question 6 *(Speaking to a Friend When You are Emotionally Upset or You are Angry)* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

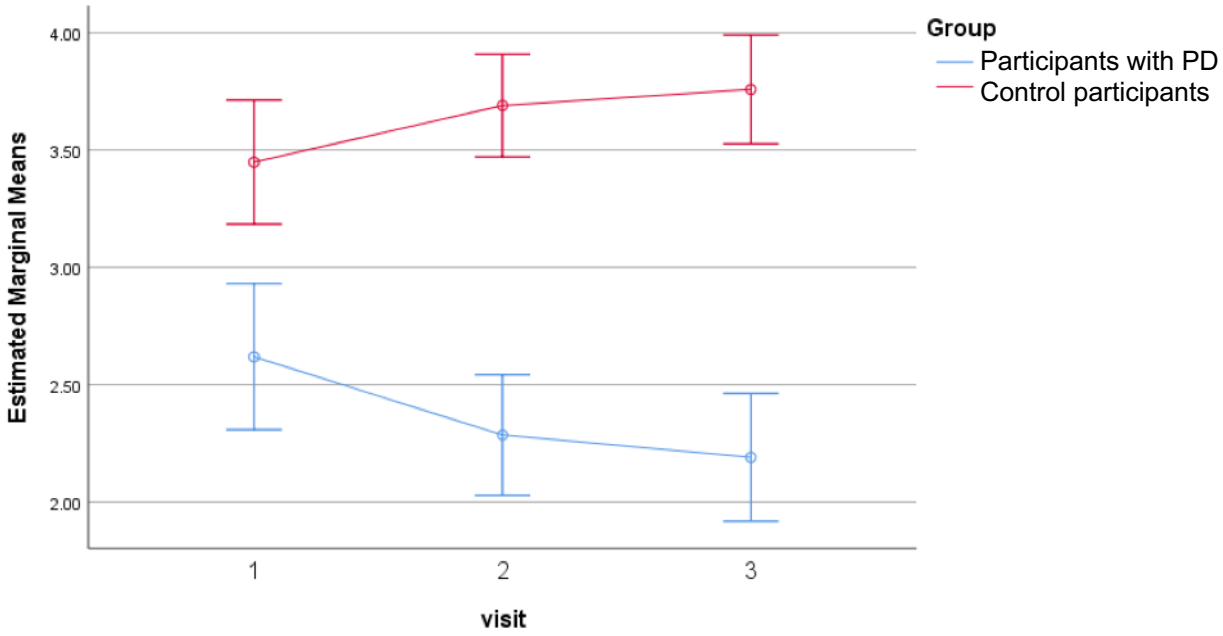

Figure 14. This figure demonstrates the changes in mean CES question 6 *(Speaking to a friend when you are emotionally upset or you are angry)* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 15: Means of Communicative Effectiveness Survey Question 7 *(Having a Conversation While Traveling in a Car)* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

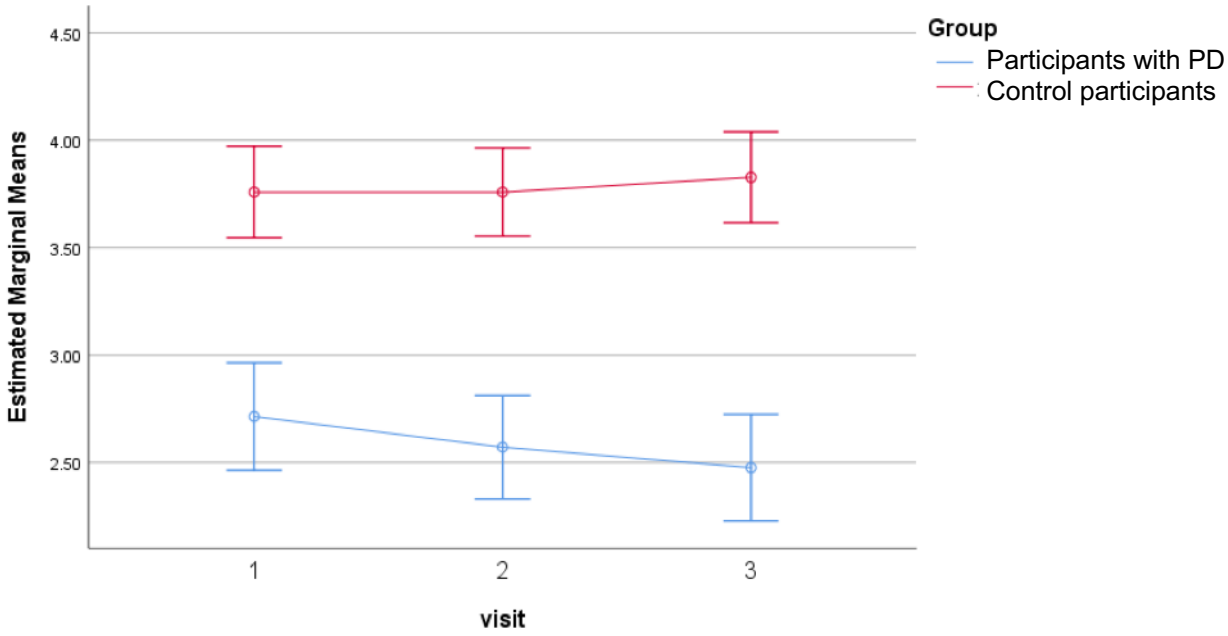

Figure 15. This figure demonstrates the changes in mean CES question 7 *(Having a conversation while traveling in a car)* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 16: Means of Communicative Effectiveness Survey Question 8 *(Having a Conversation with Someone at a Distance (Across a Room))* Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

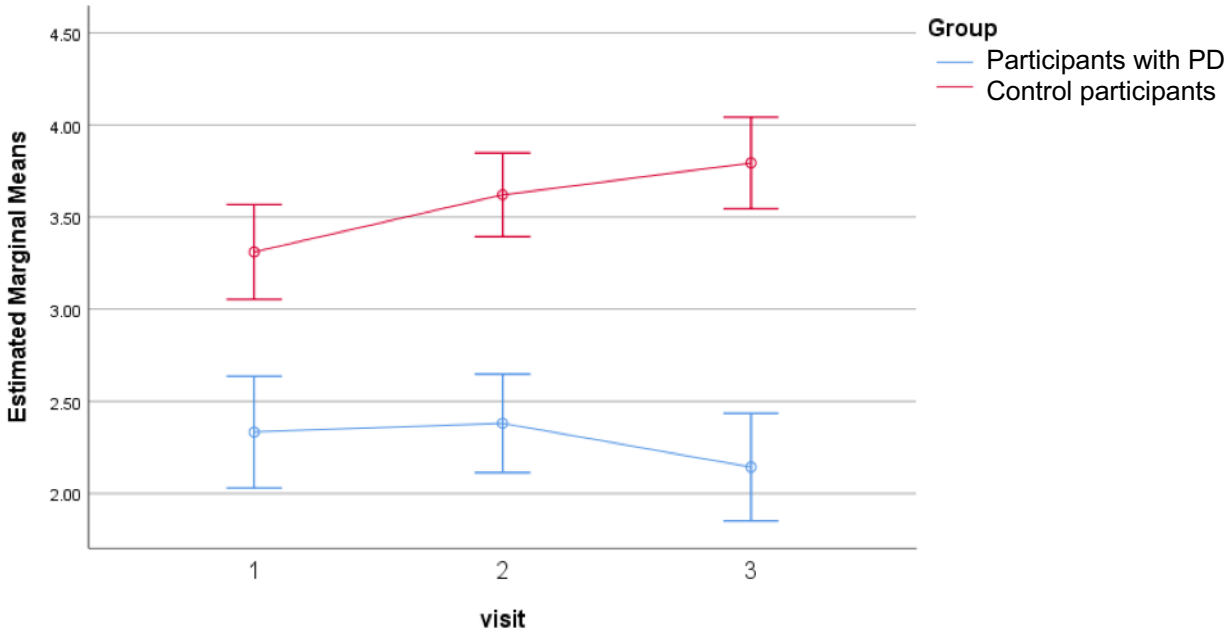

Figure 16. This figure demonstrates the changes in mean CES question 8 *(Having a conversation with someone at a distance (across a room))* scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Table 23: Univariate Testing Analysis of the Variability of Self-Rated Individual Questions of the Communicative Effectiveness Survey for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p value*
"Group"				
CES question 1	1	48	67.63	<.001
CES question 2	1	48	89.69	<.001
CES question 3	1	48	82.13	<.001
CES question 4	1	48	91.67	<.001
CES question 5	1	48	78.29	<.001
CES question 6	1	48	89.40	<.001
CES question 7	1	48	98.29	<.001
CES question 8	1	48	69.19	<.001
"Visit"				
CES question 1	2	96	0.10	.901
CES question 2	2	96	0.71	.496
CES question 3	2	96	0.58	.563
CES question 4	2	96	0.07	.934
CES question 5	2	96	0.22	.807
CES question 6	2	96	0.16	.850
CES question 7	2	96	0.49	.617
CES question 8	2	96	2.14	.123
"Group"*"Visit"				
CES question 1	2	96	4.06	.020
CES question 2	2	96	8.87	<.001
CES question 3	2	96	2.61	.078
CES question 4	2	96	5.25	.007
CES question 5	2	96	10.17	<.001
CES question 6	2	96	6.36	.003
CES question 7	2	96	1.38	.256
CES question 8	2	96	6.78	.002

Note. This table illustrates the significant and non-significant univariate effects of "Group", significant and non-significant univariate effects of "Visit", and "Group" by "Visit" interactions found from univariate analyses of the individual questions of the CES for participants with PD and control participants.

3.6.1.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of individual CES questions. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-

2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants.

3.6.1.1.1 Retest analysis of CES question 1 (*Having a conversation with a family member or friends at home*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 1 scores for participants with PD are summarized in Table 24. The ICC values related to the repeated measurement of CES question 1 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .70, *p* = .004, visit 1 versus visit 3 *ICC* = .56, *p* = .032, and visit 2 versus visit 3 *ICC* = .68, *p* = .008. These results suggest that CES question 1 scores as measured in the present study did not demonstrate good retest reliability in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

Table 24: Retest Analyses of Communicative Effectiveness Survey Question 1 (*Having a Conversation with a Family Member or Friends at Home*) for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.70[.29-.87]	0.13	0.69	0.90	.377	0.71	0.39	1.08	36.05
Visits 1 – 3	.56[-.03-.81]	0.18	0.66	1.28	.213	0.60	0.39	1.09	38.09
Visits 2 – 3	.68[.20-.87]	0.05	0.72	0.39	.771	0.71	0.40	1.11	39.37
Control participants									
Visits 1 – 2	.81[.60-.91]	-0.10	0.31	-1.80	.083	0.39	0.17	0.47	12.56
Visits 1 – 3	.35[-.23-.67]	-0.20	0.41	-2.69	.012	0.33	0.27	0.74	19.02
Visits 2 – 3	.55[.08-.78]	-0.10	0.31	-1.80	.083	0.28	0.19	0.52	13.03
Primary communication partners									
Visits 1 – 2	.79[.49-.91]	0.00	0.67	0.00	1.000	0.79	0.36	1.00	30.41
Visits 1 – 3	.46[-.35-.78]	-0.05	0.90	-0.24	.815	0.75	0.55	1.52	45.96
Visits 2 – 3	.52[-.18-.80]	-0.05	0.90	-0.24	.815	0.78	0.54	1.49	44.34

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 1 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 1 scores was 0.13 (*SD* = 0.69) and this difference was not significant $t(22) = 0.90$, $p = .377$. The mean

differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for CES question 1 scores was 0.18 (*SD* = 0.66) and this difference was not significant $t(21) = 1.28$, $p = .213$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 2 and visit 3 for CES question 1 scores was 0.05 (*SD* = 0.72) and this difference was not significant $t(21) = 0.30$, $p = .77$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.08 and *CR%* = 36.05%, for visit 1 vs visit 3, *CR* = 1.09 and *CR%* = 38.096%, and for visit 2 vs visit 3, *CR* = 1.11 and *CR%* = 39.37%. Based on the CR, an observed change in the CES question 1 scores of at least 1.08 – 1.11 would suggest a fairly large amount of measurement error ranging from 36.05 – 39.37% variation in the CES question 1 scores of IWPD. These results suggest that the measure of CES question 1 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.2 Retest analysis of CES question 1 (*Having a conversation with a family member or friends at home*) scores for control participants

Results for retest analyses of CES question 1 scores for control participants are summarized in Table 24. The ICC values related to the repeated measurement of CES question 1 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .81, $p < .001$, visit 1 versus visit 3 *ICC* = .35, $p = .095$, and visit 2 versus visit 3 *ICC* = .55, $p = .015$. These results suggest that CES question 1 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 1 scores was -0.10 (*SD* = 0.31) and this difference was not significant $t(29) = -1.80$, $p = .083$. The mean differences in retest values ranged from -1 – 0. The mean difference between visit 1 and visit 3 for CES question 1 scores was -0.20 (*SD* = 0.41) and this difference was significant $t(29) = -2.69$, $p = .012$. The mean difference between visit 2 and visit 3 for CES question 1 scores was -0.10 (*SD* = 0.31) and this difference was significant $t(29) = -$

1.80, $p = .083$. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.47 and *CR%* = 12.56%, for visit 1 vs visit 3, *CR* = 0.74 and *CR%* = 19.02%, and for visit 2 vs visit 3, *CR* = 0.52 and *CR%* = 13.03%. Based on the CR, an observed change in the CES question 1 scores of at least 0.47 – 0.74 would suggest the possibility of an acceptable amount of measurement error ranging from 12.56 – 19.02% variation in the CES question 1 scores of control participants. These results suggest that the measure of CES question 1 scores demonstrates marginal repeatability for control participants.

3.6.1.1.3 Retest analysis of CES question 2 (*Participating in conversation with strangers in a quiet place*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 2 scores for participants with PD are summarized in Table 25. The ICC values related to the repeated measurement of CES question 2 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .70, $p = .004$, visit 1 versus visit 3 *ICC* = .83, $p < .001$, and visit 2 versus visit 3 *ICC* = .65, $p = .007$. These results suggest that CES question 2 scores as measured in the present study did not demonstrate good retest reliability in IWPD because the ICC values across two of three comparisons were below our criterion of .75.

Table 25: Retest Analyses of Communicative Effectiveness Survey Question 2 (*Participating in Conversation with Strangers in a Quiet Place*) for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.70[.29-.88]	0.04	0.64	0.33	.747	0.66	0.36	1.01	35.53
Visits 1 – 3	.83[.47-.94]	0.32	0.48	3.13	.005	0.70	0.29	0.79	28.57
Visits 2 – 3	.65[.20-.85]	0.23	0.61	1.72	.096	0.64	0.38	1.04	41.69
Control participants									
Visits 1 – 2	.48[-.02-.74]	-0.23	0.50	-2.54	.017	0.45	0.32	0.89	24.30
Visits 1 – 3	.20[-.41-.58]	-0.30	0.53	-3.07	.005	0.41	0.37	1.01	26.00
Visits 2 – 3	.65[.28-.83]	-0.07	0.25	-1.44	.161	0.25	0.15	0.42	10.46
Primary communication partners									
Visits 1 – 2	.00[-1.15-.56]	0.30	0.88	1.67	.110	0.62	0.62	1.71	51.83
Visits 1 – 3	.33[-.47-.71]	0.32	0.84	1.78	.090	0.67	0.55	1.51	50.45
Visits 2 – 3	.72[.30-.88]	0.00	0.69	0.00	1.000	0.72	0.38	1.05	35.00

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 2 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 2 scores was 0.04 (*SD* = .64) and this difference was not significant $t(22) = 0.33$, $p = .747$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 1 and visit 3 for CES question 2 scores was 0.32 (*SD* = 0.48) and this difference was significant $t(21) = 3.13$, $p = .005$. The mean difference between visit 2 and visit 3 for CES question 2 scores was 0.23 (*SD* = 0.61) and this difference was not significant $t(21) = 1.74$, $p = .096$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.01 and *CR%* = 35.53%, for visit 1 vs visit 3, *CR* = 0.79 and *CR%* = 28.57%, and for visit 2 vs visit 3, *CR* = 1.04 and *CR%* = 41.69%. Based on the CR, an observed change in the CES question 2 scores of at least 0.79 – 1.04 would suggest a large amount of measurement error ranging from 28.57 – 41.69% variation in the CES question 2 scores of IWPD. These results suggest that the measure of CES question 2 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.4 Retest analysis of CES question 2 (*Participating in conversation with strangers in a quiet place*) scores for control participants

Results for retest analyses of CES question 2 scores for control participants are summarized in Table 25. The ICC values related to the repeated measurement of CES question 2 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .48, *p* = .026, visit 1 versus visit 3 *ICC* = .20, *p* = .229, and visit 2 versus visit 3 *ICC* = .65, *p* = .003. These results suggest that CES question 2 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 2 scores was -0.23 (*SD* = 0.50) and this difference was significant *t*(29) = -0.25, *p* = .017. The mean difference between visit 1 and visit 3 for CES question 2 scores was -0.30 (*SD* = 0.53) and this difference was significant *t*(29) = -3.07, *p* = .005. The mean difference between visit 2 and visit 3 for CES question 2 was -0.07 (*SD* = 0.25) and this difference was not significant *t*(29) = -1.44, *p* = .161. The mean differences in retest values ranged from -1 – 0. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.89 and *CR%* = 24.30%, for visit 1 vs visit 3, *CR* = 1.01 and *CR%* = 26.00%, and for visit 2 vs visit 3, *CR* = 0.42 and *CR%* = 10.46%. Based on the CR, an observed change in the CES question 2 scores of at least 0.42 – 1.32 would suggest a fairly large amount of measurement error ranging from 10.46 – 26.00% variation in the CES question 2 scores of control participants. These results suggest that the measure of CES question 2 scores demonstrates unacceptable repeatability for control participants.

3.6.1.1.5 Retest analysis of CES question 3 (*Conversing with a familiar person over the telephone*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 3 scores for participants with PD are summarized in Table 26. The ICC values related to the repeated measurement of CES

question 3 across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .48, *p* = .068, visit 1 versus visit 3 *ICC* = .32, *p* = .190, and visit 2 versus visit 3 *ICC* = .44, *p* = .100. These results suggest that CES question 3 scores as measured in the present study did not demonstrate good retest reliability y in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

Table 26: Retest Analyses of Communicative Effectiveness Survey Question 3 *(Conversing with a Familiar Person over the Telephone)* for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.48[-.22-.78]	0.18	0.80	1.07	.296	0.67	0.48	1.34	43.90
Visits 1 – 3	.32[-.58-.72]	0.29	0.90	1.45	.162	0.71	0.59	1.63	56.71
Visits 2 – 3	.44[-.37-.77]	0.14	0.89	0.72	.480	0.73	0.55	1.52	55.51
Control participants									
Visits 1 – 2	.77[.52-.89]	-0.10	0.31	-1.78	.083	0.36	0.17	0.48	12.71
Visits 1 – 3	.62[.23-.82]	-0.13	0.35	-2.11	.043	0.34	0.21	0.58	14.87
Visits 2 – 3	.88[.75-.94]	-0.03	0.18	-1.00	.326	0.28	0.10	0.27	6.88
Primary communication partners									
Visits 1 – 2	.33[-.64-.72]	0.04	0.98	0.21	.833	0.77	0.63	1.74	54.81
Visits 1 – 3	.37[-.58-.74]	0.00	0.87	0.0	1.000	0.69	0.55	1.52	48.51
Visits 2 – 3	.71[.29-.88]	-0.05	0.72	-0.30	.771	0.74	0.40	1.10	34.66

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 3 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 3 scores was 0.18 (*SD* = 0.80) and this difference was not significant *t*(21) = 1.07, *p* = .296. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 1 and visit 3 for CES question 3 scores was 0.29 (*SD* = 0.90) and this difference was not significant *t*(20) = 1.45, *p* = .162. The mean differences in retest values ranged from -1 – 3. The mean difference between visit 2 and visit 3 for CES question 3 scores was 0.14 (*SD* = 0.89) and this difference was not significant *t*(21) = 0.72, *p* = .480. The mean differences in retest values ranged from -1 – 2. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* =1.34 and

CR% = 43.90%, for visit 1 vs visit 3, *CR* = 1.63 and *CR%* = 56.71%, and for visit 2 vs visit 3, *CR* = 1.52 and *CR%* = 55.51%. Based on the CR, an observed change in the CES question 3 scores of at least 1.34 – 1.63 would suggest a large amount of measurement error ranging from 43.90 – 56.71% variation in the CES question 3 scores of IWPD. These results suggest that the measure of CES question 3 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.6 Retest analysis of CES question 3 (*Conversing with a familiar person over the telephone*) scores for control participants

Results for retest analyses of CES question 3 scores for control participants are summarized in Table 26. The ICC values related to the repeated measurement of CES question 3 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .77, $p < .001$, visit 1 versus visit 3 *ICC* = .62, $p = .003$, and visit 2 versus visit 3 *ICC* = .88, $p < .001$. These results suggest that CES question 3 scores as measured in the present study demonstrated good retest reliability in healthy speakers because the ICC values across two of three comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 3 scores was -0.10 (*SD* = 0.31) and this difference was not significant $t(29) = -1.80$, $p = .083$. The mean differences in retest values ranged from -1 – 0. The mean difference between visit 1 and visit 3 for CES question 3 scores was -0.13 (*SD* = 0.35) and this difference was significant $t(29) = -2.11$, $p = .043$. The mean difference between visit 2 and visit 3 for CES question 3 scores was -0.03 (*SD* = 0.18) and this difference was not significant $t(29) = -1.00$, $p = .326$. The mean differences in retest values ranged from -1 – 0. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.48 and *CR%* = 12.71%, for visit 1 vs visit 3, *CR* = 0.58 and *CR%* = 14.87%, and for visit 2 vs visit 3, *CR* = 0.27 and *CR%* = 6.88%. Based on the Cr, an observed change in the CES question 3 scores of at least 0.27 – 0.58 would suggest the possibility of an acceptable amount of measurement error ranging from 6.88 – 14.87% variation in the CES question 3 scores of control participants. These results suggest that

the measure of CES question 3 scores demonstrates marginal repeatability for control participants.

3.6.1.1.7 Retest analysis of CES question 4 (*Conversing with a stranger over the telephone*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 4 scores for participants with PD are summarized in Table 27. The ICC values related to the repeated measurement of CES question 4 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .64, *p* = .007, visit 1 versus visit 3 *ICC* = .70, *p* = .004, and visit 2 versus visit 3 *ICC* = .70, *p* = .005. These results suggest that CES question 4 scores as measured in the present study did not demonstrate good retest reliability in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

Table 27: Retest Analyses of Communicative Effectiveness Survey Question 4 *(Conversing with a Stranger over the Telephone)* for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.64[.19-.85]	0.26	0.69	1.82	.083	0.69	0.41	1.14	43.11
Visits 1 – 3	.70[.29-.87]	0.18	0.73	1.16	.257	0.76	0.42	1.15	48.26
Visits 2 – 3	.70[.26-.87]	-0.05	0.65	-0.33	.747	0.66	0.36	1.01	41.17
Control participants									
Visits 1 – 2	.33[-.27-.66]	-0.27	0.58	-2.50	.018	0.47	0.38	1.06	29.41
Visits 1 – 3	.43[-.12-.72]	-0.23	0.57	-2.25	.032	0.48	0.36	1.00	25.86
Visits 2 – 3	.53[-.01-.78]	0.03	0.41	0.44	.662	0.37	0.25	0.69	18.11
Primary communication partners									
Visits 1 – 2	.61[.12-.83]	0.30	0.8	1.67	.110	0.84	0.52	1.45	51.35
Visits 1 – 3	.55[-.10-.82]	0.09	0.97	0.44	.665	0.86	0.57	1.59	63.07
Visits 2 – 3	.81[.55-.92]	-0.23	0.69	-1.556	.135	0.87	0.38	1.04	38.26

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 4 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 4 scores was 0.26 (*SD* = 0.69) and this difference was not significant $t(22) = 1.82$, $p = .083$. The mean

differences in retest values ranged from -1 – 1. The mean difference between visit 1 and visit 3 for CES question 4 scores was 0.18 (*SD* = 0.73) and this difference was not significant $t(21) = 1.16$, $p = .257$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 2 and visit 3 for CES question 4 scores was -0.05 (*SD* = .65) and this difference was not significant $t(21) = -0.33$, $p = .747$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.14 and *CR%* = 43.11%, for visit 1 vs visit 3, *CR* = 1.15 and *CR%* = 48.26%, and for visit 2 vs visit 3, *CR* = 1.01 and *CR%* = 41.17%. Based on the CR, an observed change in the CES question 4 scores of at least 1.01 – 1.15 would suggest a fairly large amount of measurement error ranging from 41.17 – 48.26% variation in the CES question 4 scores of IWPD. These results suggest that the measure of CES question 4 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.8 Retest analysis of CES question 4 (*Conversing with a stranger over the telephone*) scores for control participants

Results for retest analyses of CES question 4 scores for control participants are summarized in Table 27. The ICC values related to the repeated measurement of CES question 4 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .33, $p = .116$, visit 1 versus visit 3 *ICC* = .43, $p = .051$, visit 2 versus visit 3 *ICC* = .53, $p = .027$. These results suggest that CES question 4 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 4 scores was -0.27 (*SD* = 0.58) and this difference was significant $t(29) = -2.50$, $p = .018$. The mean difference between visit 1 and visit 3 for CES question 4 scores was -0.23 (*SD* = 0.57) and this difference was significant $t(29) = -2.25$, $p = .032$. The mean difference between visit 2 and visit 3 for CES question 4 scores was 0.03 (*SD* = 0.41) and this difference was not significant $t(29) = 0.44$, $p = .662$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit

comparisons: for visit 1 vs visit 2, *CR* = 1.06 and *CR%* = 29.41%, for visit 1 vs visit 3, *CR* = 1.00 and *CR%* = 25.86%, and for visit 2 vs visit 3, *CR* = 0.69 and *CR%* = 18.11%. Based on the CR, an observed change in the CES question 4 scores of at least 0.69 – 1.06 would suggest a fairly large amount of measurement error ranging from 18.11 – 29.41% variation in the CES question 4 scores of control participants. These results suggest that the measure of CES question 4 scores demonstrates unacceptable repeatability for control participants.

3.6.1.1.9 Retest analysis of CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 5 scores for participants with PD are summarized in Table 28. The ICC values related to the repeated measurement of CES question 5 across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .70, *p* = .002, visit 1 versus visit 3 *ICC* = .28, *p* = .221, and visit 2 versus visit 3 *ICC* = .63, *p* = .017. These results suggest that CES question 5 scores as measured in the present study did not demonstrate good retest reliability in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

Table 28: Retest Analyses of Communicative Effectiveness Survey Question 5 *(Being Part of a Conversation in a Noisy Environment (Social Gathering))* for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.70[.32-.87]	0.26	0.62	2.02	.056	0.67	0.37	1.01	44.03
Visits 1 – 3	.28[-.67-.60]	0.27	1.03	1.24	.229	0.82	0.69	1.92	94.12
Visits 2 – 3	.63[.08-.85]	0.00	0.76	0.00	1.000	0.73	0.45	1.23	58.98
Control participants									
Visits 1 – 2	.70[.36-.86]	-0.27	0.58	-2.50	.018	0.63	0.35	0.96	29.36
Visits 1 – 3	.57[.05-.80]	-0.43	0.63	-3.79	.001	0.62	0.40	1.12	31.66
Visits 2 – 3	.69[.36-.85]	-0.17	0.53	-1.72	.096	0.55	0.31	0.85	22.94
Primary communication partners									
Visits 1 – 2	.62[.13-.84]	0.22	0.90	1.16	.260	0.86	0.53	1.47	55.43

Visits 1 – 3	.75[.40-.89]	0.32	0.78	1.91	.069	0.89	0.45	1.23	50.73
Visits 2 – 3	.80[.52-.92]	0.05	0.72	0.30	.771	0.87	0.39	1.08	45.69

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 5 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 5 scores was 0.26 (*SD* = 0.62) and this difference approached significance $t(22) = 2.02$, $p = .056$. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for CES question 5 scores was 0.27 (*SD* = 1.03) and this difference was not significant $t(21) = 1.24$, $p = .229$. The mean differences in retest values ranged from -2 – 3. The mean difference between visit 2 and visit 3 for CES question 5 scores was 0.00 (*SD* = 0.76) and this difference was not significant $t(21) = 0.00$, $p = 1.000$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.01 and *CR%* = 44.03%, for visit 1 vs visit 3, *CR* = 1.92 and *CR%* = 94.12%, and for visit 2 vs visit 3, *CR* = 1.23 and *CR%* = 58.98%. Based on the CR, an observed change in the CES question 5 scores of at least 1.01 – 1.92 would suggest a large amount of measurement error ranging from 44.03 – 94.12% variation in the CES question 5 scores of IWPD. These results suggest that the measure of CES question 5 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.10 Retest analysis of CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) scores for control participants

Results for retest analyses of CES question 5 scores for control participants are summarized in Table 28. The ICC values related to the repeated measurement of CES question 5 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .70, $p < .001$, visit 1 versus visit 3 *ICC* = .57, $p = .003$, and visit 2 versus visit 3 *ICC* = .69, $p = .001$. These results suggest that CES question 5 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 5 scores was -0.27 (*SD* = 0.58) and this difference was significant $t(29) = -2.50$, $p = .018$. The mean difference between visit 1 and visit 3 for CES question 5 scores was -0.43 (*SD* = 0.63) and this difference was significant $t(29) = -3.79$, $p = .001$. The mean difference between visit 2 and visit 3 for CES question 5 scores was -0.17 (*SD* = 0.53) and this difference was not significant $t(29) = -1.72$, $p = .096$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2. *CR* = 0.96 and *CR%* = 29.36%, for visit 1 vs visit 3, *CR* = 1.12 and *CR%* = 31.66%, and for visit 2 vs visit 3, *CR* = 0.85 and *CR%* = 22.94%. Based on the Cr, an observed change in the CES question 5 scores of at least 0.85 – 1.12 would suggest a fairly large amount of measurement error ranging from 22.94 – 31.66% variation in the CES question 5 scores of control participants. These results suggest that the measure of CES question 5 scores demonstrates unacceptable repeatability for control participants.

3.6.1.1.11 Retest analysis of CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 6 scores for participants with PD are summarized in Table 29. The ICC values related to the repeated measurement of CES question 6 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .46, $p = .074$, visit 1 versus visit 3 *ICC* = .03, $p = .469$, visit 2 versus visit 3 *ICC* = .50, $p = .064$. These results suggest that CES question 6 scores as measured in the present study did not demonstrate good retest reliability in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

Table 29: Retest Analyses of Communicative Effectiveness Survey Question 6 (*Speaking to a Friend When You are Emotionally Upset or You are Angry*) for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.46[-.23-.77]	0.26	0.92	1.37	.186	0.78	0.57	1.58	61.53
Visits 1 – 3	.03[-1.25-.59]	0.32	1.21	1.23	.231	0.86	0.84	2.34	101.56
Visits 2 – 3	.50[-.23-.80]	0.05	0.84	0.25	.803	0.73	0.51	1.42	63.73
Control participants									
Visits 1 – 2	.36[-.28-.69]	-0.37	0.96	-2.08	.046	0.55	0.44	1.22	35.35
Visits 1 – 3	.41[-.15-.71]	-0.43	0.90	-2.64	.013	0.50	0.39	1.07	29.03
Visits 2 – 3	.84[.66-.92]	-0.07	0.37	-1.00	.326	0.48	0.19	0.53	14.18
Primary communication partners									
Visits 1 – 2	.63[.13-.84]	0.45	0.86	2.49	.021	0.87	0.53	1.46	48.58
Visits 1 – 3	.50[-.16-.79]	0.33	0.97	1.58	.130	0.84	0.59	1.64	64.19
Visits 2 – 3	.65[.14-.86]	-0.14	0.85	-0.77	.452	0.83	0.49	1.36	50.98

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 6 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 6 scores was 0.26 (*SD* = 0.92) and this difference was not significant $t(22) = 1.37$, $p = .186$. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for CES question 6 scores was 0.32 (*SD* = 1.21) and this difference was not significant $t(21) = 1.23$, $p = .231$. The mean differences in retest values ranged from -2 – 3. The mean difference between visit 2 and visit 3 for CES question 6 scores was 0.05 (*SD* = 0.84) and this difference was not significant $t(21) = 0.25$, $p = .803$ The mean differences in retest values ranged from -2 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.58 and *CR%* = 61.53%, for visit 1 vs visit 3, *CR* = 2.34 and *CR%* = 101.56%, and for visit 2 vs visit 3, *CR* = 1.42 and *CR%* = 63.73%. Based on the CR, an observed change in the CES question 6 scores of at least 1.42 – 2.34 would suggest a large amount of measurement error ranging from 61.53 – 101.56% variation in the CES question 6 scores of IWPD.

These results suggest that the measure of CES question 6 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.12 Retest analysis of CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) scores for control participants

Results for retest analyses of CES question 6 scores for control participants are summarized in Table 29. The ICC values related to the repeated measurement of CES question 6 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .36$, $p = .107$, visit 1 versus visit 3 $ICC = .41$, $p = .057$, visit 2 versus visit 3 $ICC = .84$, $p < .001$. These results suggest that CES question 6 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 6 scores was -0.37 ($SD = 0.96$) and this difference was significant $t(29) = -2.08$, $p = .046$. The mean difference between visit 1 and visit 3 for CES question 6 scores was -0.43 ($SD = 0.90$) and this difference was significant $t(29) = -2.64$, $p = .013$. The mean difference between visit 2 and visit 3 for CES question 6 scores was -0.07 ($SD = 0.37$) and this difference was not significant $t(29) = -1.00$, $p = .326$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 1.22$ and $CR\% = 35.35\%$, for visit 1 vs visit 3, $CR = 1.07$ and $CR\% = 29.03\%$, and for visit 2 vs visit 3, $CR = 0.53$ and $CR\% = 14.18\%$. Based on the CR, an observed change in the CES question 6 scores of at least 0.53 – 1.22 would suggest a fairly large amount of measurement error ranging from 14.18 – 35.35% variation in the CES question 6 scores of control participants. These results suggest that the measure of CES question 6 scores demonstrates unacceptable repeatability for control participants.

3.6.1.1.13 Retest analysis of CES question 7 (*Having a conversation while travelling in a car*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 7 scores for participants with PD are summarized in Table 30. The ICC values related to the repeated measurement of CES question 7 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .37, *p* = .148, visit 1 versus visit 3 *ICC* = .57, *p* = .025, visit 2 versus visit 3 *ICC* = .21, *p* = .299. These results suggest that CES question 7 scores as measured in the present study did not demonstrate good retest reliability in IWPD because all of the ICC values across all comparisons were below our criterion of .75.

Table 30: Retest Analyses of Communicative Effectiveness Survey Question 7 (*Having a Conversation While Traveling in a Car*) for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.37[-.52-.74]	0.09	0.85	0.49	.628	0.68	0.54	1.50	55.40
Visits 1 – 3	.57[.02-.82]	0.23	0.75	1.42	.171	0.69	0.45	1.24	47.68
Visits 2 – 3	.21[-.94-.68]	0.14	0.89	0.72	.480	0.67	0.59	1.64	66.83
Control participants									
Visits 1 – 2	.80[.58-.91]	0.00	0.37	0.00	1.000	0.45	0.20	0.56	14.95
Visits 1 – 3	.55[.06-.79]	-0.07	0.52	-0.70	.489	0.47	0.31	0.86	23.18
Visits 2 – 3	.70[.37-.86]	-0.07	0.45	-0.81	.423	0.47	0.25	0.71	18.58
Primary communication partners									
Visits 1 – 2	.34[-.59-.72]	0.13	1.06	0.59	.560	0.84	0.68	1.89	62.15
Visits 1 – 3	.50[-.26-.79]	0.00	0.76	0.00	1.000	0.66	0.47	1.29	44.44
Visits 2 – 3	.43[-.41-.76]	-0.14	1.04	-0.62	.544	0.86	0.65	1.79	57.87

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 7 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 7 scores was 0.09 (*SD* = 0.85) and this difference was not significant $t(22) = 0.49$, $p = .628$. The mean differences in retest values ranged from -1 – 3. The mean difference between visit 1 and visit 3 for CES question 7 scores was 0.23 (*SD* = 0.75) and this difference was not significant $t(21) = 1.42$, $p = .171$. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 2 and visit 3 for CES question 7 scores was 0.14 (*SD* = 0.89) and this difference was not significant $t(21) = 0.72$, $p = .480$. The mean

differences in retest values ranged from -1 – 2. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.50 and *CR%* = 55.40%, for visit 1 vs visit 3, *CR* = 1.24 and *CR%* = 47.68%, and for visit 2 vs visit 3, *CR* = 1.64 and *CR%* = 66.83%. Based on the CR, an observed change in the CES question 7 scores of at least 1.24 – 1.64 would suggest a large amount of measurement error ranging from 47.68 – 66.83% variation in the CES question 7 scores of IWPD. These results suggest that the measure of CES question 7 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.14 Retest analysis of CES question 7 (*Having a conversation while traveling in a car*) scores for control participants

Results for retest analyses of CES question 7 scores for control participants are summarized in Table 30. The ICC values related to the repeated measurement of CES question 7 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .80, $p < .001$, visit 1 versus visit 3 *ICC* = .55, $p = .018$, and visit 2 versus visit 3 *ICC* = .70, $p = .001$. These results suggest that CES question 7 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 7 scores was 0.00 (*SD* = 0.37) and this difference was not significant $t(29) = 0.00$, $p = 1.000$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 1 and visit 3 for CES question 7 scores was -0.07 (*SD* = 0.52) and this difference was not significant $t(29) = -0.70$, $p = .489$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 2 and visit 3 for CES question 7 scores was -0.07 (*SD* = 0.45) and this difference was not significant $t(29) = -0.81$, $p = .423$. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.56 and *CR%* = 14.95%, for visit 1 vs visit 3, *CR* = 0.86 and *CR%* = 23.18%, and for visit 2 vs visit 3, *CR* = 0.71 and *CR%* = 18.58%. Based on the CR, an observed change in the CES

question 7 scores of at least 0.56 – 0.86 would suggest the possibility of an acceptable amount of measurement error ranging from 14.95 – 23.18% variation in the CES question 7 scores of control participants. These results suggest that the measure of CES question 7 scores demonstrates marginal repeatability for control participants.

3.6.1.1.15 Retest analysis of CES question 8 (*Having a conversation with someone at a distance (across a room)*) scores for participants with Parkinson's disease

Results for retest analyses of CES question 8 scores for participants with PD are summarized in Table 31. The ICC values related to the repeated measurement of CES question 8 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .77, *p* = .001, visit 1 versus visit 3 *ICC* = .68, *p* = .006, and visit 2 versus visit 3 *ICC* = .59, *p* = .023. These results suggest that CES question 8 scores as measured by the present study did not demonstrate good retest reliability in IWPD because the ICC values across two of three comparisons were below our criterion of .75.

Table 31: Retest Analyses of Communicative Effectiveness Survey Question 8 (*Having a Conversation with Someone at a Distance (Across a Room)*) for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.77[.45-.90]	-0.04	0.56	-0.37	.714	0.64	0.31	0.85	36.97
Visits 1 – 3	.68[.24-.87]	0.14	0.77	0.83	.418	0.78	0.44	1.23	52.18
Visits 2 – 3	.59[.03-.83]	0.18	0.85	1.00	.329	0.79	0.51	1.40	64.34
Control participants									
Visits 1 – 2	.67[.31-.85]	-0.30	0.65	-2.52	.017	0.69	0.39	1.09	33.44
Visits 1 – 3	.59[-.03-.82]	-0.50	0.57	-4.79	<.001	0.61	0.39	1.07	30.07
Visits 2 – 3	.72[.41-.87]	-0.20	0.48	-2.26	.031	0.54	0.29	0.79	20.97
Primary communication partners									
Visits 1 – 2	.65[.18-.85]	0.17	0.8	0.94	.357	0.87	0.52	1.43	55.51
Visits 1 – 3	.65[.18-.85]	0.27	0.94	1.37	.186	0.93	0.55	1.52	63.45
Visits 2 – 3	.80[.51-.92]	0.09	0.75	0.57	.576	0.90	0.40	1.11	47.88

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CES question 8 scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CES question 8 scores was -0.04 (*SD* = 0.56) and this difference was not significant $t(22) = -0.37$, $p = .714$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 1 and visit 3 for CES question 8 scores was 0.14 (*SD* = 0.77) and this difference was not significant $t(22) = 0.83$, $p = .418$. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 2 and visit 3 for CES question 8 scores was 0.18 (*SD* = 0.85) and this difference was not significant $t(21) = 1.00$, $p = .329$. The mean differences in retest values ranged from -1 – 2. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.85 and *CR%* = 36.97%, for visit 1 vs visit 3, *CR* = 1.23 and *CR%* = 52.18%, and for visit 2 vs visit 3, *CR* = 1.40 and *CR%* = 64.34%. Based on the CR, an observed change in the CES question 8 scores of at least 0.85 – 1.40 would suggest a fairly large amount of measurement error ranging from 36.97 – 64.34% variation in the CES question 8 scores of IWPD. These results suggest that the measure of CES question 8 scores demonstrates unacceptable repeatability for IWPD.

3.6.1.1.16 Retest analysis of CES question 8 (*Having a conversation with someone at a distance (across a room)*) scores for control participants

Results for retest analyses of CES question 8 scores for control participants are summarized in Table 31. The ICC values related to the repeated measurement of CES question 8 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .67, $p = .001$, visit 1 versus visit 3 *ICC* = .59, $p = .001$, and visit 2 versus visit 3 *ICC* = .72, $p < .001$. These results suggest that CES question 8 scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 8 scores was -0.30 (*SD* = 0.65) and this difference was significant $t(29) = -2.52$, $p = .017$. The mean difference between visit 1 and visit 3 for CES question 8 scores was -0.50 (*SD* = 0.57) and this difference was significant $t(29) = -4.79$, $p < .001$. The mean difference between visit 2 and visit 3 for CES question 8 scores was -0.20 (*SD* = 0.48) and this difference was significant $t(29) = -2.26$, $p = .031$. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.09 and *CR%* = 33.44%, for visit 1 vs visit 3, *CR* = 1.07 and *CR%* = 30.07%, and for visit 2 vs visit 3, *CR* = 0.79 and *CR%* = 20.97%. Based on the CR, an observed change in the CES question 8 scores of at least 0.79 – 1.09 would suggest a fairly large amount of measurement error ranging from 20.97 – 30.07% variation in the CES question 8 scores of control participants. These results suggest that the measure of CES question 8 scores demonstrates unacceptable repeatability for control participants.

3.6.2 VAPP subsection scores

Scores for the VAPP *self-perceived voice problem, daily communication, social communication*, and *emotion* subsections were obtained by summing up the converted scores for each question within a subsection. Raw scores for each question were obtained by measuring the location of an individual's response of each VAS. These raw scores were then divided by a factor of 10. VAPP *total* scores were obtained by summing up the total scores of each of the above-mentioned subsections. VAPP *activity limitation* scores were obtained by summing up the converted scores from the first question of each described situation assessing the degree of activity limitation from the *daily communication* and *social communication* subsections. VAPP *participation restriction* scores were obtained by summing up the converted scores from the second question of each described situation assessing the degree of participation restriction from the *daily communication* and *social communication* subsection. Descriptive statistics for VAPP subsection scores can be found in Table 32. The results of the repeated measures MANOVA for the dependent variables of the individual VAPP subsections showed that there was a significant multivariate effect of "Group" $F(7,43) = 14.56$, $p < .001$. The results of this repeated measures MANOVA can be found in Table 33. The results of

subsequent univariate testing for each of the individual VAPP subsections revealed the following results: 1) there was a significant univariate effect of "Group" $F(1,49) = 71.94$, $p < .001$ for the VAPP *voice problem* score with participants with PD having a greater (+3.04) marginal mean ($M = 3.29$, $SD = 1.28$) compared to control participants ($M = 0.25$, $SD = 1.26$); 2) there was a significant univariate effect of "Group" $F(1,49) = 66.31$, $p < .001$ for the VAPP *daily communication* score with participants with PD having a greater (+34.84) marginal mean ($M = 39.92$, $SD = 15.03$) compared to the control participants ($M = 5.08$, $SD = 15.06$); 3) there was a significant univariate effect of "Group" $F(1,49) = 51.45$, $p < .001$ for the VAPP *social communication* score with participants with PD having a greater (+9.56) marginal mean ($M = 10.86$, $SD = 4.67$) compared to control participants ($M = 1.30$, $SD = 4.71$); 4) there was a significant univariate effect of "Group" $F(1,49) = 46.53$, $p < .001$ for the VAPP *emotion* score with participants with PD having a greater (+17.07) marginal mean ($M = 19.38$, $SD = 8.80$) compared to control participants ($M = 2.31$, $SD = 8.82$); 5) there was a significant univariate effect of "Group" $F(1,49) = 70.39$, $p < .001$ for the VAPP *activity limitation* scores with participants with PD having a greater (+25.69) marginal mean ($M = 29.44$, $SD = 10.77$) compared to control participants ($M = 3.75$, $SD = 10.79$); 6) there was a significant univariate effect of "Group" $F(1,49) = 44.57$, $p < .001$ for the VAPP *participation restriction* scores with participants with PD having a greater (+21.02) marginal mean ($M = 23.81$, $SD = 11.04$) compared to control participants ($M = 2.79$, $SD = 11.06$); and 7) there was a significant univariate effect of "Group" $F(1,49) = 56.84$, $p < .001$ for the VAPP *total* score with participants with PD having a greater (+66.74) marginal mean ($M = 75.92$, $SD = 31.12$) compared to control participants ($M = 9.18$, $SD = 31.11$).

Table 32: Descriptive Statistics of Self- and Proxy-Rated Individual Subsections of the Voice Activity and Participation Profile for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	Voice problem mean*(SD)*	Daily comm. mean*(SD)*	Social comm. mean*(SD)*	Emotion mean*(SD)*	Activity limitation mean*(SD)*	Participation restriction mean*(SD)*	Total mean*(SD)*
Participants with PD							
Visit 1	3.43*(2.43)*	38.18*(24.93)*	10.65*(7.07)*	20.35*(15.52)*	29.01*(17.36)*	22.20*(17.73)*	74.99*(51.49)*
Visit 2	3.41*(2.10)*	39.05*(20.55)*	9.23*(6.87)*	17.77*(13.76)*	28.43*(15.05)*	21.97*(15.03)*	71.43*(43.26)*
Visit 3	3.25*(2.63)*	45.60*(30.35)*	12.85*(9.61)*	22.39*(17.89)*	32.48*(21.17)*	28.31*(21.77)*	86.43*(61.52)*
Control participants							
Visit 1	0.24*(0.30)*	5.08*(4.80)*	1.26*(1.37)*	2.39*(2.60)*	3.90*(3.83)*	2.45*(2.67)*	9.21*(8.76)*
Visit 2	0.21*(0.25)*	5.10*(4.89)*	1.16*(0.98)*	2.17*(1.80)*	3.66*(3.46)*	2.80*(2.73)*	8.84*(7.64)*
Visit 3	0.31*(0.40)*	5.06*(4.50)*	1.48*(1.48)*	2.36*(2.26)*	3.68*(3.40)*	3.14*(2.97)*	9.49*(8.40)*
Primary communication partners							
Visit 1	2.97*(2.28)*	33.63*(27.07)*	7.83*(7.05)*	15.51*(12.61)*	24.34*(19.33)*	18.86*(17.60)*	61.69*(47.73)*
Visit 2	3.38*(2.44)*	39.34*(28.56)*	9.64*(7.62)*	19.01*(14.81)*	27.89*(20.06)*	22.83*(19.15)*	73.10*(53.42)*
Visit 3	2.81*(2.20)*	39.14*(25.61)*	9.65*(7.29)*	17.57*(13.38)*	29.10*(20.47)*	21.26*(15.01)*	70.75*(47.57)*
Marginal means							
Participants with PD	3.29*(1.28)*	39.92*(15.03)*	10.86*(4.67)*	19.38*(8.80)*	29.44*(10.77)*	23.81*(11.04)*	75.92*(31.12)*
Control participants	0.25*(1.26)*	5.08*(15.06)*	1.30*(4.71)*	2.31*(8.82)*	3.75*(10.79)*	2.79*(11.06)*	9.18*(31.11)*
Primary communication partners	3.10*(2.02)*	37.59*(24.58)*	9.08*(7.13)*	17.55*(13.04)*	27.33*(18.11)*	21.08*(16.60)*	69.06*(47.80)*
PD and control visit 1	1.88*(1.57)*	21.53*(17.28)*	5.99*(4.93)*	11.27*(10.57)*	16.38*(12.14)*	12.45*(12.14)*	42.10*(35.35)*
PD and control visit 2	1.75*(1.36)*	21.70*(14.28)*	5.37*(4.64)*	9.59*(9.00)*	15.93*(10.50)*	12.41*(10.43)*	39.68*(29.71)*
PD and control visit 3	1.69*(1.71)*	24.27*(19.42)*	6.89*(6.28)*	11.67*(11.21)*	17.48*(13.85)*	15.04*(14.00)*	45.88*(39.35)*
PD and partners visit 1	3.29*(2.36)*	36.06*(26.82)*	9.33*(7.28)*	18.00*(14.43)*	26.82*(18.95)*	20.79*(18.16)*	68.90*(51.02)*
PD and partners visit 2	3.35*(2.30)*	38.91*(25.64)*	9.62*(7.48)*	18.13*(14.43)*	28.15*(18.30)*	22.44*(17.84)*	72.07*(50.10)*
PD and partners visit 3	2.94*(2.36)*	41.30*(27.54)*	10.97*(8.46)*	19.27*(15.28)*	30.19*(20.72)*	24.10*(18.36)*	76.51*(53.90)*

Note. This table illustrates the means and standard deviations for the individual subsection scores of the VAPP for participants with PD, their primary communication partners, and control participants across visits.

Table 33: Multivariate Testing Analysis of Self-Rated Individual Subsections of the Voice Activity and Participation Profile for Participants with Parkinson's Disease and Control Participants for Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p value*
"Group"	7	43	14.56	<.001
"Visit"	13	37	2.60	.011
"Group"*"Visit"	13	37	2.75	.008

Note. This table illustrates the main effects of "Group" and "Visit", and the "Group" by "Visit" interaction found from multivariate analyses of the individual questions of the VAPP for participants with PD and control participants.

The results of the repeated measures MANOVA for the dependent variables of the individual VAPP subsections showed that there was a significant multivariate effect of "Visit" $F(13,37) = 2.60$, $p = .011$. The results of this repeated measures MANOVA can be found in Table 33. A closer look at the factor "Visit", using subsequent univariate analyses revealed the following results: 1) there was no significant univariate effect of "Visit" $F(2,98) = 0.40$, $p = .674$ for the VAPP *self-perceived voice problem* score; 2) there was no significant univariate effect of "Visit" $F(2,98) = 1.35$, $p = .264$ for the VAPP *daily communication* score; 3) there was a significant univariate effect of "Visit" $F(2,98) = 3.51$, $p = .034$ for the VAPP *social communication* score which indicated that the marginal mean of VAPP *social communication* score at visit 1 ($M = 5.99$, $SD = 4.93$) was greater (+0.62) than the marginal mean of VAPP *social communication* score at visit 2 ($M = 5.37$, $SD = 4.64$) and lower (-0.90) than the marginal mean of the VAPP *social communication* score at visit 3 ($M = 6.89$, $SD = 6.28$); 4) there was no significant univariate effect of "Visit" $F(2,98) = 1.58$, $p = .210$ for the VAPP *emotion* score; 5) there was no significant univariate effect of "Visit" $F(2,98) = 0.72$, $p = .490$ for the VAPP *activity limitation* score; 6) there was a significant univariate effect of "Visit" $F(2,98) = 3.19$, $p = .046$ for the VAPP *participation restriction* score which indicated that the marginal mean of VAPP *participation restriction* score at visit 1 ($M = 12.45$, $SD = 12.14$) was greater (+0.04) than the marginal mean of VAPP *participation restriction* score at visit 2 ($M = 12.41$, $SD = 10.43$) and was lower (-2.59) than the marginal mean of the VAPP *participation restriction* score at visit 3 ($M = 15.04$, $SD = 14.00$); 7) there was no significant univariate effect of "Visit" $F(2,98) = 1.46$, $p = .238$ for the VAPP *total* score.

It is important to note that this multivariate effect of "Visit" needs to be qualified because of the finding of a significant "Group" by "Visit" interaction $F(13,37) = 2.75$, $p = .008$ for the individual VAPP subsections. The results of this repeated measures MANOVA can be found in Table 33. Subsequent univariate analyses revealed the following results: 1) There was no significant "Group" by "Visit" interaction $F(2,98) = 0.74$, $p = .480$ for the VAPP *self-perceived voice problem* score. This interaction is illustrated in Figure 17. 2) There was no significant "Group" by "Visit" interaction $F(2,98) = 1.37$, $p = .258$ for the VAPP *daily communication* score. This interaction is illustrated in Figure 18. 3) There was no significant "Group" by "Visit" interaction $F(2,98) = 2.18$, $p = .119$ for the VAPP *social communication* score. This interaction is illustrated in Figure 19. 4) There was no significant "Group" by "Visit" interaction $F(2,98) = 1.27$, $p = .286$ for the VAPP *emotion* score. This interaction is illustrated in Figure 20. 5) There was no significant "Group" by "Visit" interaction $F(2,98) = 0.78$, $p = .463$ for the VAPP *activity limitation* score. This interaction is illustrated in Figure 21. 6) There was no significant "Group" by "Visit" interaction $F(2,98) = 2.11$, $p = .127$ for the VAPP *participation restriction* score. This interaction is illustrated in Figure 22. 7) There was no significant "Group" by "Visit" interaction $F(2,98) = 1.17$, $p = .313$ for the VAPP *total* score. This interaction is illustrated in Figure 23. The results from the univariate analysis for the individual VAPP subsection scores are provided in Table 34.

Figure 17: Means of Voice Activity and Participation Profile *Self-Perceived Voice Problem* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

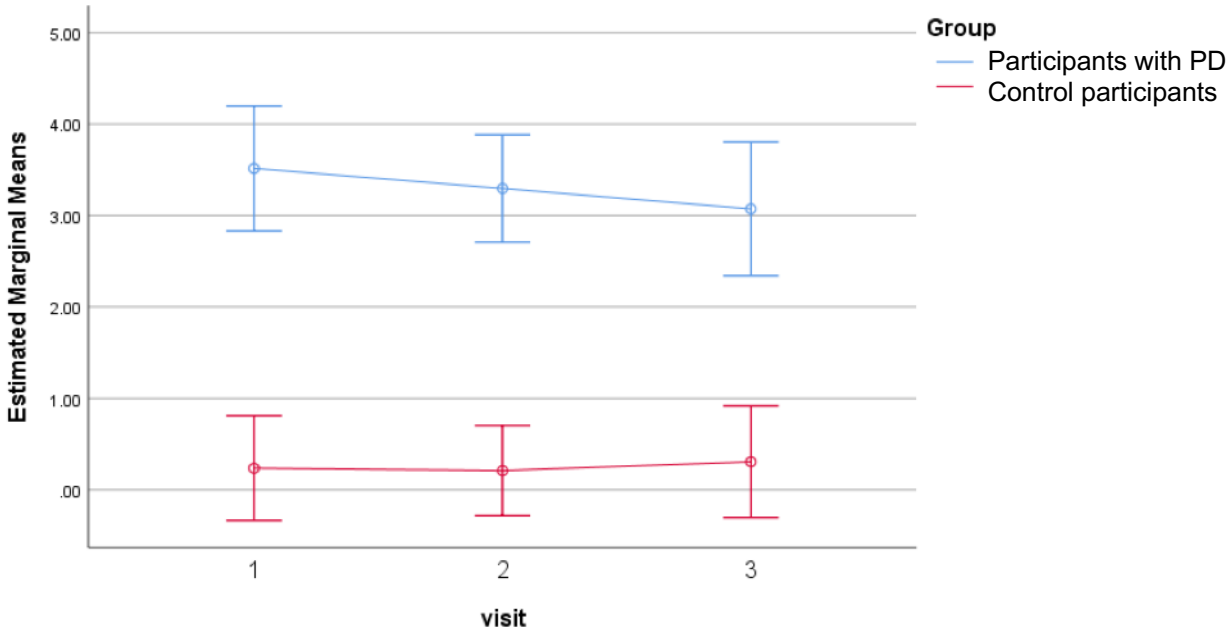

Figure 17. This figure demonstrates the changes in mean VAPP *self-perceived voice problem* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 18: Means of Voice Activity and Participation Profile *Daily Communication* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

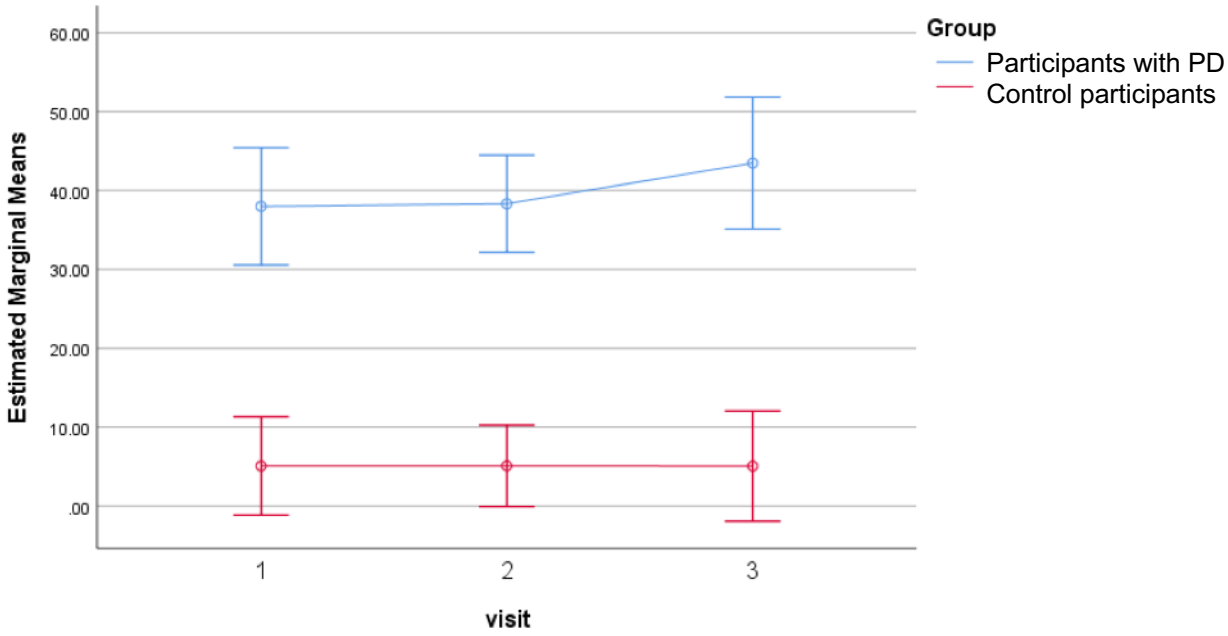

Figure 18. This figure demonstrates the changes in mean VAPP *daily communication* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 19: Means of Voice Activity and Participation Profile *Social Communication* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

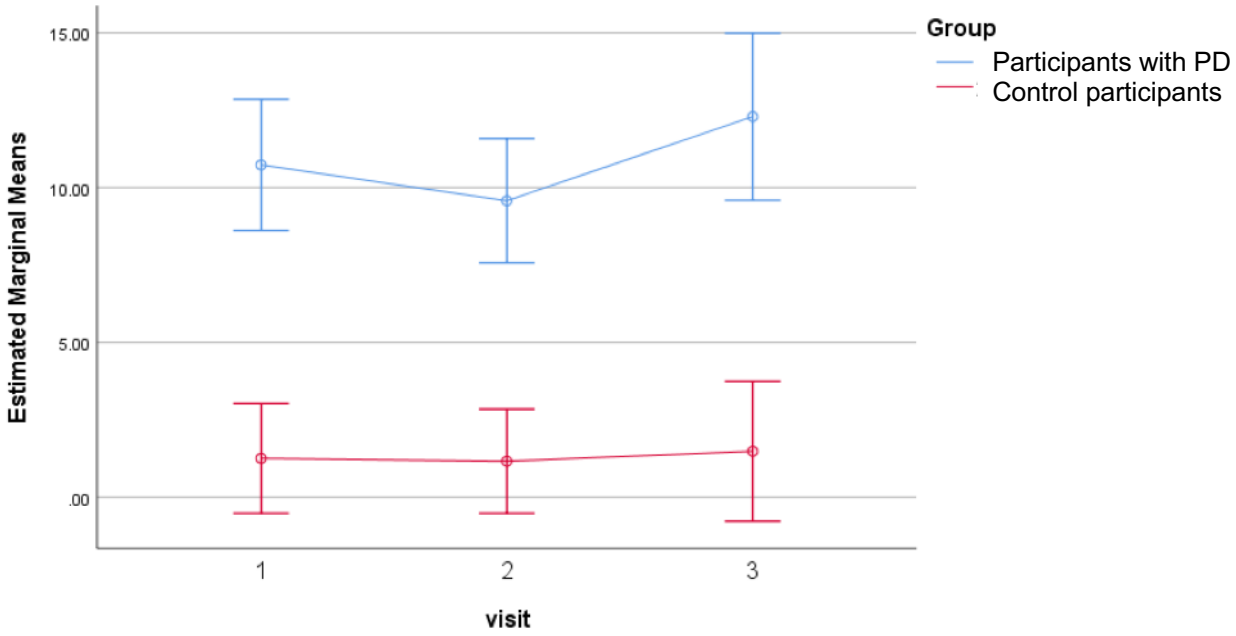

Figure 19. This figure demonstrates the changes in mean VAPP *social communication* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 20: Means of Voice Activity and Participation Profile *Emotion* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

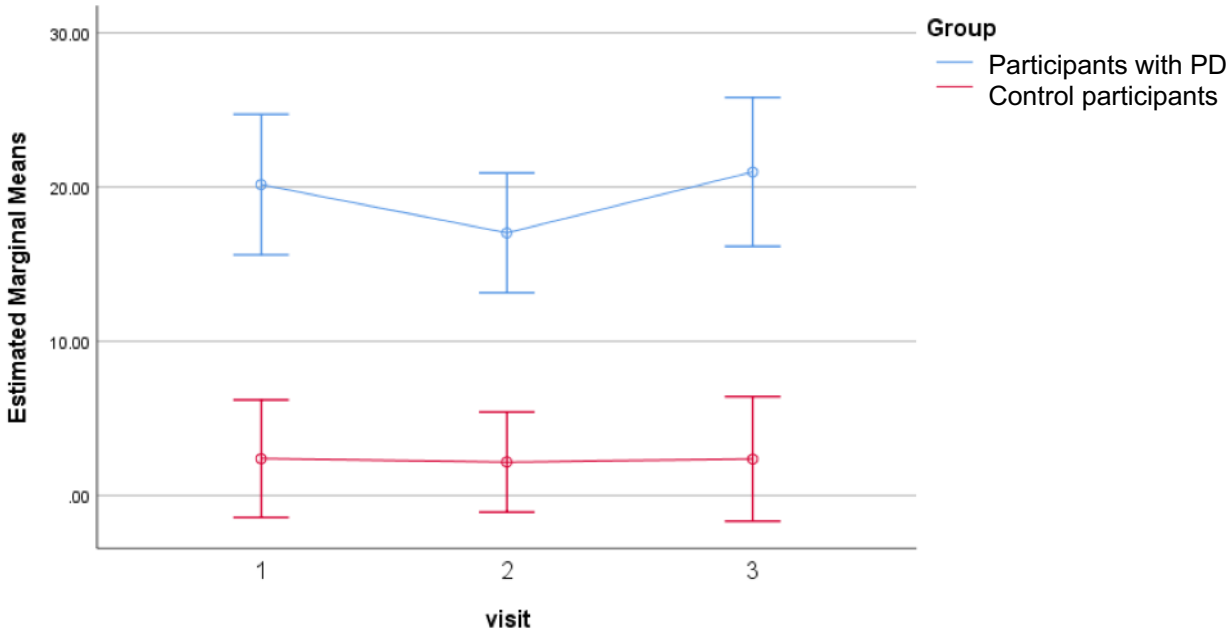

Figure 20. This figure demonstrates the changes in mean VAPP *emotion* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 21: Means of Voice Activity and Participation Profile *Activity Limitation* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

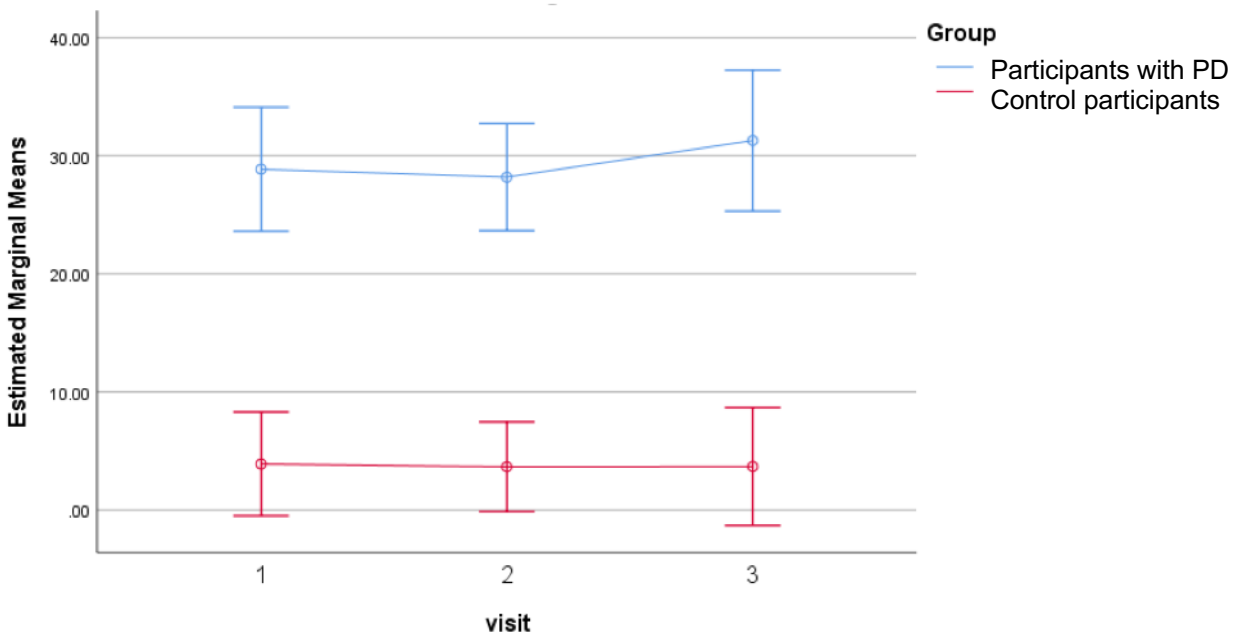

Figure 21. This figure demonstrates the changes in mean VAPP *activity limitation* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 22: Means of Voice Activity and Participation Profile *Participation Restriction* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

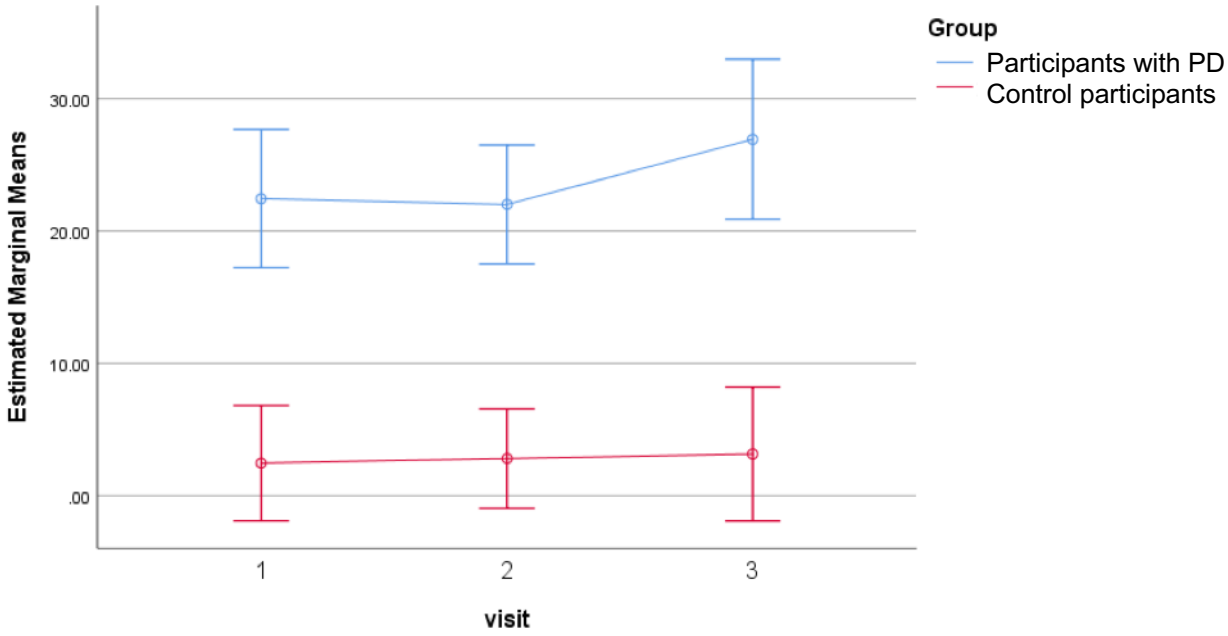

Figure 22. This figure demonstrates the changes in mean VAPP *participation restriction* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Figure 23: Means of Voice Activity and Participation Profile *Total* Subsection Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

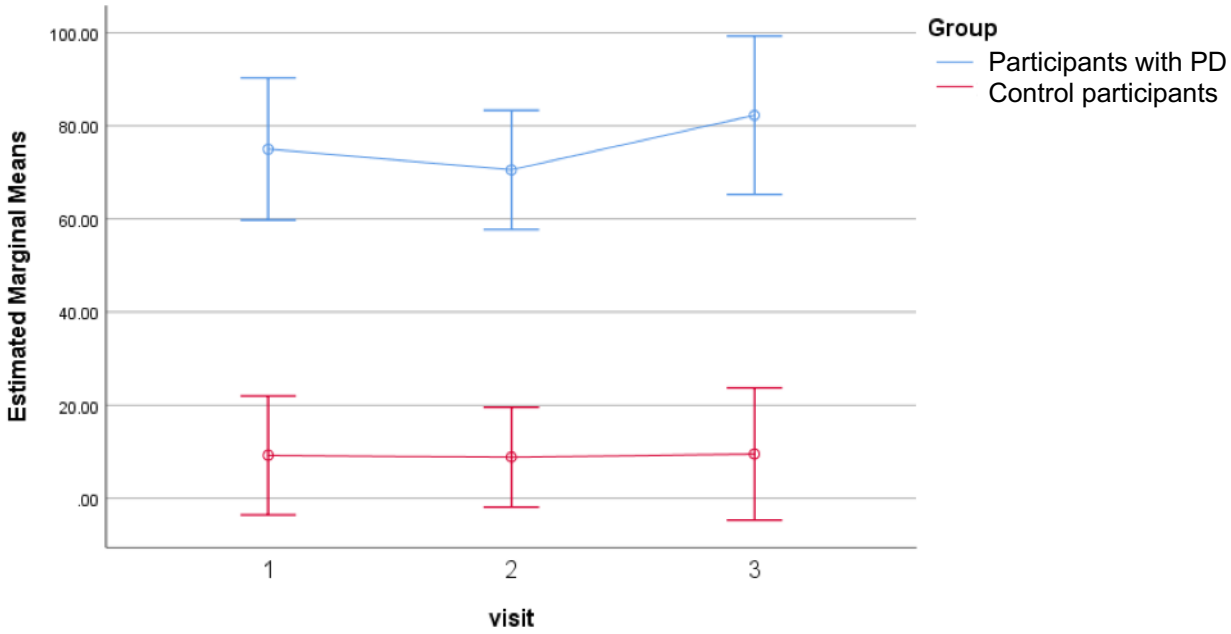

Figure 23. This figure demonstrates the changes in mean VAPP *total* subsection scores for participants with PD and control participants across visits. Error bars represent standard deviations.

Table 34: Univariate Testing Analysis of Self-Rated Individual Subsections of the Voice Activity and Participation Profile for Participants with Parkinson's Disease and Control Participants for Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p* value
"Group"				
Voice problem	1	49	71.94	<.001
Daily communication	1	49	66.31	<.001
Social communication	1	49	51.45	<.001
Emotion	1	49	46.53	<.001
Activity limitation	1	49	70.39	<.001
Participation restriction	1	49	44.57	<.001
Total	1	49	56.84	<.001
"Visit"				
Voice problem	2	98	0.40	.674
Daily communication	2	98	1.35	.264
Social communication	2	98	3.51	.034
Emotion	2	98	1.58	.210
Activity limitation	2	98	0.72	.490
Participation restriction	2	98	3.19	.046
Total	2	98	1.46	.238
"Group"*"Visit"				
Voice problem	2	98	0.74	.480
Daily communication	2	98	1.37	.258
Social communication	2	98	2.18	.119
Emotion	2	98	1.27	.286
Activity limitation	2	98	0.78	.463
Participation restriction	2	98	2.11	.127
Total	2	98	1.17	.313

Note. This table illustrates the significant and non-significant univariate effects of "Group", significant and non-significant univariate effects of "Visit", and "Group" by "Visit" interactions found from univariate analyses of the individual subsections of the VAPP for participants with PD and control participants.

3.6.2.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of individual VAPP subsection scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants.

3.6.2.1.1 Retest analysis of VAPP *self-perceived voice problem* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *self-perceived voice problem* scores for participants with PD are summarized in Table 35. The ICC values related to the repeated measurement of VAPP *self-perceived voice problem* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .77, *p* = .001, visit 1 versus visit 3 *ICC* = .59, *p* = .026, and visit 2 versus visit 3 *ICC* = .65, *p* = .013. These results suggest that VAPP *self-perceived voice problem* scores as measured in the present study did not demonstrate good retest reliability in IWPD because the ICC values across two of three comparisons were below our criterion of .75.

Table 35: Retest Analyses of Voice Activity and Participation Profile *Self-Perceived Voice Problem* Subsection Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.77[.44-.91]	-0.05	2.01	-0.13	.900	2.27	1.09	3.02	87.96
Visits 1 – 3	.59[-.00-.83]	0.34	2.73	0.58	.569	2.53	1.62	4.49	131.70
Visits 2 – 3	.65[.13-.86]	0.22	2.40	0.43	.673	2.38	1.41	3.90	120.00
Control participants									
Visits 1 – 2	.67[.31-.84]	0.03	0.28	0.53	.601	0.28	0.16	0.44	183.08
Visits 1 – 3	.29[-.50-.66]	-0.07	0.46	-0.84	.407	0.35	0.30	0.83	392.96
Visits 2 – 3	.23[-.59-.63]	-0.10	0.44	-1.21	.236	0.33	0.29	0.81	261.53
Primary communication partners									
Visits 1 – 2	.80[.53-.91]	-0.40	1.95	-1.00	.330	2.36	1.06	2.93	98.49
Visits 1 – 3	.92[.80-.97]	0.26	1.25	0.97	.343	2.24	0.63	1.76	51.93
Visits 2 – 3	.80[.53-.92]	0.59	1.88	1.47	.156	2.32	1.04	2.88	102.41

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *self-perceived voice problem* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *self-perceived voice problem* scores was -0.05 (*SD* = 2.01) and this difference was not significant $t(21) = -0.13$, $p = .900$. The mean differences in retest values ranged from -5.8 – 3.8. The mean difference

between visit 1 and visit 3 for VAPP *self-perceived voice problem* scores was 0.34 (*SD* = 2.73) and this difference was not significant $t(21) = 0.58$, $p = .569$. The mean differences in retest values ranged from -7.0 – 4.5. The mean difference between visit 2 and visit 3 for VAPP *self-perceived voice problem* scores was 0.22 (*SD* = 2.40) and this difference was not significant $t(20) = 0.43$, $p = .673$. The mean differences in retest values ranged from -6.60 – 3.70. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 3.02 and *CR%* = 87.96%, for visit 1 vs visit 3, *CR* = 4.49 and *CR%* = 131.70%, and for visit 2 vs visit 3, *CR* = 3.90 and *CR%* = 120.00%. Based on the CR, an observed change in the VAPP *self-perceived voice problem* scores of at least 3.02 – 4.49 would suggest a large amount of measurement error ranging from 87.96 – 131.70% variation in the VAPP *self-perceived voice problem* scores of IWPD. These results suggest that the measure of VAPP *self-perceived voice problem* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.2 Retest analysis of VAPP *self-perceived voice problem* scores for control participants

Results for retest analyses of the VAPP *self-perceived voice problem* scores for control participants are summarized in Table 35. The ICC values related to the repeated measurement of VAPP *self-perceived voice problem* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .67, $p = .002$, visit 1 versus visit 3 *ICC* = .29, $p = .182$, and visit 2 versus visit 3 *ICC* = .23, $p = .241$. These results suggest that VAPP *self-perceived voice problem* scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *self-perceived voice problem* scores was 0.03 (*SD* = 0.28) and this difference was not significant $t(29) = .53$, $p = .601$. The mean differences in retest values ranged from -0.70 – 0.50. The mean difference between visit 1 and visit 3 for VAPP *self-perceived voice problem* scores was -0.07 (*SD* = 0.46) and this difference was not significant $t(29) = -0.84$, $p = .407$. The mean differences in retest values ranged from -0.70 – 0.90. The mean difference between visit 2 and visit 3 for VAPP *self-perceived voice problem* scores was -0.10 (*SD* = 0.44) and this difference

was not significant $t(29) = -1.21$, $p = .236$. The mean differences in retest values ranged from -1.90 – 1.10. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.44 and *CR%* = 183.08%, for visit 1 vs visit 3, *CR* = 0.83 and *CR%* = 392.96%, and for visit 2 vs visit 3, *CR* = 0.81 and *CR%* = 261.53%. Based on the CR, an observed change in the VAPP *self-perceived voice problem* scores of at least 0.44 – 0.83 would suggest a large amount of measurement error ranging from 183.33 – 392.96% variation in the VAPP *self-perceived voice problem* scores of control participants. These results suggest that the measure of VAPP *self-perceived voice problem* scores demonstrates unacceptable repeatability for control participants.

3.6.2.1.3 Retest analysis of VAPP *daily communication* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *daily communication* scores for participants with PD are summarized in Table 36. The ICC values related to the repeated measurement of VAPP *daily communication* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .88$, $p < .001$, visit 1 versus visit 3 $ICC = .76$, $p = .001$, and visit 2 versus visit 3 $ICC = .81$, $p < .001$. These results suggest that VAPP *daily communication* scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

Table 36: Retest Analyses of Voice Activity and Participation Profile *Daily Communication* Subsection Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Partners Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.88[.72-.95]	-0.87	15.05	-0.28	.784	22.85	7.91	21.92	57.42
Visits 1 – 3	.76[.44-.90]	-7.05	24.29	-1.36	.188	27.77	13.61	37.69	96.51
Visits 2 – 3	.81[.56-.92]	-6.53	20.38	-1.50	.148	25.92	11.30	31.29	68.63
Control participants									
Visits 1 – 2	.78[.54-.90]	-0.02	4.14	-0.02	.983	4.85	2.27	6.30	123.92
Visits 1 – 3	.79[.57-.90]	0.02	3.89	0.02	.981	4.65	2.13	5.91	115.80
Visits 2 – 3	.84[.66-.92]	0.03	3.54	0.05	.959	4.70	1.88	5.21	102.90
Primary communication partners									
Visits 1 – 2	.88[.72-.95]	-5.71	17.84	-1.54	.139	27.82	9.64	26.70	79.39
Visits 1 – 3	.90[.77-.96]	-4.99	15.51	-1.51	.146	26.35	8.33	23.08	58.67
Visits 2 – 3	.95[.87-.98]	0.36	12.76	0.13	.895	27.13	6.07	16.80	42.93

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *daily communication* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *daily communication* scores was -0.87 (*SD* = 15.05) and this difference was not significant $t(22) = -0.28$, $p = .784$. The mean differences in retest values ranged from -32.20 – 24.40. The mean difference between visit 1 and visit 3 for VAPP *daily communication* scores was -7.05 (*SD* = 24.29) and this difference was not significant $t(21) = -1.36$, $p = .188$. The mean differences in retest values ranged from -71.60 – 37.60. The mean difference between visit 2 and visit 3 for VAPP *daily communication* scores was -6.53 (*SD* = 20.38) and this difference was not significant $t(21) = -1.50$, $p = .148$. The mean differences in retest values ranged from -49.10 – 23.20. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 21.92 and *CR%* = 57.42%, for visit 1 vs visit 3, *CR* = 37.69 and *CR%* = 96.51%, and for visit 2 vs visit 3, *CR* = 31.29 and *CR%* = 68.63%. Based on the CR, an observed change in the VAPP *daily communication* scores of at least 21.92 – 37.69 would suggest a large amount of measurement error ranging from 57.42 – 96.51% variation in the VAPP *daily communication* scores of IWPD. These

results suggest that the measure of VAPP *daily communication* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.4 Retest analysis of VAPP *daily communication* scores for control participants

Results for retest analyses of the VAPP *daily communication* scores for control participants are summarized in Table 36. The ICC values related to the repeated measurement of VAPP *daily communication* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .78$, $p < .001$, visit 1 versus visit 3 $ICC = .79$, $p < .001$, and visit 2 versus visit 3 $ICC = .84$, $p < .001$. These results suggest that VAPP *daily communication* scores as measured in the present study demonstrated good retest reliability in healthy speakers because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *daily communication* scores was -0.02 ($SD = 4.14$) and this difference was not significant $t(29) = -0.02$, $p = .983$. The mean differences in retest values ranged from -12.20 – 9.60. The mean difference between visit 1 and visit 3 for VAPP *daily communication* scores was 0.02 (SD = 3.89) and this difference was not significant $t(29) = 0.02$, $p = .981$. The mean differences in retest values ranged from -7.80 – 9.10. The mean difference between visit 2 and visit 3 for VAPP *daily communication* scores was 0.03 ($SD = 3.54$) and this difference was not significant $t(29) = 0.05$, $p = .959$. The mean differences in retest values ranged from -6.3 – 8.1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 6.30$ and $CR\% = 123.92\%$, for visit 1 vs visit 3, $CR = 5.91$ and $CR\% = 115.80\%$, and for visit 2 vs visit 3, $CR = 5.21$ and $CR\% = 102.90\%$. Based on the CR, an observed change in the VAPP *daily communication* scores of at least 5.21 – 6.30 would suggest a large amount of measurement error ranging from 102.90 – 123.92% variation in the VAPP *daily communication* scores of control participants. These results suggest that the measure of VAPP *daily communication* scores demonstrates unacceptable repeatability for control participants.

3.6.2.1.5 Retest analysis of VAPP *social communication* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *social communication* scores for participants with PD are summarized in Table 37. The ICC values related to the repeated measurement of VAPP *social communication* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .84, $p < .001$, visit 1 versus visit 3 *ICC* = .82, $p < .001$, and visit 2 versus visit 3 *ICC* = .74, $p = .001$. These results suggest that VAPP *social communication* scores as measured in the present study demonstrated good retest reliability in IWPD because the ICC values across all comparisons were above or approached our criterion of .75.

Table 37: Retest Analyses of Voice Activity and Participation Profile *Social Communication* Subsection Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.84[.62-.93]	1.41	5.17	1.31	.203	6.97	2.79	7.72	72.52
Visits 1 – 3	.82[.58-.93]	-2.06	6.51	-1.48	.153	8.44	3.58	9.91	107.41
Visits 2 – 3	.74[.38-.89]	-3.34	7.21	-2.17	.041	8.35	4.26	11.80	91.81
Control participants									
Visits 1 – 2	.57[.09-.80]	0.09	1.31	0.39	.699	1.19	0.78	2.16	171.70
Visits 1 – 3	.60[.15-.81]	-0.23	1.53	-0.81	.424	1.43	0.90	2.50	215.37
Visits 2 – 3	.78[.53-.89]	-0.32	1.05	-1.67	.107	1.26	0.59	1.63	110.19
Primary communication partners									
Visits 1 – 2	.90[.75-.96]	-1.81	4.19	-2.07	.050	7.34	2.32	6.43	82.12
Visits 1 – 3	.92[.79-.97]	-1.71	3.69	-2.18	.041	7.17	2.03	5.62	58.28
Visits 2 – 3	.89[.74-.96]	0.014	4.81	0.01	.990	7.46	2.47	6.85	70.99

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *social communication* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *social communication* scores was 1.41 (*SD* = 5.17) and this difference was not significant $t(22) = 1.31$, $p = .203$. The mean differences in retest values ranged from -4.70 – 14.10. The mean difference between visit 1 and visit 3 for VAPP *social communication* scores was -2.06 (*SD* = 6.51)

and this difference was not significant $t(21) = -1.48$, $p = .153$. The mean differences in retest values ranged from -14.30 – 12.30. The mean difference between visit 2 and visit 3 for VAPP *social communication* scores was -3.34 ($SD = 7.21$) and this difference was significant $t(21) = -2.17$, $p = .041$. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 7.72$ and $CR\% = 72.52\%$, for visit 1 vs visit 3, $CR = 9.91$ and $CR\% = 107.41\%$, and for visit 2 vs visit 3, $CR = 11.80$ and $CR\% = 91.81\%$. Based on the CR, an observed change in the VAPP *social communication* scores of at least 7.72 – 11.80 would suggest a large amount of measurement error ranging from 72.52 – 107.41% variation in the VAPP *social communication* scores of IWPD. These results suggest that the measure of VAPP *social communication* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.6 Retest analysis of VAPP *social communication* scores for control participants

Results for retest analyses of the VAPP *social communication* scores for control participants are summarized in Table 37. The ICC values related to the repeated measurement of VAPP *social communication* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .57$, $p = .015$, visit 1 versus visit 3 $ICC = .60$, $p = .009$, and visit 2 versus visit 3 $ICC = .78$, $p < .001$. These results suggest that VAPP *social communication* scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *social communication* scores was 0.09 ($SD = 1.31$) and this difference was not significant $t(29) = 0.39$, $p = .699$. The mean differences in retest values ranged from -2.10 – 4.60. The mean difference between visit 1 and visit 3 for VAPP *social communication* scores was -0.223 ($SD = 1.53$) and this difference was not significant $t(29) = -0.81$, $p = .424$. The mean differences in retest values ranged from -3.20 – 4.70. The mean difference between visit 2 and visit 3 for VAPP *social communication* scores was -0.32 ($SD = 1.05$) and this difference was not significant $t(29) = -1.67$, $p = .107$. The mean differences in retest values ranged from -3.5 – 1.3. Additionally, the following CR values were obtained for the following visit

comparisons: for visit 1 vs visit 2, *CR* = 2.16 and *CR%* = 171.70%, for visit 1 vs visit 3, *CR* = 2.50 and *CR%* = 215.37%, and for visit 2 vs visit 3, *CR* = 1.63 and *CR%* = 110.19%. Based on the CR, an observed change in the VAPP *social communication* scores of at least 1.63 – 2.50 would suggest a large amount of measurement error ranging from 110.19 – 215.37% variation in the VAPP *social communication* scores of control participants. These results suggest that the measure of VAPP *social communication* scores demonstrates unacceptable repeatability for control participants.

3.6.2.1.7 Retest analysis of VAPP *emotion* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *emotion* scores for participants with PD are summarized in Table 38. The ICC values related to the repeated measurement of VAPP *emotion* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .81, $p < .001$, visit 1 versus visit 3 *ICC* = .75, $p = .002$, and visit 2 versus visit 3 *ICC* = .81, $p < .001$. These results suggest that VAPP *emotion* scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

Table 38: Retest Analyses of Voice Activity and Participation Profile *Emotion* Subsection Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.81[.56-.92]	2.59	11.71	1.06	.301	14.67	6.39	17.71	87.02
Visits 1 – 3	.75[.38-.89]	-1.70	15.37	-0.52	.609	16.75	8.37	23.19	130.53
Visits 2 – 3	.81[.55-.92]	-4.44	12.55	-1.66	.112	15.96	6.96	19.27	86.06
Control participants									
Visits 1 – 2	.64[.23-.83]	0.22	2.32	0.52	.607	2.24	1.34	3.72	155.50
Visits 1 – 3	.77[.52-.89]	0.02	2.12	0.06	.952	2.44	1.17	3.24	149.12
Visits 2 – 3	.76[.48-.88]	-0.20	1.82	-0.59	.559	2.04	1.00	2.77	117.47
Primary communication partners									
Visits 1 – 2	.75[.43-.90]	-3.50	12.08	-1.39	.179	13.75	6.88	19.05	122.82
Visits 1 – 3	.93[.84-.97]	-1.73	6.55	-1.24	.230	13.00	3.44	9.53	50.12
Visits 2 – 3	.84[.62-.93]	1.68	10.65	0.74	.468	14.11	5.65	15.64	89.00

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *emotion* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *emotion* scores was 2.59 (*SD* = 11.71) and this difference was not significant $t(22) = 1.06$, $p = .301$. The mean differences in retest values ranged from -16.40 – 24.30. The mean difference between visit 1 and visit 3 for VAPP *emotion* scores was -1.70 (*SD* = 15.37) and this difference was not significant $t(21) = -0.52$, $p = .609$. The mean differences in retest values ranged from -53.50 – 25.20. The mean difference between visit 2 and visit 3 for VAPP *emotion* scores was -4.44 (*SD* = 12.55) and this difference was not significant $t(21) = -1.66$, $p = .112$. The mean differences in retest values ranged from -42.80 – 12.40. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 17.71 and *CR%* = 87.02%, for visit 1 vs visit 3, *CR* = 23.19 and *CR%* = 130.53%, and for visit 2 vs visit 3, *CR* = 19.27 and *CR%* = 86.06%. Based on the CR, an observed change in the VAPP *emotion* scores of at least 17.71 – 23.19 would suggest a large amount of measurement error ranging from 86.06 – 130.53% variation in the VAPP

emotion scores of IWPD. These results suggest that the measure of VAPP *emotion* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.8 Retest analysis of VAPP *emotion* scores for control participants

Results for retest analyses of the VAPP *emotion* scores for control participants are summarized in Table 38. The ICC values related to the repeated measurement of VAPP *emotion* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .64$, $p = .004$, visit 1 versus visit 3 $ICC = .77$, $p < .001$, and visit 2 versus visit 3 $ICC = .76$, $p < .001$. These results suggest that VAPP *emotion* scores as measured in the present study demonstrated good retest reliability in healthy speakers because the ICC values across two of three comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *emotion* scores was 0.22 ($SD = 2.32$) and this difference was not significant $t(29) = 0.52$, $p = .607$. The mean differences in retest values ranged from -3.10 – 8.20. The mean difference between visit 1 and visit 3 for VAPP *emotion* scores was 0.02 ($SD = 2.12$) and this difference was not significant $t(29) = 0.06$, $p = .952$. The mean differences in retest values ranged from -4.60 – 4.40. The mean difference between visit 2 and visit 3 for VAPP *emotion* scores was -0.20 ($SD = 1.82$) and this difference was not significant $t(29) = -0.59$, $p = .559$. The mean differences in retest values ranged from -4.30 – 5.10. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 3.72$ and $CR\% = 155.50\%$, for visit 1 vs visit 3, $CR = 3.24$ and $CR\% = 149.12\%$, and for visit 2 vs visit 3, $CR = 2.77$ and $CR\% = 117.47\%$. Based on the CR, an observed change in the VAPP *emotion* scores of at least 2.77 – 3.72 would suggest a large amount of measurement error ranging from 117.47 – 155.50% variation in the VAPP *emotion* scores of control participants. These results suggest that the measure of VAPP *emotion* scores demonstrates unacceptable repeatability for control participants.

3.6.2.1.9 Retest analysis of VAPP *activity limitation* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *activity limitation* scores for participants with PD are summarized in Table 39. The ICC values related to the repeated measurement of VAPP *activity limitation* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .89, $p < .001$, visit 1 versus visit 3 *ICC* = .78, $p = .001$, and visit 2 versus visit 3 *ICC* = .80, $p < .001$. These results suggest that VAPP *activity limitation* scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

Table 39: Retest Analyses of Voice Activity and Participation Profile *Activity Limitation* Subsection Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.89[.74-.95]	0.58	10.42	0.27	.791	16.25	5.39	14.93	51.45
Visits 1 – 3	.78[.46-.91]	-3.37	16.79	-0.94	.357	19.36	9.08	25.15	88.47
Visits 2 – 3	.80[.54-.92]	-3.90	14.87	-1.23	.232	18.37	8.21	22.75	70.05
Control participants									
Visits 1 – 2	.77[.52-.89]	0.24	3.19	0.41	.687	3.65	1.75	4.85	124.32
Visits 1 – 3	.83[.65-.92]	0.22	2.77	0.43	.672	3.62	1.49	4.14	113.01
Visits 2 – 3	.80[.57-.90]	-0.02	2.84	-0.04	.969	3.43	1.53	4.25	115.47
Primary communication partners									
Visits 1 – 2	.93[.83-.97]	-3.54	9.77	-1.74	.096	19.70	5.21	14.44	59.31
Visits 1 – 3	.94[.83-.98]	-4.33	9.03	-2.25	.036	19.91	4.88	13.51	48.43
Visits 2 – 3	.97[.92-.97]	-1.00	7.43	-0.63	.537	20.27	3.51	9.72	33.41

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *activity limitation* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *activity limitation* scores was 0.58 (*SD* = 10.42) and this difference was not significant $t(22) = 0.27$, $p = .791$. The mean differences in retest values ranged from -20.30 – 23.20. The mean difference between visit 1 and visit 3 for VAPP *activity limitation* scores was -3.37 (*SD* = 16.79) and this difference was not significant $t(21) = -0.94$, $p = .357$. The mean differences in retest values ranged from -45.10 – 28.60. The mean difference between visit 2 and visit 3 for

VAPP *activity limitation* scores was -3.90 (*SD* = 14.87) and this difference was not significant $t(21) = -1.23$, $p = .232$. The mean differences in retest values ranged from -36.20 – 23.20. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 14.93 and *CR%* = 51.45%, for visit 1 vs visit 3, *CR* = 25.15 and *CR%* = 88.47%, and for visit 2 vs visit 3, *CR* = 22.75 and *CR%* = 70.05%. Based on the CR, an observed change in the VAPP *activity limitation* scores of at least 14.93 – 25.15 would suggest a large amount of measurement error ranging from 51.45 – 88.47% variation in the VAPP *activity limitation* scores of IWPD. These results suggest that the measure of VAPP *activity limitation* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.10 Retest analysis of VAPP *activity limitation* scores for control participants

Results for retest analyses of the VAPP *activity limitation* scores for control participants are summarized in Table 39. The ICC values related to the repeated measurement of VAPP *activity limitation* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .77, $p < .001$, visit 1 versus visit 3 *ICC* = .83, $p < .001$, and visit 2 versus visit 3 *ICC* = .80, $p < .001$. These results suggest that VAPP *activity limitation* scores as measured in the present study demonstrated good retest reliability in healthy speakers because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *activity limitation* scores was 0.24 (*SD* = 3.19) and this difference was not significant $t(29) = 0.41$, $p = .687$. The mean differences in retest values ranged from -7.20 – 6.80. The mean difference between visit 1 and visit 3 for VAPP *activity limitation* scores was 0.22 (*SD* = 2.77) and this difference was not significant $t(29) = 0.43$, $p = .672$. The mean differences in retest values ranged from -5.10 – 6.80. The mean difference between visit 2 and visit 3 for VAPP *activity limitation* scores was -0.02 (*SD* = 2.84) and this difference was not significant $t(29) = -0.04$, $p = .969$. The mean differences in retest values ranged from -6.60 – 4.20. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 4.85 and *CR%* = 124.32%, for visit 1 vs visit 3, *CR* = 4.14 and

CR% = 113.01%, and for visit 2 vs visit 3, *CR* = 4.25 and *CR%* = 115.47%. Based on the CR, an observed change in the VAPP *activity limitation* scores of at least 4.14 – 4.85 would suggest a large amount of measurement error ranging from 113.01 – 124.32% variation in the VAPP *activity limitation* scores of control participants. These results suggest that the measure of VAPP *activity limitation* scores demonstrates unacceptable repeatability for control participants.

3.6.2.1.11 Retest analysis of VAPP *participation restriction* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *participation restriction* scores for participants with PD are summarized in Table 40. The ICC values related to the repeated measurement of VAPP *participation restriction* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .90, $p < .001$, visit 1 versus visit 3 *ICC* = .83, $p < .001$, and visit 2 versus visit 3 *ICC* = .82, $p < .001$. These results suggest that VAPP *participation restriction* scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

Table 40: Retest Analyses of Voice Activity and Participation Profile Participation *Restriction Subsection* Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.90[.76-.96]	0.22	10.07	0.11	.917	16.44	5.20	14.40	64.85
Visits 1 – 3	.83[.59-.93]	-5.60	14.72	-1.79	.089	19.85	8.19	22.67	103.21
Visits 2 – 3	.82[.56-.93]	-6.10	14.12	-2.03	.056	18.71	7.94	21.98	77.65
Control participants									
Visits 1 – 2	.64[.24-.83]	-0.35	2.80	-0.68	.503	2.70	1.62	4.49	183.17
Visits 1 – 3	.53[.02-.77]	-0.69	3.20	-1.18	.249	2.82	1.94	5.36	191.53
Visits 2 – 3	.87[.73-.94]	-0.34	1.91	-0.98	.337	2.85	1.03	2.85	90.73
Primary communication partners									
Visits 1 – 2	.88[.72-.95]	-3.97	11.67	-1.63	.117	18.39	6.37	17.65	93.57
Visits 1 – 3	.84[.61-.93]	-2.14	12.44	-0.81	.430	16.36	6.54	18.12	79.38
Visits 2 – 3	.85[.65-.94]	1.60	12.64	0.59	.559	17.20	6.66	18.46	86.82

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *participation restriction* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *participation restriction* scores was 0.22 (*SD* = 10.07) and this difference was not significant $t(22) = 0.11$, $p = .917$. The mean differences in retest values ranged from -16.60 – 23.30. The mean difference between visit 1 and visit 3 for VAPP *participation restriction* scores was -5.60 (*SD* = 14.72) and this difference was not significant $t(21) = -1.79$, $p = .089$. The mean differences in retest values ranged from -40.80 – 22.80. The mean difference between visit 2 and visit 3 for VAPP *participation restriction* scores was -6.10 (*SD* = 14.12) and this difference approached significance $t(21) = -2.03$, $p = .056$. The mean differences in retest values ranged from -37.20 – 19.40. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 14.40 and *CR%* = 64.85%, for visit 1 vs visit 3, *CR* = 22.67 and *CR%* = 103.21%, and for visit 2 vs visit 3, *CR* = 21.98 and *CR%* = 77.65%. Based on the CR, an observed change in the VAPP *participation restriction* scores of at least 14.40 – 22.67 would suggest a large amount of measurement error ranging from 64.85 – 103.21% variation in the VAPP *participation restriction* scores of IWPD. These results suggest that the measure of VAPP *participation restriction* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.12 Retest analysis of VAPP *participation restriction* scores for control participants

Results for retest analyses of the VAPP *participation restriction* scores for control participants are summarized in Table 40. The ICC values related to the repeated measurement of VAPP *participation restriction* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .64, $p = .004$, visit 1 versus visit 3 *ICC* = .53, $p = .023$, and visit 2 versus visit 3 *ICC* = .87, $p < .001$. These results suggest that VAPP *participation restriction* scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *participation restriction* scores was -0.35 (*SD* = 2.80) and this difference was not significant $t(29) = -0.68$, $p = .503$. The mean differences in retest values ranged from -9.20 – 7.60. The mean difference between visit 1 and visit 3 for VAPP *participation restriction* scores was -0.69 (*SD* = 3.20) and this difference was not significant $t(29) = -1.18$, $p = .249$. The mean differences in retest values ranged from -7.70 – 9.60. The mean difference between visit 2 and visit 3 for VAPP *participation restriction* scores was -0.34 (*SD* = 1.91) and this difference was not significant $t(29) = -0.98$, $p = .337$. The mean differences in retest values ranged from -4.60 – 3.10. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 4.49 and *CR%* = 183.17%, for visit 1 vs visit 3, *CR* = 5.36 and *CR%* = 191.53%, and for visit 2 vs visit 3, *CR* =2.85 and *CR%* = 90.73%. Based on the CR, an observed change in the VAPP *participation restriction* scores of at least 2.85 – 5.36 would suggest a large amount of measurement error ranging from 90.73 – 191.53% variation in the VAPP *participation restriction* scores of control participants. These results suggest that the measure of VAPP *participation restriction* scores demonstrates unacceptable repeatability for control participants.

3.6.2.1.13 Retest analysis of VAPP *total* scores for participants with Parkinson's disease

Results for retest analyses of the VAPP *total* scores for participants with PD are summarized in Table 41. The ICC values related to the repeated measurement of VAPP *total* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .88, $p < .001$, visit 1 versus visit 3 *ICC* = .80, $p < .001$, and visit 2 versus visit 3 *ICC* = .82, $p < .001$. These results suggest that VAPP *total* scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

Table 41: Retest Analyses of Voice Activity and Participation Profile *Total* Subsection Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.88[.73-.95]	3.56	31.14	0.55	.589	47.55	16.47	45.63	60.85
Visits 1 – 3	.80[.53-.92]	-10.33	46.54	-1.04	.310	56.73	25.37	70.27	98.38
Visits 2 – 3	.82[.58-.93]	-14.55	40.45	-1.69	.106	53.18	22.56	62.50	72.31
Control participants									
Visits 1 – 2	.74[.45-.88]	0.37	7.54	0.27	.788	8.22	4.19	11.61	126.05
Visits 1 – 3	.76[.50-.89]	-0.28	7.58	-0.20	.841	8.58	4.20	11.65	131.74
Visits 2 – 3	.84[.66-.92]	-0.65	6.02	-0.59	.557	8.03	3.21	8.90	93.74
Primary communication partners									
Visits 1 – 2	.88[.71-.95]	-11.41	32.55	-1.68	.107	50.65	17.55	48.61	78.79
Visits 1 – 3	.94[.85-.98]	-7.93	22.36	-1.66	.111	47.65	11.67	32.33	44.23
Visits 2 – 3	.92[.81-.97]	2.87	27.87	0.48	.634	50.58	14.31	39.63	56.01

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of VAPP *total* subsection scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for VAPP *total* scores was 3.56 (*SD* = 31.14) and this difference was not significant $t(22) = 0.55$, $p = .589$. The mean differences in retest values ranged from -54.50 – 68.50. The mean difference between visit 1 and visit 3 for VAPP *total* scores was -10.33 (*SD* = 46.54) and this difference was not significant $t(21) = -1.04$, $p = .310$. The mean differences in retest values ranged from -144.90 – 73.00. The mean difference between visit 2 and visit 3 for VAPP *total* scores was -14.55 (*SD* = 40.45) and this difference was not significant $t(21) = -1.69$, $p = .106$. The mean differences in retest values ranged from -114.80 – 35.00. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 45.63 and *CR%* = 60.85%, for visit 1 vs visit 3, *CR* = 70.27 and *CR%* = 98.38%, and for visit 2 vs visit 3, *CR* = 62.50 and *CR%* = 72.31%. Based on the CR, an observed change in the VAPP *total* scores of at least 45.63 – 70.27 would suggest a large amount of measurement error ranging from 60.85 – 98.38% variation in the VAPP *total* scores of

IWPD. These results suggest that the measure of VAPP *total* scores demonstrates unacceptable repeatability for IWPD.

3.6.2.1.14 Retest analysis of VAPP *total* scores for control participants

Results for retest analyses of the VAPP *total* scores for control participants are summarized in Table 41. The ICC values related to the repeated measurement of VAPP *total* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .74$, $p < .001$, visit 1 versus visit 3 $ICC = .76$, $p < .001$, and visit 2 versus visit 3 $ICC = .84$, $p < .001$. These results suggest that VAPP *total* scores as measured in the present study demonstrated good retest reliability in healthy speakers because the ICC values across all comparisons either were above or approached our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *total* scores was 0.37 (SD = 7.54) and this difference was not significant $t(29) = 0.27$, $p = .788$. The mean differences in retest values ranged from -17.80 – 17.80. The mean difference between visit 1 and visit 3 for VAPP *total* scores was -0.28 (SD = 7.58) and this difference was not significant $t(29) = -0.20$, $p = .841$. The mean differences in retest values ranged from -15.30 – 21.60. The mean difference between visit 2 and visit 3 for VAPP *total* scores was -0.65 (SD = 6.02) and this difference was not significant $t(29) = -0.59$, $p = .557$. The mean differences in retest values ranged from -12.60 – 14.90. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 11.61$ and $CR\% = 126.05\%$, for visit 1 vs visit 3, $CR = 11.65$ and $CR\% = 131.74\%$, and for visit 2 vs visit 3, $CR = 8.90$ and $CR\% = 93.74\%$. Based on the CR, an observed change in the VAPP *total* scores of at least 8.90 – 11.65 would suggest a large amount of measurement error ranging from 93.74 – 131.74% variation in the VAPP *total* scores of control participants. These results suggest that the measure of VAPP *total* scores demonstrates unacceptable repeatability for control participants.

3.6.3 CPIB scores

CPIB scores were obtained by converting the raw total score to a standardized score using the conversion table indicated by Baylor and colleagues (2013). Descriptive

statistics for CPIB standardized scores can be found in Table 42. The results of the two-way repeated measures ANOVA for the dependent variable of the CPIB standardized scores showed that there was a significant main effect of "Group" $F(1, 50) = 130.74$, $p < .001$ with participants with PD having a lower (-17.06) marginal mean ($M = 52.01$, $SD = 5.30$) compared to control participants ($M = 69.07$, $SD = 5.31$). In contrast, there was no significant main effect of "Visit" on the CPIB score $F(2,100) = 2.44$, $p = .092$. It is important to note that this main effect of "Visit" needs to be qualified because of the finding of a significant "Group" by "Visit" interaction $F(2,49) = 8.20$, $p = .001$ for the CPIB score. This significant interaction is illustrated in Figure 24. A closer look at the factor "Visit", using pairwise post-hoc analyses indicated that the marginal mean of the CPIB score at visit 1 ($M = 61.02$, $SD = 6.56$) was similar to the marginal mean of the CPIB score at visit 2 ($M = 60.79$, $SD = 5.34$) and visit 3 ($M = 59.82$, $SD = 5.70$). Results of post-hoc analyses involving comparisons of these group differences at each visit provided additional information about this significant "Group" by "Visit" interaction. For the post-hoc comparison related to visit 1, the participants with PD had a mean CPIB score ($M = 53.44$, $SD = 6.49$) that was significantly lower (-15.15) than the mean CPIB score of control participants ($M = 68.59$, $SD = 6.50$), $t(2,51) = -8.36$, $p < .001$. For the post-hoc comparison related to visit 2, the participants with PD had a mean CPIB score ($M = 52.61$, $SD = 5.27$) that was significantly lower (-16.35) than the mean CPIB score of control participants ($M = 68.96$, $SD = 5.27$), $t(2,51) = -10.69$, $p < .001$. For the post-hoc comparison related to visit 3, the participants with PD had a mean CPIB score ($M = 49.96$, $SD = 5.64$) that was significantly lower (-19.71) than the mean CPIB score of control participants ($M = 69.67$, $SD = 5.64$), $t(2,50) = 12.46$, $p < .001$. Thus, it appears that the CPIB score of participants with PD decreased from visit 1 to visit 3, whereas the CPIB scores of control participants increased from visit 1 to visit 3.

Table 42: Descriptive Statistics of Self- and Proxy-Rated Communicative Participation Item Bank for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	CPIB Standardized Score mean(*SD*)
Participants with PD	
Visit 1	53.63*(8.41)*
Visit 2	53.03*(6.81)*
Visit 3	49.96*(7.80)*
Control participants	
Visit 1	68.59*(4.44)*
Visit 2	68.96*(3.97)*
Visit 3	69.67*(3.27)*
Primary communication partners	
Visit 1	56.27*(8.94)*
Visit 2	54.83*(8.58)*
Visit 3	52.84*(10.77)*
Marginal means	
Participants with PD	52.01*(5.30)*
Control participants	69.07*(5.31)*
Primary communication partners	54.58*(7.83)*
PD and control visit 1	61.02*(6.56)*
PD and control visit 2	60.79*(5.34)*
PD and control visit 3	59.82*(5.70)*
PD and communication partners visit 1	54.85*(8.89)*
PD and communication partners visit 2	53.62*(7.76)*
PD and communication partners visit 3	51.40*(9.42)*

Note. This table illustrates the means and standard deviations for the standardized scores of the CPIB for participants with PD, their primary communication partners, and control participants across visits.

Figure 24: Means of Communicative Participation Item Bank Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

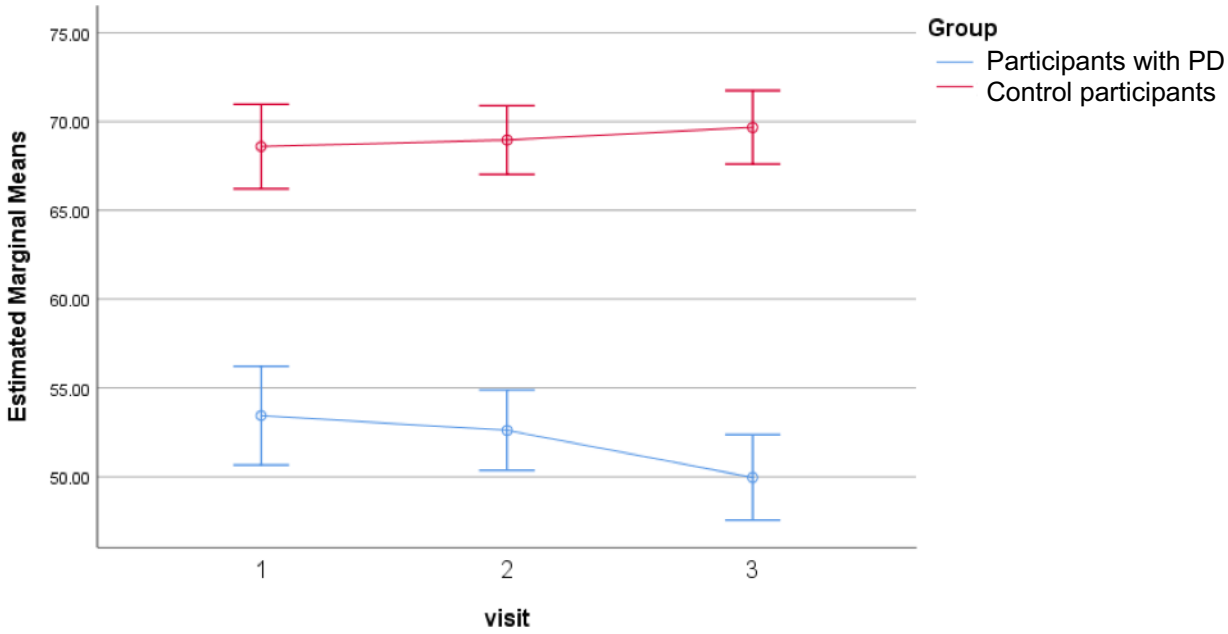

Figure 24. This figure demonstrates the changes in mean CPIB scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.6.3.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of CPIB scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants.

3.6.3.1.1 Retest analysis of CPIB scores for participants with Parkinson's disease

Results for retest analyses of CPIB scores for participants with PD are summarized in Table 43.The ICC values related to the repeated measurement of CPIB scores across the

three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .87, $p < .001$, visit 1 versus visit 3 *ICC* = .84, $p < .001$, and visit 2 versus visit 3 *ICC* = .84, $p < .001$. These results suggest that CPIB scores as measured in the present study demonstrated good retest reliability in IWPD because all of the ICC values across all comparisons were above our criterion of .75.

Table 43: Retest Analyses of Communicative Participation Item Bank Scores for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.87[.68-.94]	0.60	5.34	0.54	.596	7.65	2.64	7.64	14.25
Visits 1 – 3	.84[.52-.94]	3.48	5.43	3.00	.007	8.11	3.24	8.99	16.95
Visits 2 – 3	.84[.41-.88]	2.65	4.96	2.50	.021	7.32	2.93	8.11	16.24
Control participants									
Visits 1 – 2	.87[.73-.94]	-0.36	2.87	-0.69	.493	4.21	1.52	4.21	6.13
Visits 1 – 3	.84[.65-.92]	-1.08	2.81	-2.10	.045	3.90	1.56	4.32	6.26
Visits 2 – 3	.73[.45-.87]	-0.71	3.33	-1.17	.250	3.64	1.89	5.23	7.51
Primary communication partners									
Visits 1 – 2	.78[.49-.91]	1.44	7.47	0.93	.364	8.76	4.11	11.38	20.23
Visits 1 – 3	.76[.43-.90]	3.42	8.48	1.89	.072	9.90	4.85	13.43	24.50
Visits 2 – 3	.87[.70-.95]	1.79	6.54	1.29	.213	9.74	3.51	9.72	18.40

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of CPIB scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for CPIB standardized scores was 0.60 (*SD* = 5.34) and this difference was not significant $t(22) = 0.54$, $p = .596$. The mean differences in retest values ranged from -7.50 – 14.50. The mean difference between visit 1 and visit 3 for CPIB scores was 3.48 (*SD* = 5.43) and this difference was significant $t(21) = 3.00$, $p = .007$. The mean difference between visit 2 and visit 3 for CPIB scores was 2.65 (*SD* = 4.96) and this difference was significant $t(21) = 2.50$, $p = .021$. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 7.64 and *CR%* = 14.25%, for visit 1 vs visit 3, *CR* = 8.99 and *CR%* = 16.95%, and for visit 2 vs visit 3, *CR* = 8.11 and *CR%* = 16.24%. Based on the

CR, an observed change in the CPIB scores of at least 7.64 – 8.99 would suggest the possibility of an acceptable amount of measurement error ranging from 14.25 – 16.95% variation in the CPIB scores of IWPD. These results suggest that the measure of CPIB scores demonstrates marginal repeatability for IWPD.

3.6.3.1.2 Retest analysis of CPIB scores for control participants

Results for retest analyses of CPIB scores for control participants are summarized in Table 43. The ICC values related to the repeated measurement of CPIB scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .87$, $p < .001$, visit 1 versus visit 3 $ICC = .84$, $p < .001$, and visit 2 versus visit 3 $ICC = .73$, $p < .001$. These results suggest that CPIB scores as measured in the present study demonstrated good retest reliability in control participants because the ICC values across all comparisons either were above or approached our criterion of .75.

The mean difference between visit 1 and visit 2 for CPIB scores was -0.36 ($SD = 2.87$) and this difference was not significant $t(29) = -0.69$, $p = .493$. The mean differences in retest values ranged from -8.80 – 6.80. The mean difference between visit 1 and visit 3 for CPIB was -1.08 ($SD = 2.81$) and this difference was significant $t(29) = -2.10$, $p = .045$. The mean difference between visit 2 and visit 3 for CPIB scores was -0.71 ($SD = 3.33$) and this difference was not significant $t(29) = -1.17$, $p = .250$. The mean differences in retest values ranged from -8.80 – 8.80. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 4.21$ and $CR\% = 6.13\%$, for visit 1 vs visit 3, $CR = 4.32$ and $CR\% = 6.26\%$, and for visit 2 vs visit 3, $CR = 5.23$ and $CR\% = 7.51\%$. Based on the CR, an observed change in CPIB scores of at least 4.21 – 5.23 would suggest an acceptable amount of measurement error ranging from 6.13 – 7.51% variation in CPIB scores of control participants. These results suggest that the measure of CPIB scores demonstrates good repeatability for control participants.

3.6.4 LSUS scores

LSUS scores were obtained based on whether participants indicated that the first, second, third, fourth, or fifth categories was indicative of their typical level of speech usage.

Descriptive statistics for LSUS scores can be found in Table 44. The results of the two-way repeated measures ANOVA for the dependent variables of the LSUS score showed that there was a significant main effect of "Group" $F(1,50) = 12.30$, $p = .001$ with participants with PD having a lower (-0. 58) marginal mean ($M = 1.99$, $SD = 0.61$) compared to control participants ($M = 2.57$, $SD = 0.60$). In contrast, there was no significant main effect of "Visit" on the LSUS score $F(2,100) = 0.77$, $p = .465$. It is important to note that this main effect of "Visit" needs to be qualified because of the finding of a "Group" by "Visit" interaction that trended towards significance $F(2,100) = 2.79$, $p = .066$ for the LSUS score. This significant interaction is illustrated in Figure 25. A closer look at the factor "Visit", using pairwise post-hoc analyses indicated that the marginal mean of the LSUS score at visit 1 ($M = 2.27$, $SD = 0.72$) was similar to the marginal mean of the LSUS score at visit 2 ($M = 2.34$, $SD = 0.65$) and visit 3 ($M = 2.22$, $SD = 0.72$). It appears that at visit 1 the group difference between the LSUS scores of participants PD and control participants greater than the group differences that were found at visit 2 and visit 3. Results of a post-hoc analyses involving comparisons of these group differences at each visit provided additional information about this significant "Group" by "Visit" interaction. For the post-hoc comparison related to visit 1, the participants with PD had a mean LSUS score ($M = 1.91$, $SD = 0.73$) that was significantly lower (-0.67) than the mean LSUS score of control participants ($M = 2.67$, $SD = 0.76$), $t(2,51) = 3.64$, $p = .001$. For the post-hoc comparison related to visit 2, the participants with PD had a mean LSUS score ($M = 2.00$, $SD = 0.74$) that was significantly lower (-0.63) than the mean LSUS score of control participants ($M = 2.63$, $SD = 0.61$), $t(2,51) = -3.41$, $p = .001$. In contrast, the post-hoc comparisons related to group differences were not significant at visit 3 $t(2,50) = -1.75$, $p = .087$. Thus, it appears that although the LSUS score of participants with PD was significantly lower than that of control participants at visit 1 and visit 2, it increased to a level that was not significantly different from control participants at visit 3.

Table 44: Descriptive Statistics of Self- and Proxy-Rated Level of Speech Usage Scale for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	LSUS score mean(*SD*)
Participants with PD	
Visit 1	1.91*(0.73)*
Visit 2	2.00*(0.74)*
Visit 3	2.05*(0.72)*
Control participants	
Visit 1	2.67*(0.76)*
Visit 2	2.63*(0.61)*
Visit 3	2.40*(0.72)*
Primary communication partners	
Visit 1	2.13*(0.92)*
Visit 2	2.04*(0.93)*
Visit 3	1.82*(0.73)*
Marginal means	
Participants with PD	1.99*(0.61)*
Control participants	2.57*(0.60)*
Primary communication partners	1.99*(0.70)*
PD and control visit 1	2.27*(0.72)*
PD and control visit 2	2.34*(0.65)*
PD and control visit 3	2.22*(0.72)*
PD and communication partners visit 1	1.98*(0.80)*
PD and communication partners visit 2	2.05*(0.86)*
PD and communication partners visit 3	1.93*(0.73)*

Note. This table illustrates the means and standard deviations for the LSUS for participants with PD, their primary communication partners, and control participants across visits.

Figure 25: Means of Level of Speech Usage Scale Scores for Participants with Parkinson's Disease and Control Participants Across Visits 1, 2, and 3

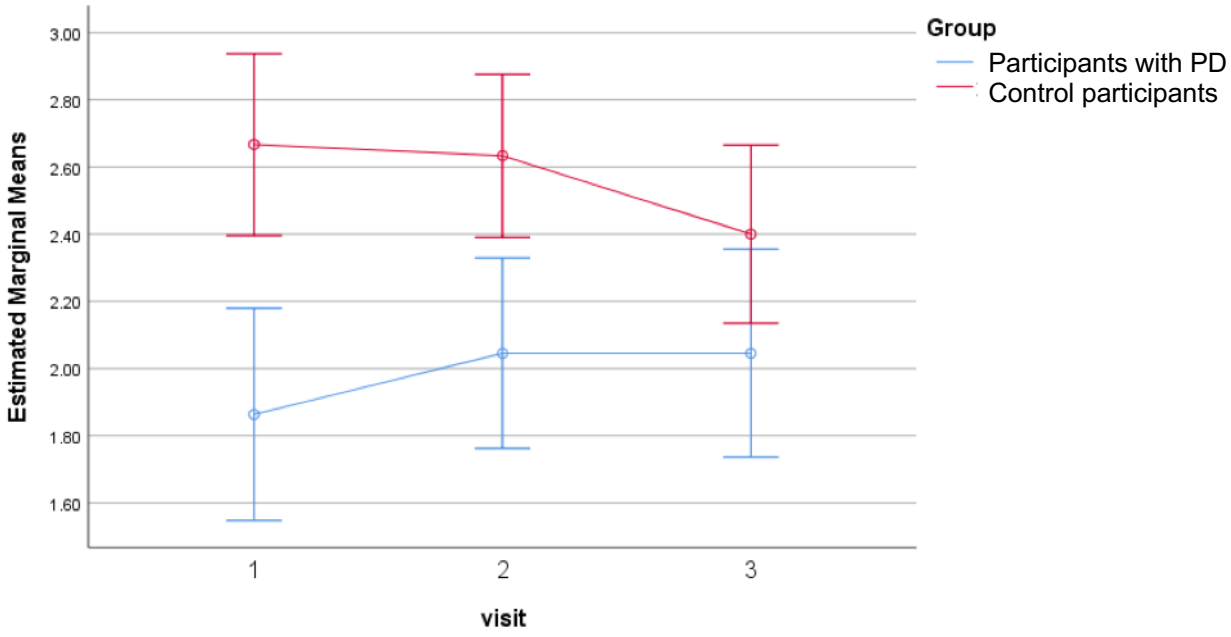

Figure 25. This figure demonstrates the changes in mean LSUS scores for participants with PD and control participants across visits. Error bars represent standard deviations.

3.6.4.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of LSUS scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and control participants.

3.6.4.1.1 Retest analysis of LSUS scores for participants with Parkinson's disease

Results for retest analyses of LSUS scores for participants with PD are summarized in Table 45. The ICC values related to the repeated measurement of LSUS scores across the

three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .75, *p* = .001, visit 1 versus visit 3 *ICC* = .72, *p* = .002, and visit 2 versus visit 3 *ICC* = .85, *p* < .001. These results suggest that LSUS scores as measured in the present study demonstrated good retest reliability in IWPD because the ICC values across all comparisons were above or approached our criterion of .75.

Table 45: Retest Analyses of Level of Speech Usage Scale for Participants with Parkinson's Disease, Primary Communication Partners of Participants with Parkinson's Disease, and Control Participants Across Visits 1, 2, and 3

	ICC [95% CI]	Average mean difference	SD difference	t value	*p* value	SD pooled	SEM	CR	CR%
Participants with PD									
Visits 1 – 2	.75[.40-.89]	-0.09	0.67	-0.62	.539	0.74	0.37	1.02	53.30
Visits 1 – 3	.72[.34-.88]	-0.18	0.66	-1.28	.213	0.73	0.38	1.06	53.13
Visits 2 – 3	.85[.63-.94]	0.00	0.53	0.00	1.000	0.73	0.28	0.78	38.21
Control participants									
Visits 1 – 2	.76[.49-.89]	0.03	0.61	0.30	.769	0.69	0.34	0.94	35.02
Visits 1 – 3	.46[-.08-.74]	0.27	0.87	1.68	.103	0.74	0.54	1.51	57.29
Visits 2 – 3	.57[.12-.79]	0.23	0.73	1.76	.090	0.67	0.48	1.21	50.50
Primary communication partners									
Visits 1 – 2	.70[.28-.87]	0.09	0.90	0.46	.648	0.93	0.51	1.40	65.89
Visits 1 – 3	.60[.07-.83]	0.27	0.88	1.45	.162	0.83	0.53	1.45	71.32
Visits 2 – 3	.80[.52-.91]	0.23	0.69	1.56	.135	0.84	0.37	1.04	56.90

Note. This table illustrates the retest analyses performed to examine the repeatability and retest reliability of LSUS scores for participants with PD, their primary communication partners, and control participants.

The mean difference between visit 1 and visit 2 for LSUS scores was -0.09 (*SD* = 0.67) and this difference was not significant *t*(22) = -0.62, *p* = .539. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for LSUS scores was -0.18 (*SD* = 0.66) and this difference was not significant *t*(21) = -1.28, *p* = .213. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 2 and visit 3 for LSUS scores was 0.00 (*SD* = 0.53) and this difference was not significant *t*(21) = 0.00, *p* = 1.00. The mean differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.02 and *CR%* = 53.30%, for visit 1 vs visit 3, *CR* = 1.06 and *CR%* = 53.13%, and for visit 2 vs visit 3, *CR* = 0.78 and *CR%* = 38.21%.

Based on the CR, an observed change in the LSUS scores of at least 0.78 – 1.06 would suggest a fairly large amount of measurement error ranging from 38.21 – 53.30% variation in the LSUS scores of IWPD. These results suggest that the measure of LSUS scores demonstrates unacceptable repeatability for IWPD.

3.6.4.1.2 Retest analysis of LSUS scores for control participants

Results for retest analyses of LSUS scores for control participants are summarized in Table 45. The ICC values related to the repeated measurement of LSUS scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .76, $p < .001$, visit 1 versus visit 3 *ICC* = .46, $p = .043$, and visit 2 versus visit 3 *ICC* = .57, $p = .011$. These results suggest that LSUS scores as measured in the present study did not demonstrate good retest reliability in healthy speakers because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for LSUS scores was 0.03 (*SD* = 0.61) and this difference was not significant $t(29) = 0.30$, $p = .769$. The mean differences in retest values ranged from -1 – 1. The mean difference between visit 1 and visit 3 for LSUS scores was 0.27 (*SD* = 0.87) and this difference was not significant $t(29) = 1.68$, $p = .103$. The mean differences in retest values ranged from -1 – 3. The mean difference between visit 2 and visit 3 for LSUS scores was 0.23 (*SD* = 0.73) and this difference was not significant $t(29) = 1.76$, $p = .090$. The mean differences in retest values ranged from -1 – 2. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 0.94 and *CR%* = 35.02%, for visit 1 vs visit 3, *CR* = 1.51 and *CR%* = 57.29%, and for visit 2 vs visit 3, *CR* = 1.21 and *CR%* = 50.50%. Based on the CR, an observed change in the LSUS of at least 0.94 – 1.51 would suggest a fairly large amount of measurement error ranging from 35.02 – 57.29% variation in the LSUS scores of control participants. These results suggest that the measure of LSUS scores demonstrates unacceptable repeatability for control participants.

3.7 Statistical Analysis for Objective 5: Self- and Proxy-Rated Communicative Participation

In order to answer the question '*Do ratings of communicative participation differ over time between and within participants with PD and their primary communication partners?',* a repeated measures MANOVA was performed to analyze the variability of self-rated communicative participation of the participants with PD and their primary communication partner over three time points for two of the four dependent measures related to communicative participation (CES question scores, VAPP subtest scores). The following factors were used in this analysis: one between-group independent factor with two levels "Group" [PD, communication partners], one within-group independent factor with three levels "Visit" [visit 1, visit 2, visit 3]. Additionally, a repeated measures ANOVA was performed to analyze the variability of self-rated communicative participation of the participants with PD and their primary communication partner over three time points for the remaining two of the four dependent measures related to communicative participation (CPIB scores, LSUS scores). The following factors were used in this analysis: one between-group independent factor with two levels "Group" [PD, communication partners], one within-group independent factors with three levels "Visit" [visit 1, visit 2, visit 3].

3.7.1 CES question scores

Scores for each of the eight individual CES questions were obtained based on participants' responses of 1, 2, 3, or 4 for each question. Descriptive statistics for individual CES question scores can be found in Table 21. The results of the repeated measures MANOVA for the dependent variables of the individual CES questions showed that there was no significant multivariate effect of "Group" $F(8,33) = 1.83$, $p = .106$. The results of this repeated measures MANOVA can be found in Table 46. The results of subsequent univariate testing for each of the individual CES questions revealed the following results: 1) there was a significant univariate effect of "Group" $F(1,40) = 5.48$, $p = .024$ for CES question 1 (*Having a conversation with a family member or friends at home*) with participants with PD having a lower (-0.43) marginal mean ($M = 2.92$, $SD = 0.60$) compared to their primary communication partners ($M = 3.35$, $SD = 0.60$); 2) there was a significant univariate effect of "Group" $F(1,40) = 6.47$, $p = .015$ for CES question 2 (*Participating in conversation with strangers in a quiet place*) with participants with

PD having a lower (-0.42) marginal mean (M = 2.71, SD = 0.55) compared to their primary communication partners (M = 3.13, SD = 0.55); 3) there was a univariate effect of "Group" that trended towards significance for $F(1,40) = 3.65$, $p = .063$ for CES question 3 (*Conversing with a familiar person over the telephone*) with participants with PD having a lower (-0.32) marginal mean (M = 2.86, SD = 0.55) compared to their primary communication partners (M = 3.18, SD = 0.5); 4) there was no significant univariate effect of "Group" $F(1,40) = 1.03$, $p = .316$ for CES question 4 (*Conversing with a stranger over the telephone*); 5) there was no significant univariate effect of "Group" $F(1,40) = 2.38$, $p = .131$ for CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*); 6) there was a univariate effect of "Group" that trended towards significance $F(1,40) = 3.68$, $p = .062$ for CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) with participants with having a lower (-0.36) marginal mean (M = 2.37, SD = 0.64) compared to their primary communication partners (M = 2.73, SD = 0.64); 7) there was a significant univariate effect of "Group" $F(1,40) = 7.50$, $p = .009$ for CES question 7 (*Having a conversation while traveling in a car*) with participants with PD having a lower (-0.44) marginal mean (M = 2.59, SD = 0.55) compared to their primary communication partners (M = 2.59, SD = 0.55); and 8) there was no significant univariate effect of "Group" $F(1,40) = 0.66$, $p = .423$ for CES question 8 (*Having a conversation with someone at a distance (across a room)*).

Table 46: Multivariate Testing Analysis of the Variability of Self-Rated Individual Questions of the Communicative Effectiveness Survey for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p value*
"Group"	8	33	1.83	.106
"Visit"	16	148	2.12	.010
"Group"*"Visit"	16	148	0.41	.977

Note. This table illustrates the main effects of "Group" and "Visit", and the "Group" by "Visit" interaction found from multivariate analyses of the individual questions of the CES for participants with PD and their primary communication partners.

In contrast, the results of the repeated measures MANOVA for the dependent variables of the individual CES questions showed that there was a significant main effect of "Visit" $F(16,148) = 2.12$, $p = .010$. The results of this repeated measures MANOVA can be found in Table 46. A closer look at the factor "Visit" via subsequent univariate testing for each of the individual CES questions revealed the following results: 1) there was no significant univariate effect of "Visit" $F(2,80) = 0.34$, $p = .710$ for CES question 1 (*Having a conversation with a family member or friends at home*); 2) there was a significant univariate effect of "Visit" $F(2,80) = 5.73$, $p = .005$ for CES question 2 (*Participating in conversation with strangers in a quiet place*), which indicated that the marginal mean of the CES question 2 score at visit 1 ($M = 3.12$, $SD = 0.65$) was greater than the mean of the CES question 2 score at visit 2 ($M = 2.88$, $SD = 0.65$) and visit 3 ($M = 2.76$, $SD = 0.71$); 3) there was no significant univariate effect of "Visit" $F(2,80) = 0.57$, $p = .567$ for CES question 3 (*conversing with a familiar person over the telephone*); 4) there was a significant univariate effect of "Visit" $F(2,80) = 3.29$, $p = .043$, which indicated that the marginal mean of the CES question 4 (*Conversing with a stranger over the telephone*) score at visit 1 ($M = 2.76$, $SD = 0.84$) was greater than the mean of the CES question 4 score at visit 2 ($M = 2.45$, $SD = 0.78$) and visit 3 ($M = 2.62$, $SD = 0.84$); 5) there was a significant univariate effect of "Visit" $F(2,80) = 5.09$, $p = .008$, which indicated that the marginal mean of the CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) score at visit 1 ($M = 2.57$, $SD = 0.78$) was significantly greater than the mean of the CES question 5 score at visit 2 ($M = 2.26$, $SD = 0.71$) and visit 3 ($M = 2.21$, $SD = 0.91$); 6) there was a significant univariate effect of "Visit" $F(2,80) = 2.17$, $p = .009$, which indicated that the marginal mean of the CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) score at visit 1 ($M = 2.81$, $SD = 0.91$) was greater than the mean of the CES question 6 score at visit 2 ($M = 2.41$, $SD = 0.78$) and visit 3 ($M = 2.43$, $SD = 0.78$); 7) there was no significant univariate effect of "Visit" $F(2,80) = 0.40$, $p = .670$ for CES question 7 (*Having a conversation while traveling in a car*), and 8) there was no significant univariate effect of "Visit" $F(2,80) = 1.90$, $p = .157$ for CES question 8 (*Having a conversation with someone at a distance (across a room)*). The univariate analysis results of the individual CES question scores are provided in Table 47.

Table 47: Univariate Testing Analysis of the Variability of Self-Rated Individual Questions of the Communicative Effectiveness Survey for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p value*
"Group"				
CES question 1	1	40	5.48	.024
CES question 2	1	40	6.47	.015
CES question 3	1	40	3.65	.063
CES question 4	1	40	1.03	.316
CES question 5	1	40	2.38	.131
CES question 6	1	40	3.68	.062
CES question 7	1	40	7.50	.009
CES question 8	1	40	0.66	.423
"Visit"				
CES question 1	2	80	0.34	.710
CES question 2	2	80	5.73	.005
CES question 3	2	80	0.57	.567
CES question 4	2	80	3.29	.043
CES question 5	2	80	5.09	.008
CES question 6	2	80	2.17	.009
CES question 7	2	80	0.40	.670
CES question 8	2	80	1.90	.157
"Group"*"Visit"				
CES question 1	2	80	0.71	.493
CES question 2	2	80	1.02	.367
CES question 3	2	80	0.57	.567
CES question 4	2	80	0.18	.835
CES question 5	2	80	0.03	.975
CES question 6	2	80	0.35	.707
CES question 7	2	80	0.59	.555
CES question 8	2	80	0.46	.635

Note. This table illustrates the significant and non-significant univariate effects of "Group", significant and non-significant univariate effects of "Visit", and "Group" by "Visit" interactions found from univariate analyses of the individual questions of the CES for participants with PD and their primary communication partners.

Additionally, the results of the repeated measures MANOVA for the dependent variables of the individual CES questions showed that there was no significant "Group" by "Visit" interaction $F(16,148) = 0.41$, $p = .977$. The results of this repeated measures MANOVA can be found in Table 46. The non-significant interactions for CES question 1, 2, 3, 4, 5, 6, 7, and 8 scores can be found in Figures 26, 27, 28, 29, 30, 31, 32, and 33, respectively. Subsequent univariate testing for each of the individual CES questions revealed no

significant “Group” by “Visit” interaction for each of the individual CES questions. These results suggest that self- and proxy-ratings of the CES may be consistent over time for participants with PD and their primary communication partners.

Figure 26: Means of Communicative Effectiveness Survey Question 1 (*Having a Conversation with a Family Member or Friends at Home*) scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson’s Disease Across Visits 1, 2, and 3

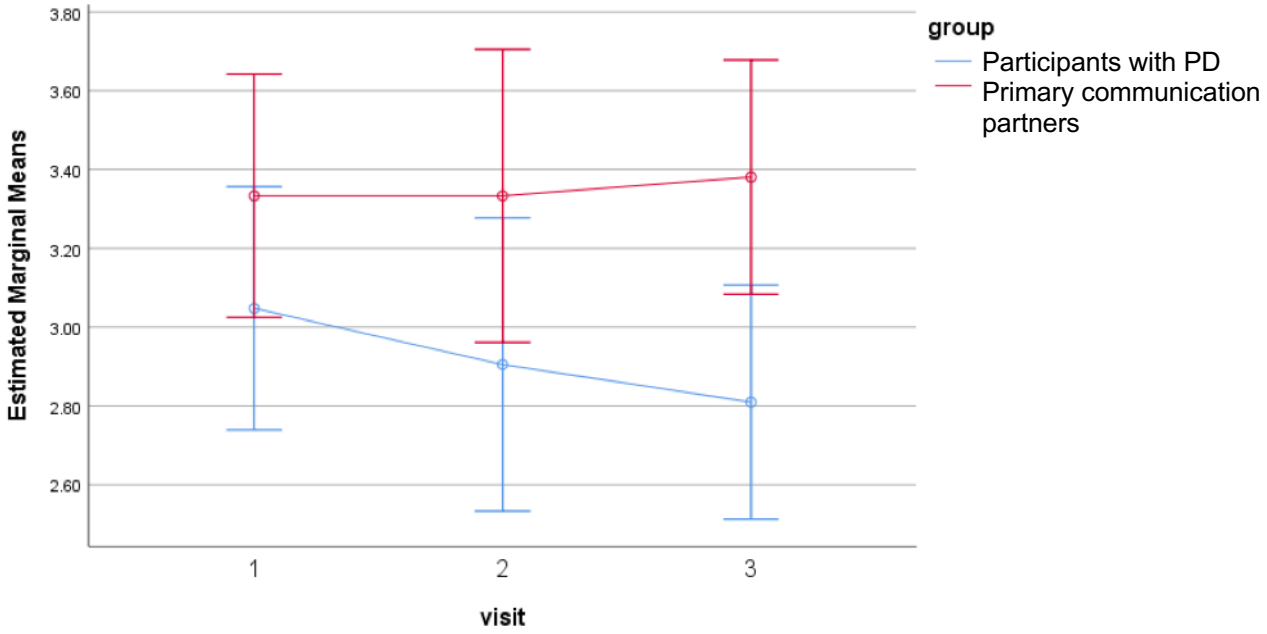

Figure 26. This figure demonstrates the changes in mean CES question 1 (*Having a conversation with a family member or friends at home*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 27: Means of Communicative Effectiveness Survey Question 2 (*Participating in Conversation with Strangers in a Quiet Place*) Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

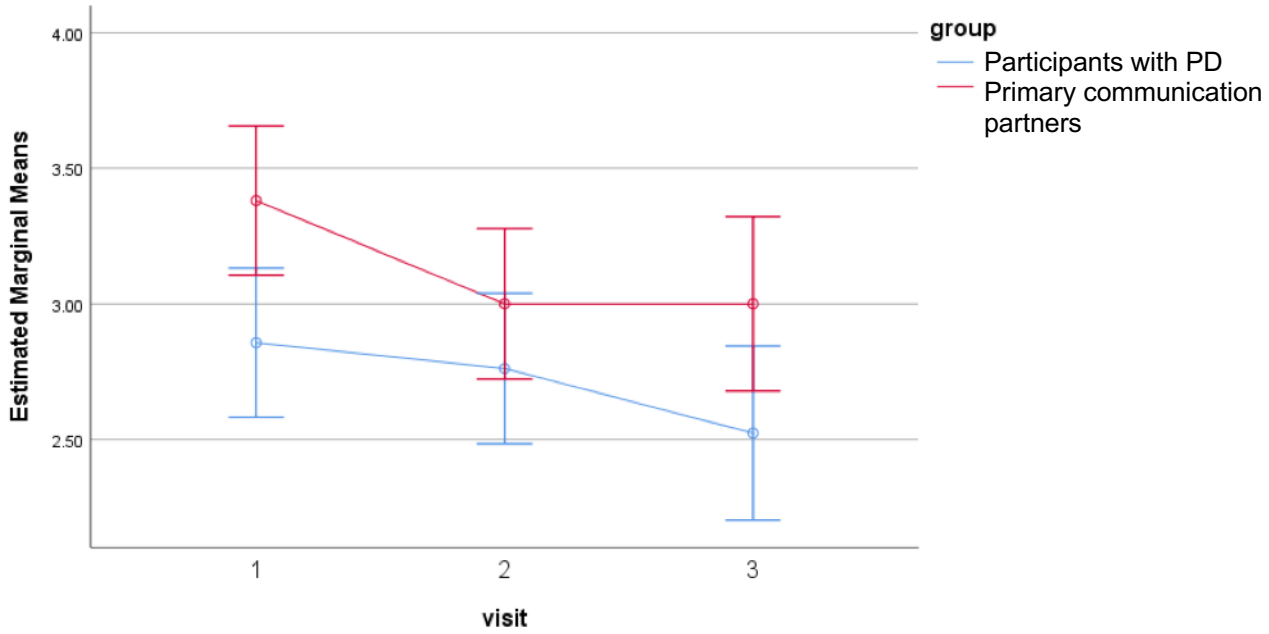

Figure 27. This figure demonstrates the changes in mean CES question 2 (*Participating in conversation with strangers in a quiet place*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 28: Means of Communicative Effectiveness Survey Question 3 (*Conversing with a Familiar Person over the Telephone*) Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

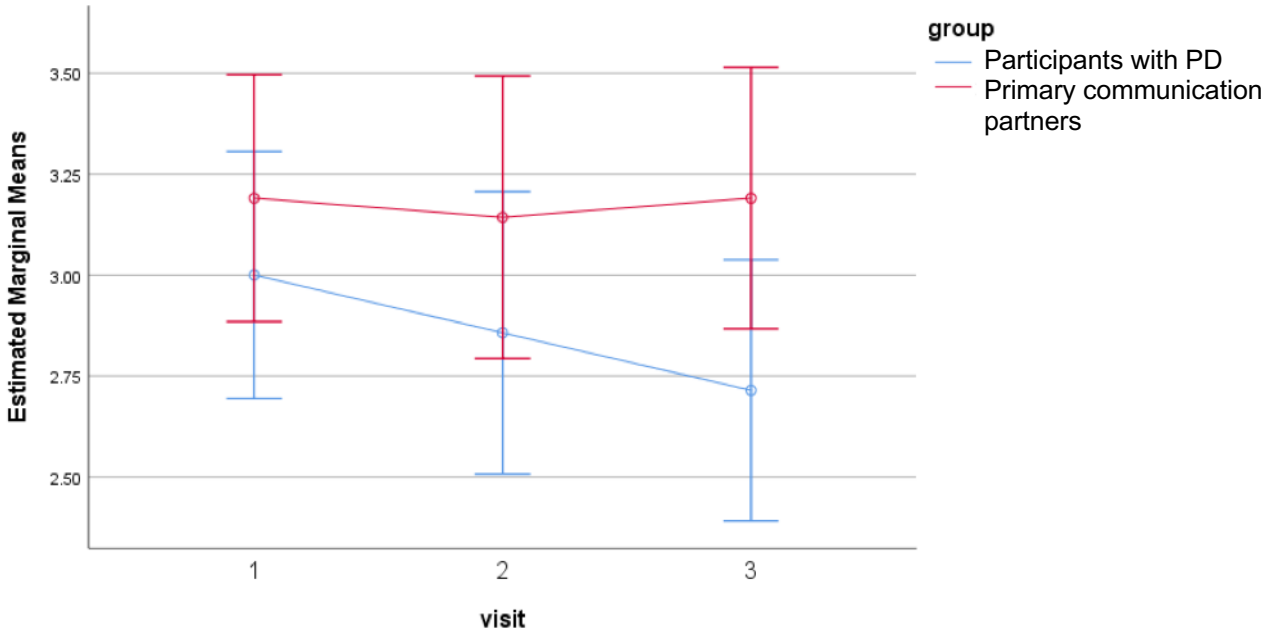

Figure 28. This figure demonstrates the changes in mean CES question 3 (*Conversing with a familiar person over the telephone*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 29: Means of Communicative Effectiveness Survey Question 4 (*Conversing with a Stranger over the Telephone*) Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

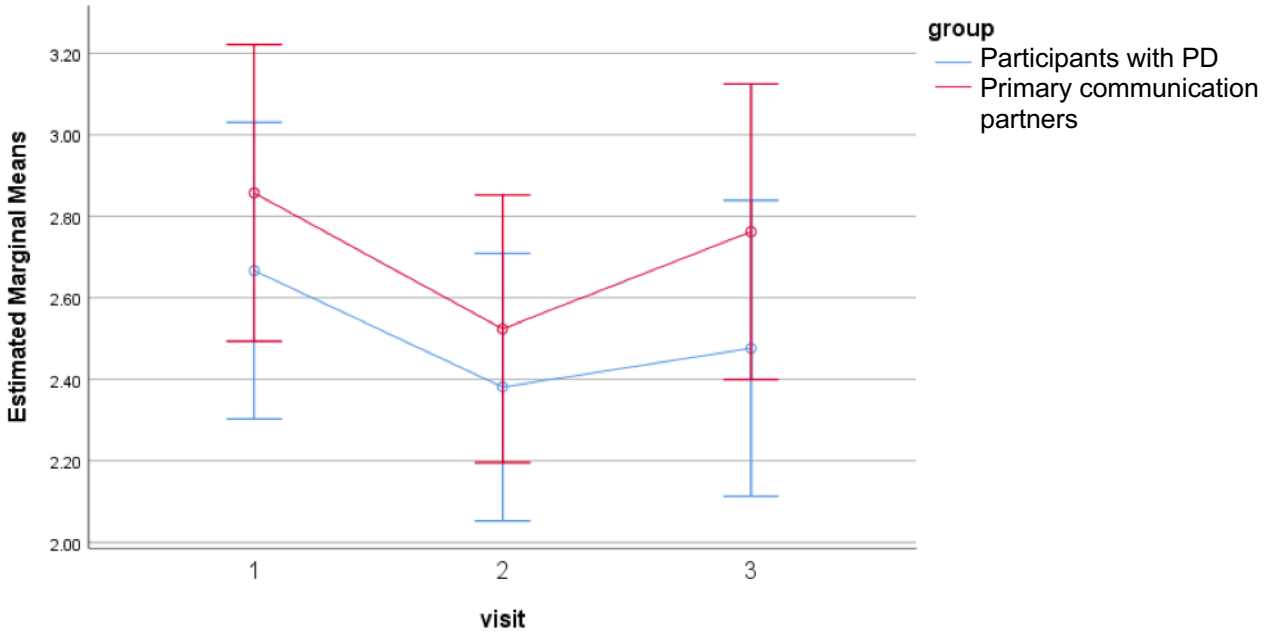

Figure 29. This figure demonstrates the changes in mean CES question 4 (*Conversing with a stranger over the telephone*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 30: Means of Communicative Effectiveness Survey Question 5 *(Being Part of a Conversation in a Noisy Environment (Social Gathering)*) Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

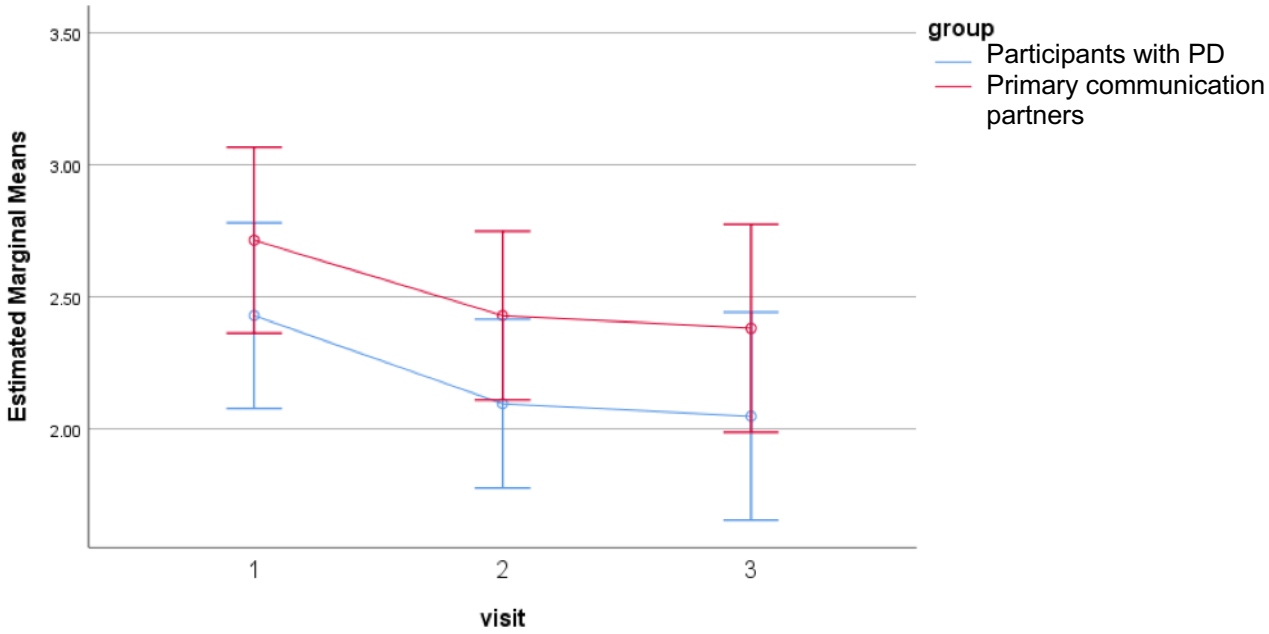

Figure 30. This figure demonstrates the changes in mean CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 31: Means of Communicative Effectiveness Survey Question 6 (*Speaking to a Friend When You are Emotionally Upset or You are Angry*) Scores for Participant with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

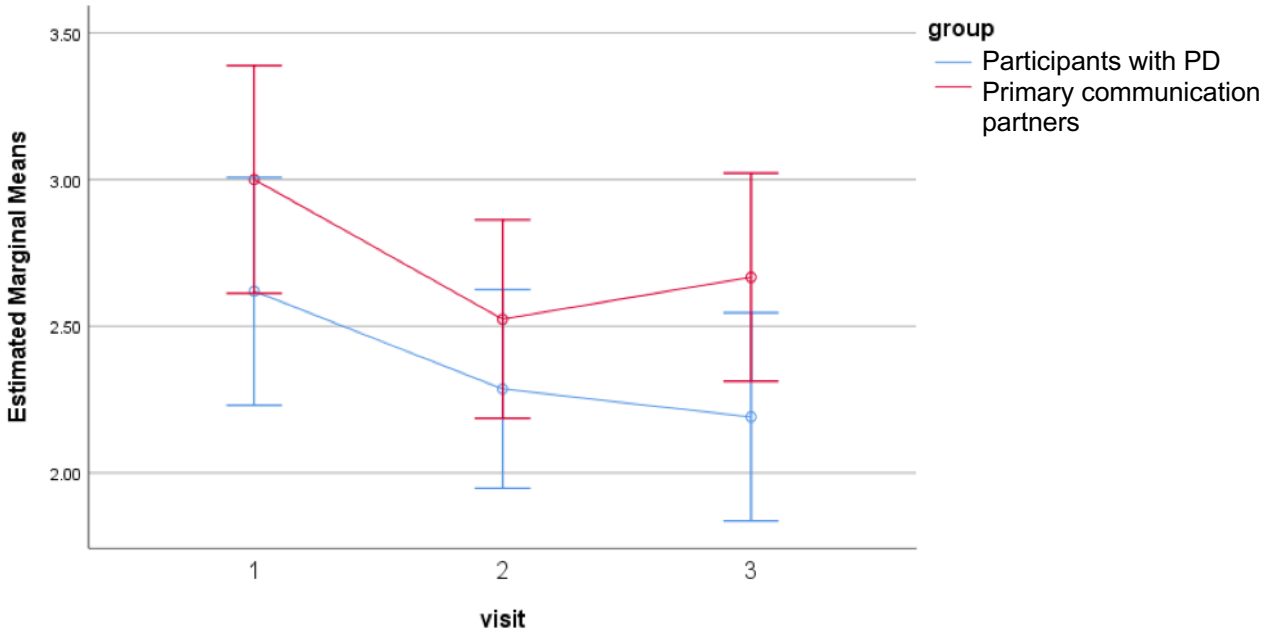

Figure 31. This figure demonstrates the changes in mean CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 32: Means of Communicative Effectiveness Survey Question 7 (*Having a Conversation While Traveling in a Car*) Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, 3

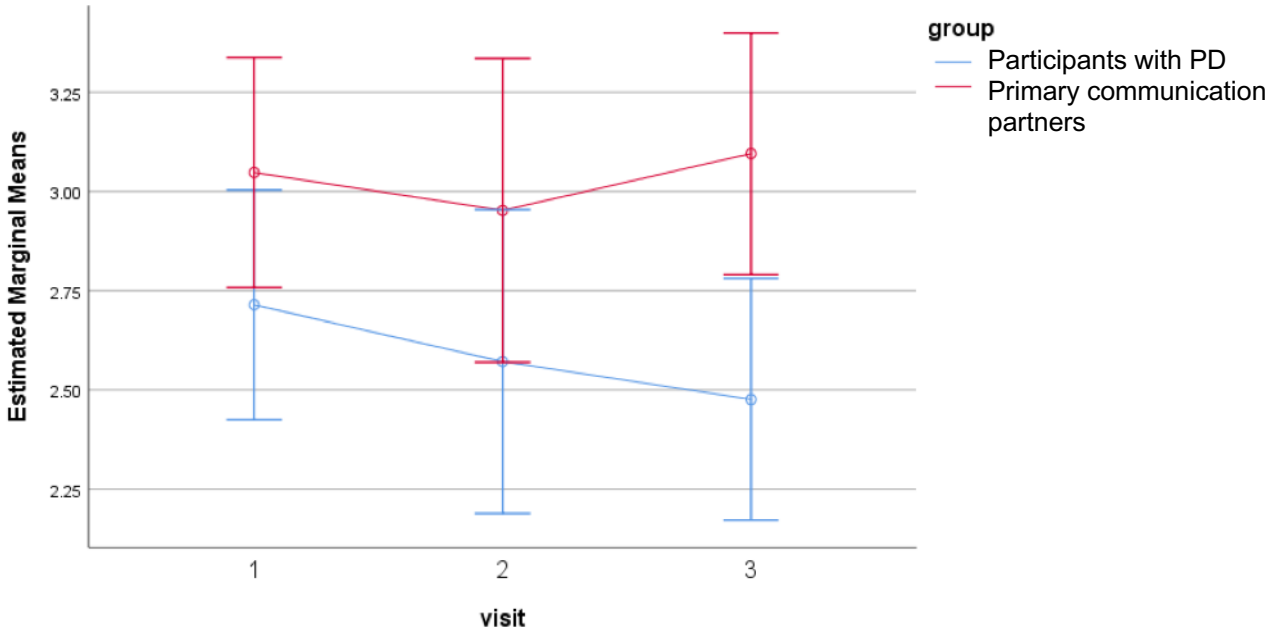

Figure 32. This figure demonstrates the changes in mean CES question 7 (*Having a conversation while traveling in a car*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 33: Means of Communicative Effectiveness Survey Question 8 (*Having a Conversation with Someone at a Distance (Across a Room)*) Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

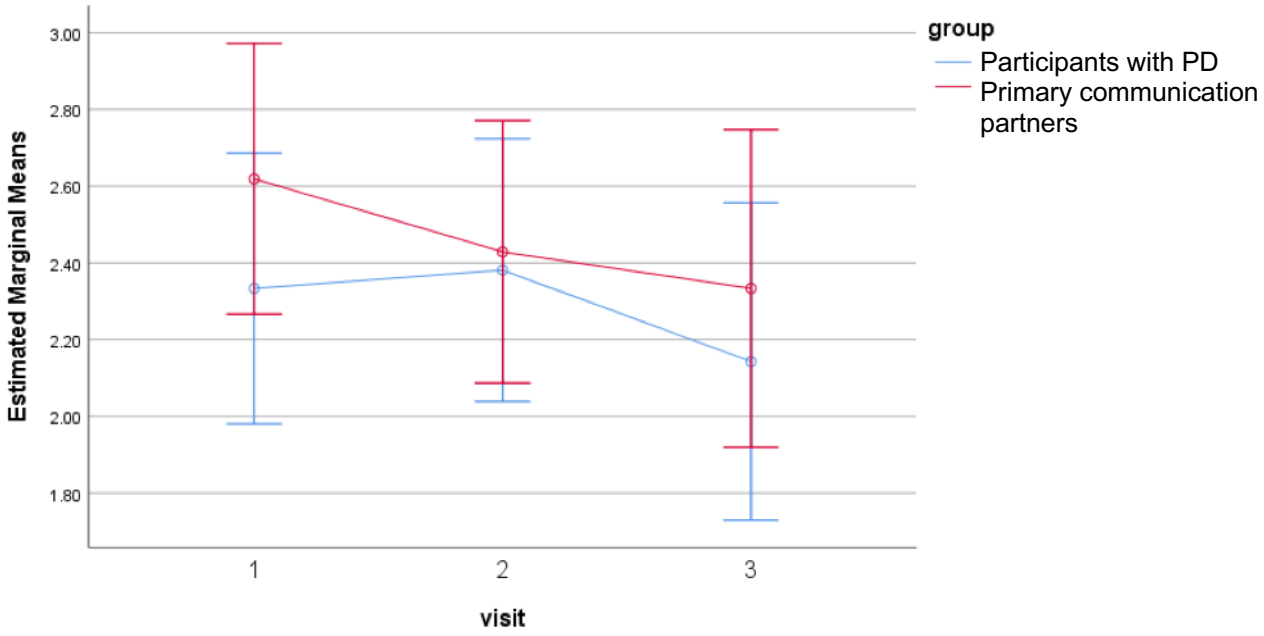

Figure 33. This figure demonstrates the changes in mean CES question 8 (*Having a conversation with someone at a distance (across a room)*) scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

3.7.1.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of individual CES question scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and their primary communication partner.

3.7.1.1.1 Retest analysis of CES question 1 (*Having a conversation with a family member or friends at home*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 1 for participants with PD.

3.7.1.1.2 Retest analysis of CES question 1 (*Having a conversation with a family member or friends at home*) scores for primary communication partners

Results for retest analyses of CES question 1 scores for primary communication partners are summarized in Table 24. The ICC values related to the repeated measurement of CES question 1 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .79, $p < .001$, visit 1 versus visit 3 *ICC* = .46, $p = .090$, visit 2 versus visit 3 *ICC* = .60, $p = .010$. These results suggest that CES question 1 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 1 scores was 0.00 (*SD* = 0.67) and this difference was not significant $t(22) = 0.00$, $p = 1.000$. The mean differences in retest values ranged from -2 – 1. The mean difference between visit 1 and visit 3 for CES question 1 scores was -0.05 (*SD* = 0.90) and this difference was not significant $t(21) = -0.24$, $p = .815$. The mean differences in retest values ranged from -3 – 1. The mean difference between visit 2 and visit 3 for CES question 1 scores was -0.05 (*SD* = 0.90) and this difference was not significant $t(21) = -0.24$, $p = .815$. The mean differences in retest values ranged from -3 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.00 and *CR%* = 30.41%, for visit 1 vs visit 3, *CR* = 1.52 and *CR%* = 45.96%, and for visit 2 vs visit 3, *CR* = 1.49 and *CR%* = 44.34%. Based on the CR, an observed change in the CES question 1 scores of at least 1.00 – 1.52 would suggest a fairly large amount of measurement error ranging from 30.41 – 45.96% variation in the CES question 1 scores

of primary communication partners of IWPD. These results suggest that the measure of CES question 1 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.3 Retest analysis of CES question 2 (*Participating in conversation with strangers in a quiet place*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 2 for participants with PD.

3.7.1.1.4 Retest analysis of CES question 2 (*Participating in conversation with strangers in a quiet place*) scores for primary communication partners

Results for retest analyses of CES question 2 scores for primary communication partners are summarized in Table 25. The ICC values related to the repeated measurement of CES question 2 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .00, *p* = .500, visit 1 versus 3 *ICC* = .33, *p* = .164, visit 2 versus visit 3 *ICC* = .72, *p* = .004. These results suggest that CES question 2 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 2 scores was 0.30 (*SD* = 0.88) and this difference was not significant *t*(22) = 1.67, *p* = .110. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for CES question 2 scores was 0.32 (*SD* = 0.84) and this difference was not significant *t*(21) = 1.78, *p* = .090. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 2 and visit 3 for CES question 2 scores was 0.00 (*SD* = 0.69) and this difference was not significant *t*(21) = 0.00, *p* = 1.000. The mean differences in retest values ranged from -2 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.71 and *CR%* = 51.83%, for visit 1 vs visit 3, *CR* = 1.51 and *CR%* = 50.45%, and for visit 2 vs visit 3, *CR* = 1.05 and *CR%* = 35.00%. Based on the CR, an observed change in the CES question 2 scores of at least 1.05 – 1.71 would suggest a large amount of measurement

error ranging from 35.00 – 51.83% variation in the CES question 2 scores of primary communication partners of IWPD. These results suggest that the measure of CES question 2 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.5 Retest analysis of CES question 3 (*Conversing with a familiar person over the telephone*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 3 for participants with PD.

3.7.1.1.6 Retest analysis of CES question 3 (*Conversing with a familiar person over the telephone*) scores for primary communication partners

Results for retest analyses of CES question 3 scores for primary communication partners are summarized in Table 26. The ICC values related to the repeated measurement of CES question 3 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .33, *p* = .185, visit 1 versus visit 3 *ICC* = .37, *p* = .155, and visit 2 versus visit 3 *ICC* = .71, *p* = .004.. These results suggest that CES question 3 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 3 scores was 0.13 (*SD* = 0.69) and this difference was not significant *t*(22) = 0.21, *p* = .833. The mean differences in retest values ranged from -2 – 3. The mean difference between visit 1 and visit 3 for CES question 3 scores was 0.18 (*SD* = 0.66) and this difference was not significant *t*(21) = 0.00, *p* = 1.000. The mean differences in retest values ranged from -2 – 1. The mean difference between visit 2 and visit 3 for CES question 3 scores was 0.05 (*SD* = 0.72) and this difference was not significant *t*(21) = -0.30, *p* = .771. The mean differences in retest values ranged from -2 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.74 and *CR%* = 54.81%, for visit 1 vs visit 3, *CR* = 1.52 and *CR%* = 48.51% and for visit 2 vs visit 3, *CR* = 1.10 and *CR%* = 34.66%. Based on the CR, an observed change in the CES

question 3 scores of at least 1.10 – 1.74 would suggest a large amount of measurement error ranging from 34.66 – 54.81% variation in the CES question 3 scores of primary communication partners of IWPD. These results suggest that the measure of CES question 3 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.7 Retest analysis of CES question 4 (*Conversing with a stranger over the telephone*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 4 for participants with PD.

3.7.1.1.8 Retest analysis of CES question 4 (*Conversing with a stranger over the telephone*) scores for primary communication partners

Results for retest analyses of CES question 4 scores for primary communication partners are summarized in Table 27. The ICC values related to the repeated measurement of CES question 4 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .61$, $p = .012$, visit versus and visit 3 $ICC = .55$, $p = .040$, and visit 2 versus visit 3 $ICC = .81$, $p < .001$. These results suggest that CES question 4 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because the ICC values across two of three comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 4 scores was 0.30 ($SD = 0.88$) and this difference was not significant $t(22) = 1.67$, $p = .110$. The mean differences in retest values ranged from -1 – 3. The mean difference between visit 1 and visit 3 for CES question 4 scores was 0.09 ($SD = 0.97$) and this difference was not significant $t(21) = 0.44$, $p = .665$. The mean differences in retest values ranged from -2 – 1. The mean difference between visit 2 and visit 3 for CES question 4 scores was -0.23 ($SD = 0.69$) and this difference was not significant $t(21) = -1.56$, $p = .135$. The mean differences in retest values ranged from -2 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 1.45$ and $CR\% = 51.35\%$, for visit 1 vs visit 3, $CR = 1.59$ and $CR\% = 63.07\%$, and for visit 2 vs

visit 3, *CR* = 1.04 and *CR%* = 38.26%. Based on the Cr, an observed change in the CES question 4 scores of at least 1.04 – 1.59 would suggest a large amount of measurement error ranging from 38.26 – 63.07% variation in the CES question 4 scores of primary communication partners of IWPD. These results suggest that the measure of CES question 4 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.9 Retest analysis of CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 5 for participants with PD.

3.7.1.1.10 Retest analysis of CES question 5 scores (*Being part of a conversation in a noisy environment (social gathering)*) for primary communication partners

Results for retest analyses of CES question 5 for primary communication partners scores are summarized in Table 28. The ICC values related to the repeated measurement of CES question 5 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .62, *p* = .013, visit 1 versus visit 3 *ICC* = .75, *p* = .001, and visit 2 versus visit 3 *ICC* = .80, *p* < .001. These results suggest that CES question 5 scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because the across two of three comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 5 scores was 0.22 (*SD* = 0.90) and this difference was not significant *t*(22) = 1.16, *p* = .260. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for CES question 5 scores was 0.32 (*SD* = 0.78) and this difference approached significance *t*(21) = 1.91, *p* = .069. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 2 and visit 3 for CES question 5 scores was 0.05 (*SD* = 0.72) and this difference was not significant *t*(21) = 0.30, *p* = .771. The mean

differences in retest values ranged from -1 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.47 and *CR%* = 55.43%, for visit 1 vs visit 3, *CR* = 1.23 and *CR%* = 50.73%, and for visit 2 vs visit 3, *CR* = 1.08 and *CR%* = 45.69%. Based on the CR, an observed change in the CES question 5 scores of at least 1.08 – 1.47 would suggest a fairly large amount of measurement error with CR% values ranging from 45.69 – 55.43% for the CES question 5 of primary communication partners of IWPD. These results suggest that the measure of CES question 5 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.11 Retest analysis of CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 6 for participants with PD.

3.7.1.1.12 Retest analysis of CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*) scores for primary communication partners

Results for retest analyses of CES question 6 scores for primary communication partners are summarized in Table 29. The ICC values related to the repeated measurement of CES question 6 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .63, *p* = .007, visit 1 versus visit 3 *ICC* = .50, *p* = .056, and visit 2 versus visit 3 *ICC* = .65, *p* = .012. These results suggest that CES question 6 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 6 scores was 0.45 (*SD* = 0.86) and this difference was significant $t(21) = 2.49$, $p = .021$. The mean difference between visit 1 and visit 3 for CES question 6 scores was 0.33 (*SD* = 0.97) and this difference was not significant $t(20) = 1.58$, $p = .130$. The mean differences in retest

values ranged from -2 – 2. The mean difference between visit 2 and visit 3 for CES question 6 scores was -0.14 (*SD* = 0.85) and this difference was not significant $t(20) = -0.77$, $p = .452$. The mean differences in retest values ranged from -2 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.46 and *CR%* = 48.58%, for visit 1 vs visit 3, *CR* = 1.64 and *CR%* = 64.19%, and for visit 2 vs visit 3, *CR* = 1.36 and *CR%* = 50.98%. Based on the CR, an observed change in the CES question 6 scores of at least 1.36 – 1.64 would suggest a large amount of measurement error ranging from 48.58 – 64.19% variation in the CES question 6 scores of primary communication partners of IWPD. These results suggest that the measure of CES question 6 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.13 Retest analysis of CES question 7 (*Having a conversation while traveling in a car*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 7 for participants with PD.

3.7.1.1.14 Retest analysis of CES question 7 (*Having a conversation while traveling in a care*) scores for primary communication partners

Results for retest analyses of CES question 7 scores for primary communication partners are summarized in Table 30. The ICC values related to the repeated measurement of CES question 7 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .34, $p = .174$, visit 1 versus visit 3 *ICC* = .50, $p = .069$, and visit 2 versus visit 3 *ICC* = .43, $p = .110$. These results suggest that CES question 7 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for CES question 7 scores was 0.13 (*SD* = 1.06) and this difference was not significant $t(22) = 0.59$, $p = .560$. The mean differences in retest values ranged from -2 – 3. The mean difference between visit 1 and visit 3 for CES question 7 scores was 0.00 (*SD* = 0.76) and this difference was not

significant $t(21) = 0.00$, $p = 1.000$. The mean differences in retest values ranged from -2 – 1. The mean difference between visit 2 and visit 3 for CES question 7 scores was -0.14 ($SD = 1.04$) and this difference was not significant $t(21) = -0.617$, $p = .544$. The mean differences in retest values ranged from -3 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 1.89$ and $CR\% = 62.15\%$, for visit 1 vs visit 3, $CR = 1.29$ and $CR\% = 44.44\%$, and for visit 2 vs visit 3, $CR = 1.79$ and $CR\% = 57.87\%$. Based on the CR, an observed change in the CES question 7 scores of at least 1.29 – 1.89 would suggest a large amount of measurement error ranging from 44.44 – 62.15% variation in the CES question 7 scores of primary communication partners of IWPD. These results suggest that the measure of CES question 7 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.1.1.15 Retest analysis of CES question 8 (*Having a conversation with someone at a distance (across a room)*) scores for participants with Parkinson's disease

See objective 4 for retest analyses of CES question 8 for participants with PD.

3.7.1.1.16 Retest analysis of CES question 8 (*Having a conversation with someone at a distance (across a room)*) scores for primary communication partners

Results for retest analyses of CES question 8 scores for primary communication partners are summarized in Table 31. The ICC values related to the repeated measurement of CES question 8 scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .65$, $p = .009$, visit 1 versus visit 3 $ICC = .65$, $p = .009$, and visit 2 versus visit 3 $ICC = .80$, $p < .001$. These results suggest that CES question 8 scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because the ICC values across two of three comparisons were below .8075

The mean difference between visit 1 and visit 2 for CES question 8 scores was 0.17 (*SD* = 0.89) and this difference was not significant $t(22) = 0.94$, $p = .357$. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 1 and visit 3 for CES question 8 scores was 0.27 (*SD* = 0.94) and this difference was not significant $t(21) = 1.37$, $p = .186$. The mean differences in retest values ranged from -2 – 2. The mean difference between visit 2 and visit 3 for CES question 8 scores was 0.09 (*SD* = 0.75) and this difference was not significant $t(21) = 0.57$, $p = .576$. The mean differences in retest values ranged from -2 – 1. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.43 and *CR%* = 55.51%, for visit 1 vs visit 3, *CR* = 1.52 and *CR%* = 63.45%, and for visit 2 vs visit 3, *CR* = 1.11 and *CR%* = 47.88%. Based on the CR, an observed change in the CES question 8 scores of at least 1.11 – 1.52 would suggest a large amount of measurement error ranging from 47.88 – 63.45% variation in the CES question 8 scores of primary communication partners of IWPD. These results suggest that the measure of CES question 8 scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2 VAPP subsection scores

Scores for the VAPP *self-perceived voice problem, daily communication, social communication*, and *emotion* subsections were obtained by summing up the converted scores for each question within a subsection. Raw scores for each question were obtained by measuring the location of an individual's response of each VAS. These raw scores were then divided by a factor of 10. VAPP *total* scores were obtained by summing up the total scores of each of the above-mentioned subsections. VAPP *activity limitation* scores were obtained by summing up the converted scores from the first question of each described situation assessing the degree of activity limitation from the *daily communication* and *social communication* subsections. VAPP *participation restriction* scores were obtained by summing up the converted scores from the second question of each described situation assessing the degree of participation restriction from the *daily communication* and *social communication* subsection. Descriptive statistics for VAPP subsection scores can be found in Table 32. The results of the repeated measures

MANOVA for the dependent variables of the individual VAPP subsections showed that there was no significant multivariate effect of "Group" $F(6,36) = 0.43$, $p = .852$. Additionally, the results of the repeated measures MANOVA for the dependent variables of the individual VAPP subsections showed that there was no significant multivariate effect of "Visit" $F(12,156) = 1.58$, $p = .103$ and no significant "Group" by "Visit" interaction $F(12,156) = 0.81$, $p = .640$. The results of this repeated measures MANOVA can be found in Table 48. Subsequent univariate testing for each of the individual VAPP subsections revealed no significant univariate effects of visit and no significant "Group" by "Visit" interactions for each of the individual VAPP subsections. These values are described in Table 49. The non-significant interactions for VAPP *self-perceived voice problem, daily communication, social communication, emotion, activity limitation, participation restriction,* and *total* scores can be found in Figures 34, 35, 36, 37, 38, 39, and 40, respectively. These results suggest that self- and proxy- ratings of the VAPP may be consistent over time for participants with PD and their primary communication partners.

Table 48: Multivariate Testing Analysis of the Variability of Self-Rated Individual Subsections of the Voice Activity and Participation Profile for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p* value
"Group"	6	36	0.43	.852
"Visit"	12	156	1.58	.103
"Group"*"Visit"	12	156	0.81	.640

Note. This table illustrates the main effects of "Group" and "Visit", and the "Group" by "Visit" interaction found from multivariate analyses of the individual questions of the VAPP for participants with PD and control participants.

Table 49: Univariate Testing Analysis of the Variability of Self-Rated Individual Subsections of the Voice Activity and Participation Profile for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

	Df hypothesis	Df error	F value	*p* value
"Group"				
Voice problem	1	41	0.10	.753
Daily communication	1	41	0.10	.758
Social communication	1	41	0.69	.416
Emotion	1	41	0.21	.648
ALS	1	41	0.15	.704
PRS	1	41	0.29	.593
Total	1	41	0.22	.640
"Visit"				
Voice problem	2	82	1.02	.364
Daily communication	2	82	1.85	.164
Social communication	2	82	2.37	.100
Emotion	2	82	0.300	.742
ALS	2	82	1.779	.175
PRS	2	82	1.503	.229
Total	2	82	1.080	.344
"Group"*"Visit"				
Voice problem	2	82	0.413	.663
Daily communication	2	82	0.625	. 538
Social communication	2	82	2.038	.137
Emotion	2	82	1.925	.152
ALS	2	82	0.617	.542
PRS	2	82	1.496	.230
Total	2	82	1.378	.258

Note. This table illustrates the significant and non-significant univariate effects of "Group", significant and non-significant univariate effects of "Visit", and "Group" by "Visit" interactions found from univariate analyses of the individual subsections of the VAPP for participants with PD and their primary communication partners.

Figure 34: Means of Voice Activity and Participation Profile *Self-Perceived Voice Problem* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

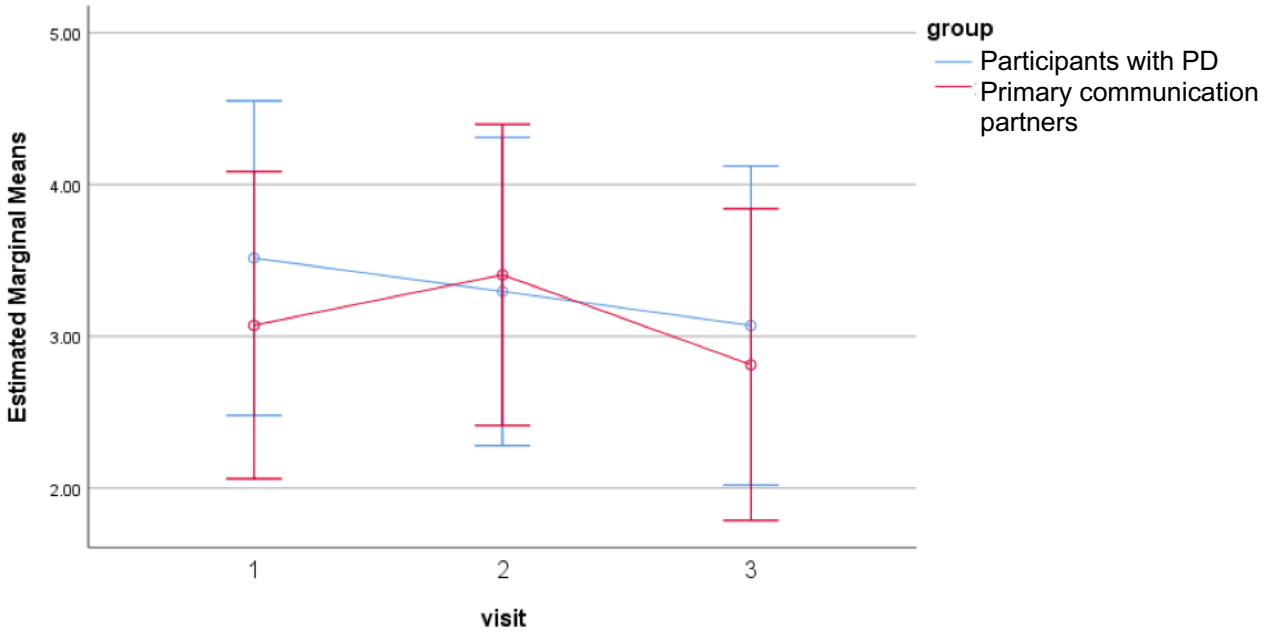

Figure 34. This figure demonstrates the changes in mean VAPP *self-perceived voice problem* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 35: Means of Voice Activity and Participation Profile *Daily Communication* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

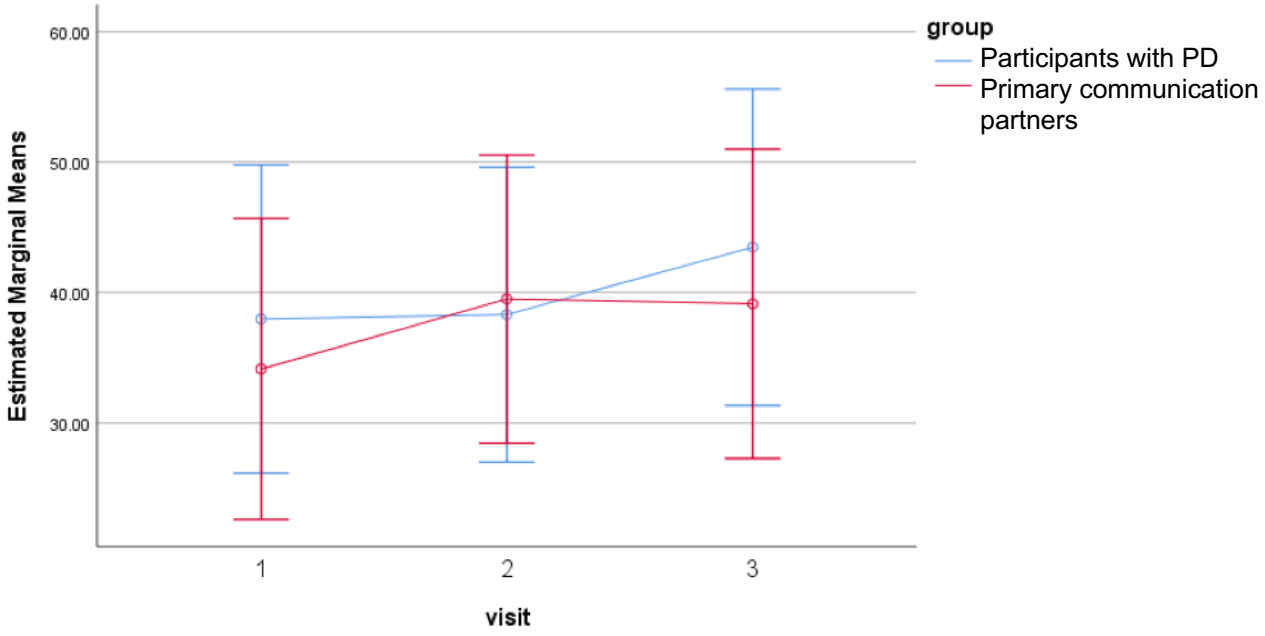

Figure 35. This figure demonstrates the changes in mean VAPP *daily communication* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 36: Means of Voice Activity and Participation Profile *Social Communication* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

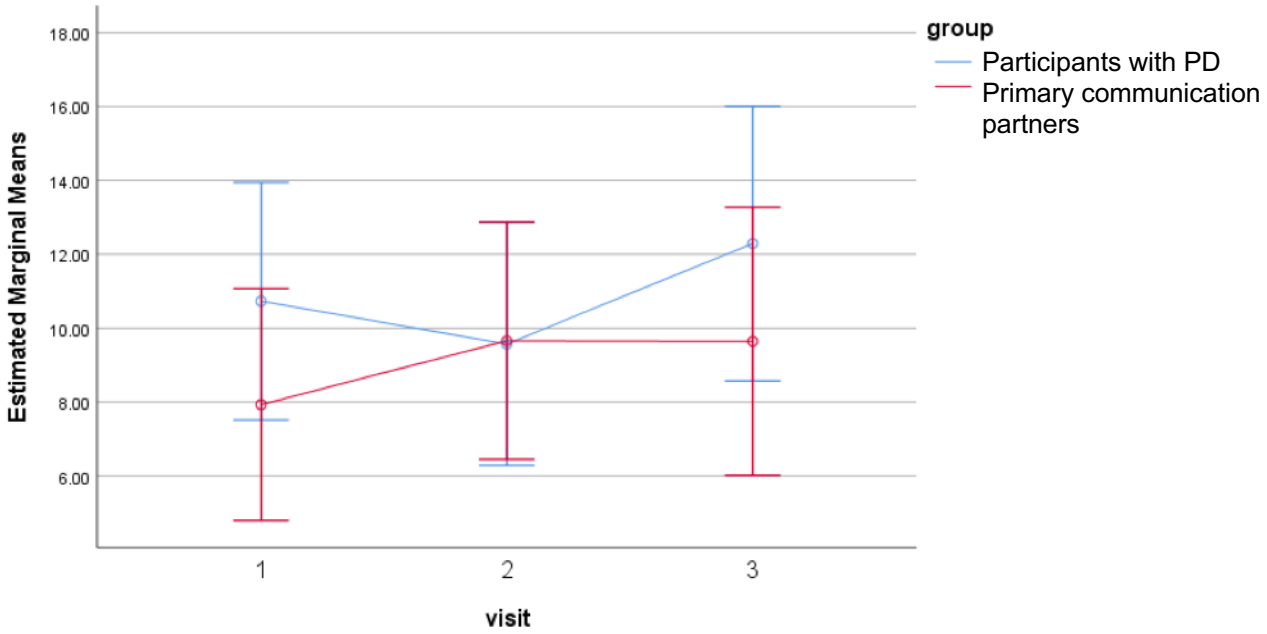

Figure 36. This figure demonstrates the changes in mean VAPP *social communication* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 37: Means of Voice Activity and Participation Profile *Emotion* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

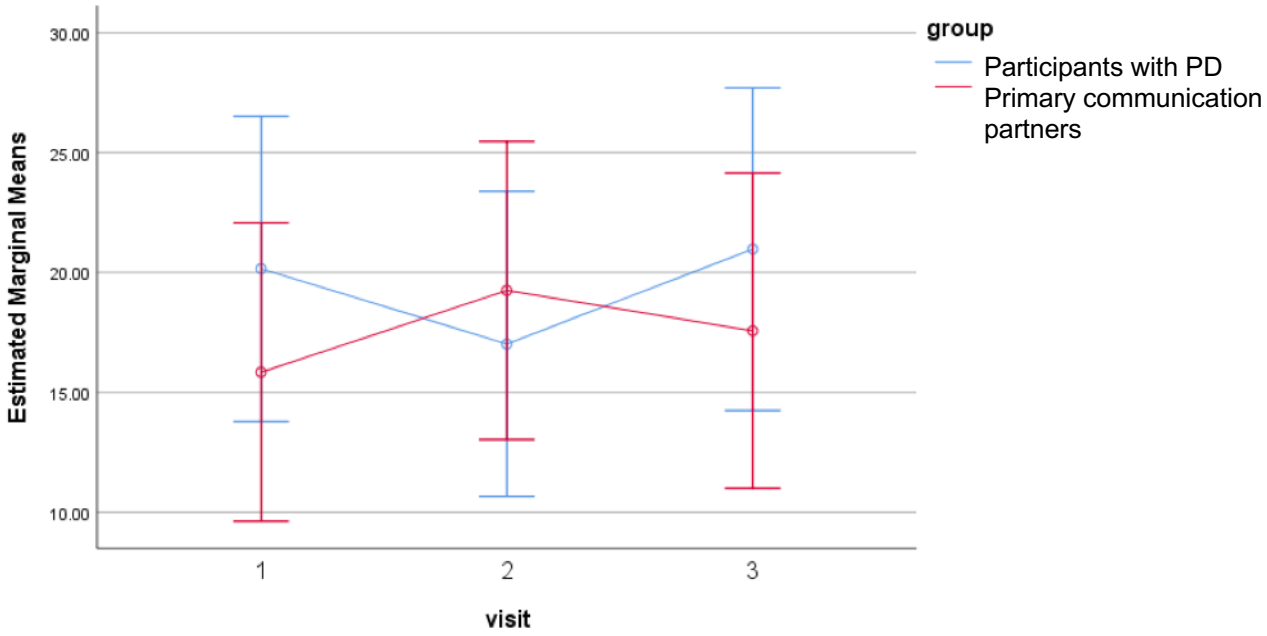

Figure 37. This figure demonstrates the changes in mean VAPP *emotion* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 38: Means of Voice Activity and Participation Profile *Activity Limitation* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

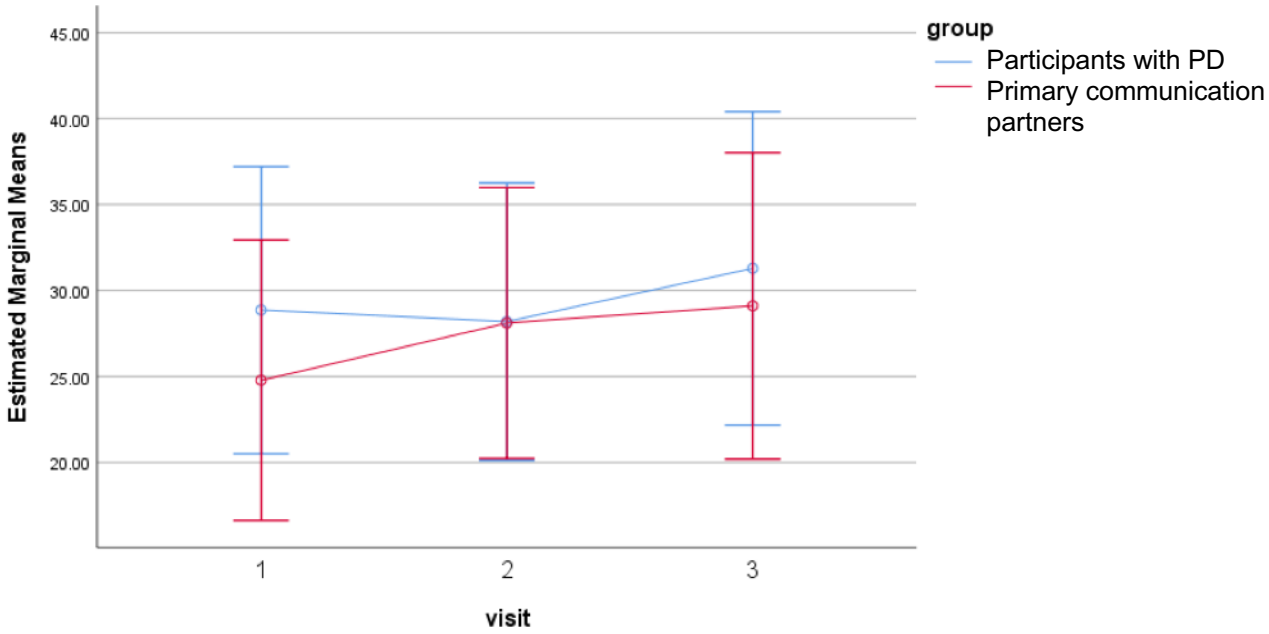

Figure 38. This figure demonstrates the changes in mean VAPP *activity limitation* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 39: Means of Voice Activity and Participation Profile *Participation Restriction* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

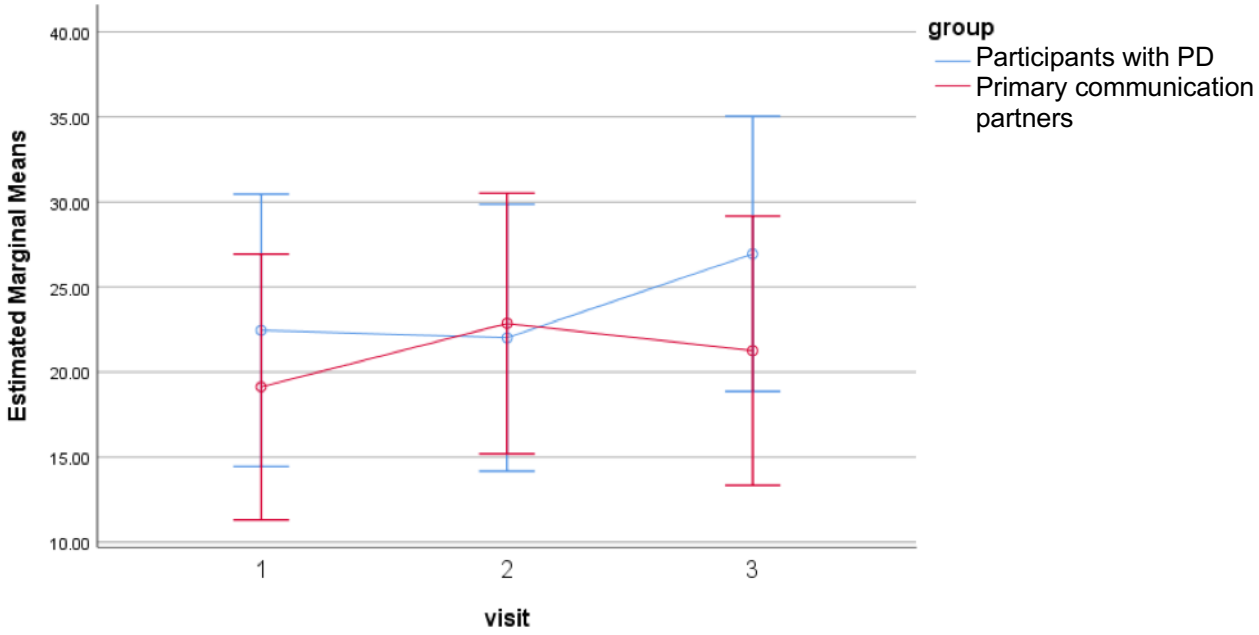

Figure 39. This figure demonstrates the changes in mean VAPP *participation restriction* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

Figure 40: Means of Voice Activity and Participation Profile *Total* Subsection Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

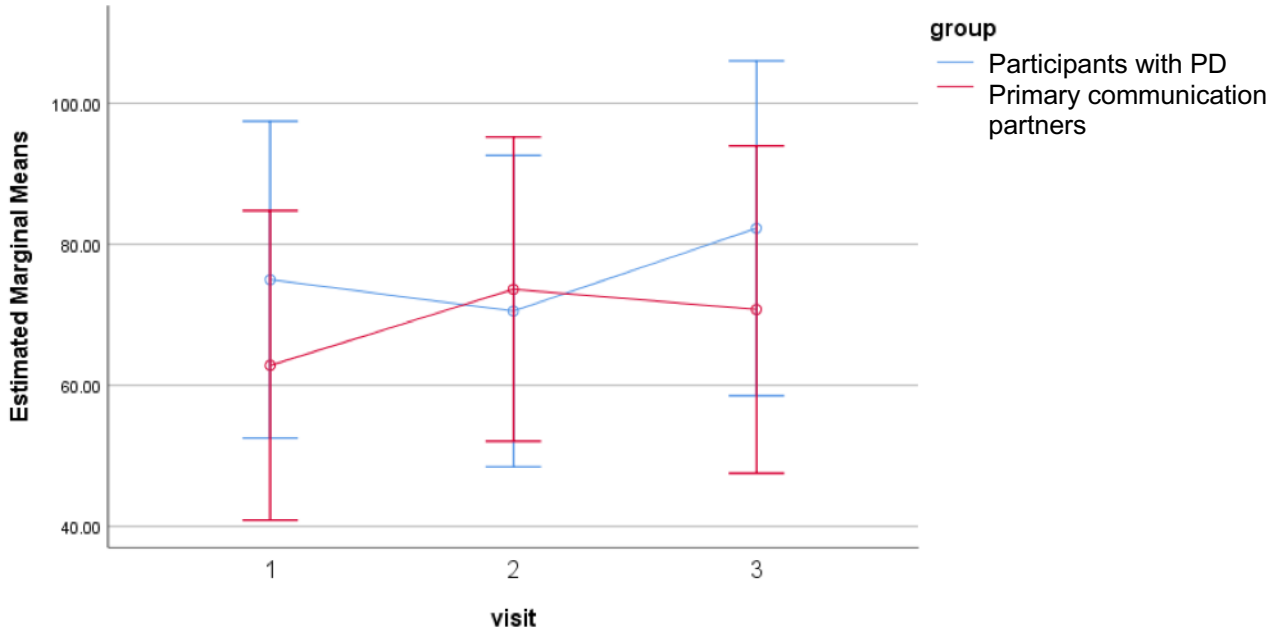

Figure 40. This figure demonstrates the changes in mean VAPP *total* subsection scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

3.7.2.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of individual VAPP subsection scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and their primary communication partner.

3.7.2.1.1 Retest analysis of VAPP *self-perceived voice problem* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *self-perceived voice problem* scores for participants with PD.

3.7.2.1.2 Retest analysis of VAPP *self-perceived voice problem* scores for primary communication partners

Results for retest analyses of the VAPP *self-perceived voice problem* scores are summarized in Table 35. The ICC values related to the repeated measurement of VAPP *self-perceived voice problem* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .80, $p < .001$, visit 1 versus visit 3 *ICC* = .92, $p < .001$, and visit 2 versus visit 3 *ICC* = .80, $p < .001$. These results suggest that VAPP *self-perceived voice problem* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *self-perceived voice problem* scores was -0.40 (*SD* = 1.95) and this difference was not significant $t(22) = -1.00$, $p = .330$. The mean differences in retest values ranged from -7.20 – 3.80. The mean difference between visit 1 and visit 3 for VAPP *self-perceived voice problem* scores was 0.26 (*SD* = 1.25) and this difference was not significant $t(21) = 0.97$, $p = .343$. The mean differences in retest values ranged from -1.40 – 3.60. The mean difference between visit 2 and visit 3 for VAPP *self-perceived voice problem* scores was 0.59 (*SD* = 1.88) and this difference was not significant $t(21) = 1.47$, $p = .156$. The mean differences in retest values ranged from -1.80 – 7.10. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 2.93 and *CR%* = 98.49%, for visit 1 vs visit 3, *CR* = 1.76 and *CR%* = 51.93%, and for visit 2 vs visit 3, *CR* = 2.88 and *CR%* = 102.41%. Based on the CR, an observed change in the VAPP *self-perceived voice problem* scores of at least 1.76 – 2.93 would suggest a large amount of measurement error ranging from 51.93 – 102.41% variation in the VAPP *self-perceived voice problem* of

primary communication partners of IWPD. These results suggest that the measure of VAPP *self-perceived voice problem* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2.1.3 Retest analysis of VAPP *daily communication* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *daily communication* scores for participants with PD.

3.7.2.1.4 Retest analysis of VAPP *daily communication* scores for primary communication partners

Results for retest analyses of the VAPP *daily communication* scores are summarized in Table 36. The ICC values related to the repeated measurement of VAPP *daily communication* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .88$, $p < .001$, visit 1 versus visit 3 $ICC = .90$, $p < .001$, and visit 2 versus visit 3 $ICC = .95$, $p < .001$. These results suggest that VAPP *daily communication* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *daily communication* scores was -5.71 ($SD = 17.84$) and this difference was not significant $t(22) = -1.54$, $p = .139$. The mean differences in retest values ranged from -78.20 – 10.10. The mean difference between visit 1 and visit 3 for VAPP *daily communication* scores was -4.99 ($SD = 3.31$) and this difference was not significant $t(21) = -1.51$, $p = .146$. The mean differences in retest values ranged from -43.80 – 17.10. The mean difference between visit 2 and visit 3 for VAPP *daily communication* scores was 0.36 ($SD = 12.76$) and this difference was not significant $t(21) = 0.13$, $p = .895$. The mean differences in retest values ranged from -25.70 – 34.40. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 26.70$ and $CR\% = 79.39\%$, for visit 1 vs visit 3, $CR = 23.08$ and $CR\% = 58.67\%$, and for visit 2 vs visit 3, $CR = 16.80$ and $CR\% = 42.93\%$. Based on the CR, an observed change in the VAPP *daily communication* scores

of at least 16.80 – 26.70 would suggest a fairly large amount of measurement error ranging from 42.93 – 79.39% variation in the VAPP *daily communication* scores of primary communication partners of IWPD. These results suggest that the measure of VAPP *daily communication* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2.1.5 Retest analysis of VAPP *social communication* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *social communication* scores for participants with PD.

3.7.2.1.6 Retest analysis of VAPP *social communication* scores for primary communication partners

Results for retest analyses of the VAPP *social communication* scores for primary communication partners are summarized in Table 37. The ICC values related to the repeated measurement of VAPP *social communication* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .90$, $p < .001$, visit 1 versus visit 3 $ICC = .92$, $p < .001$, and visit 2 versus visit 3 $ICC = .89$, $p < .001$. These results suggest that VAPP *social communication* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *social communication* scores was -1.81 ($SD = 4.19$) and this difference was significant $t(22) = -2.07$, $p = .050$. The mean difference between visit 1 and visit 3 for VAPP *social communication* scores was -1.71 ($SD = 3.69$) and this difference was significant $t(21) = -2.18$, $p = .041$. The mean difference between visit 2 and visit 3 for VAPP *social communication* scores was 0.01 ($SD = 4.81$) and this difference was not significant $t(21) = 0.01$, $p = .990$. The mean differences in retest values ranged from -10.60 – 9.70. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 6.43$ and $CR\% = 82.12\%$, for visit 1 vs visit 3, $CR = 5.62$ and $CR\% = 58.28\%$, and for visit 2 vs visit 3, $CR = 6.85$ and $CR\% = 70.99\%$. Based on the CR, an observed change in the

VAPP *social communication* scores of at least 5.62 – 6.85 would suggest a fairly large amount of measurement error ranging from 58.28 – 82.12% variation in the VAPP *social communication* scores of primary communication partners of IWPD. These results suggest that the measure of VAPP *social communication* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2.1.7 Retest analysis of VAPP *emotion* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *emotion* scores for participants with PD.

3.7.2.1.8 Retest analysis of VAPP *emotion* scores for primary communication partners

Results for retest analyses of the VAPP *emotion* scores for primary communication partners are summarized in Table 38. The ICC values related to the repeated measurement of VAPP *emotion* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .75$, $p = .001$, visit 1 versus visit 3 $ICC = .93$, $p < .001$, and visit 2 versus visit 3 $ICC = .84$, $p < .001$. These results suggest that VAPP *emotion* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *emotion* scores was -3.50 ($SD = 12.08$) and this difference was not significant $t(22) = -1.39$, $p = .179$. The mean differences in retest values ranged from -43.80 – 15.20. The mean difference between visit 1 and visit 3 for VAPP *emotion* scores was -1.73 ($SD = 6.55$) and this difference was not significant $t(21) = -1.24$, $p = .230$. These results indicate that the mean differences in retest values ranged from -15.6 – 8.6. The mean difference between visit 2 and visit 3 for VAPP *emotion* scores was 1.68 ($SD = 10.65$) and this difference was not significant $t(21) = 0.74$, $p = .468$. The mean differences in retest values ranged from -15.50 – 12.60. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 19.05$ and $CR\% = 122.82\%$, for visit 1 vs visit 3, $CR = 9.53$ and $CR\% = 50.12\%$, and for visit 2 vs visit 3, $CR = 15.64$ and $CR\% = 89.00\%$. Based on the CR, an observed change in the VAPP *emotion* scores of at least

9.53 – 19.05 would suggest a fairly large amount of measurement error ranging from 50.12 – 122.82% variation in the VAPP *emotion* scores of primary communication partners of IWPD. These results suggest that the measure of VAPP *emotion* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2.1.9 Retest analysis of VAPP *activity limitation* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *activity limitation* scores for participants with PD.

3.7.2.1.10 Retest analysis of VAPP *activity limitation* scores for primary communication partners

Results for retest analyses of the VAPP *activity limitation* scores for primary communication partners are summarized in Table 39. The ICC values related to the repeated measurement of VAPP *activity limitation* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .93$, $p < .001$, visit 1 versus visit 3 $ICC = .98$, $p < .001$, and visit 2 versus visit 3 $ICC = .97$, $p < .001$. These results suggest that VAPP *activity limitation* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *activity limitation* scores was -3.54 ($SD = 9.77$) and this difference was not significant $t(22) = -1.74$, $p = .096$. The mean differences in retest values ranged from -42.30 – 7.20. The mean difference between visit 1 and visit 3 for VAPP *activity limitation* scores was -4.33 ($SD = 9.03$) and this difference was significant $t(21) = -2.25$, $p = .036$. The mean difference between visit 2 and visit 3 for VAPP *activity limitation* scores was -1.00 ($SD = 7.43$) and this difference was not significant $t(21) = -0.63$, $p = .537$. The mean differences in retest values ranged from -16.20 – 18.60. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 14.44$ and $CR\% = 59.31\%$, for visit 1 vs visit 3, $CR = 13.51$ and $CR\% = 48.43\%$, and for visit 2 vs visit 3, $CR = 9.72$ and $CR\% = 33.41\%$. Based on the CR, an observed change in the VAPP *activity limitation* scores of

at least 9.72 – 14.44 would suggest a fairly large amount of measurement error ranging from 33.41 – 59.31% variation in the VAPP *activity limitation* scores of primary communication partners of IWPD. These results suggest that the measure of VAPP *activity limitation* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2.1.11 Retest analysis of VAPP *participation restriction* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *participation restriction* scores for participants with PD.

3.7.2.1.12 Retest analysis of VAPP *participation restriction* scores for primary communication partners

Results for retest analyses of the VAPP *participation restriction* scores for primary communication partners are summarized in Table 40. The ICC values related to the repeated measurement of VAPP *participation restriction* score across the three pairwise visits was found to be the following: visit 1 versus visit 2 $ICC = .88$, $p < .001$, visit 1 versus visit 3 $ICC = .84$, $p < .001$, and visit 2 versus visit 3 $ICC = .85$, $p < .001$. These results suggest that VAPP *participation restriction* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *participation restriction* scores was -3.97 ($SD = 11.67$) and this difference was not significant $t(22) = -1.63$, $p = .117$. The mean differences in retest values ranged from -49.50 – 16.00. The mean difference between visit 1 and visit 3 for VAPP *participation restriction* scores was -2.14 ($SD = 12.44$) and this difference was not significant $t(21) = -0.81$, $p = .430$. The mean differences in retest values ranged from -26.80 – 33.30. The mean difference between visit 2 and visit 3 for VAPP *participation restriction* scores was 1.60 ($SD = 12.64$) and this difference was not significant $t(21) = 0.59$, $p = .559$. The mean differences in retest values ranged from -16.40 – 36.80. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, $CR = 17.65$ and $CR\% = 93.57\%$,

for visit 1 vs visit 3, *CR* = 18.12 and *CR%* = 79.38%, and for visit 2 vs visit 3, *CR* = 18.46 and *CR%* = 86.82%. Based on the CR, an observed change in the VAPP *participation restriction* scores of at least 17.65 – 18.46 would suggest a large amount of measurement error ranging from 79.38 – 93.57% variation in the VAPP *participation restriction* scores of primary communication partners of IWPD. These results suggest that the measure of VAPP *participation restriction* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.2.1.13 Retest analysis of VAPP *total* scores for participants with Parkinson's disease

See objective 4 for retest analyses of VAPP *total* scores for participants with PD.

3.7.2.1.14 Retest analysis of VAPP *total* scores for primary communication partners

Results for retest analyses of the VAPP *total* scores for primary communication partners are summarized in Table 41. The ICC values related to the repeated measurement of VAPP *total* scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .88, $p < .001$, visit 1 versus visit 3 *ICC* = .94, $p < .001$, and visit 2 versus visit 3 *ICC* = .92, $p < .001$. These results suggest that VAPP *total* scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for VAPP *total* scores was -11.41 (*SD* = 32.55) and this difference was not significant $t(22) = -1.68$, $p = .107$. The mean differences in retest values ranged from -142.80 – 22.40. The mean difference between visit 1 and visit 3 for VAPP *total* scores was -7.93 (*SD* = 22.36) and this difference was not significant $t(21) = -1.66$, $p = .111$. The mean differences in retest values ranged from -57.4 – 19.6. The mean difference between visit 2 and visit 3 for VAPP *total* scores was 2.87 (*SD* = 27.87) and this difference was not significant $t(21) = 0.48$, $p = .634$. The mean differences in retest values ranged from -41.10 – 85.40. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* =

48.61 and *CR%* = 78.79%, for visit 1 vs visit 3 *CR* = 32.33 and *CR%* = 44.23%, and for visit 2 vs visit 3, *CR* = 39.63 and *CR%* = 56.01%. Based on the CR, an observed change in the VAPP *total* scores of at least 32.33 – 48.61 would suggest a fairly large amount of measurement error ranging from 44.23 – 78.79% variation the VAPP *total* scores of primary communication partners of IWPD. These results suggest that the measure of VAPP *total* scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.7.3 CPIB scores

CPIB scores were obtained by converting the raw total score to a standardized score using the conversion table indicated by Baylor and colleagues (2013). Descriptive statistics for CPIB standardized scores can be found in Table 42. The results of the two-way repeated measures ANOVA for the dependent variable of the CPIB standardized scores showed that there was no significant main effect of "Group" $F(1,42) = 1.18$, $p = .283$ with participants with PD having a similar marginal mean ($M = 52.01$, $SD = 7.83$) compared to their primary communication partners ($M = 54.58$, $SD = 7.83$). In contrast, there was a significant main effect of "Visit" on the CPIB standardized scores $F(2, 84) = 6.33$, $p = .003$. A closer look at the factor "Visit", using pairwise post-hoc analyses indicated that the marginal mean of the CPIB score at visit 1 ($M = 54.85$, $SD = 8.89$) was greater than the marginal mean of the CPIB score at visit 2 ($M = 53.62$, $SD = 7.76$) and visit 3 ($M = 51.40$, $SD = 9.42$). Finally, there was no significant "Group" by "Visit" interaction $F(2,84) = 0.12$, $p = .887$ for the CPIB score. This interaction is illustrated in Figure 41. These results suggest that self-and proxy-rated CPIB scores are consistent over time.

Figure 41: Means of Communicative Participation Item Bank Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

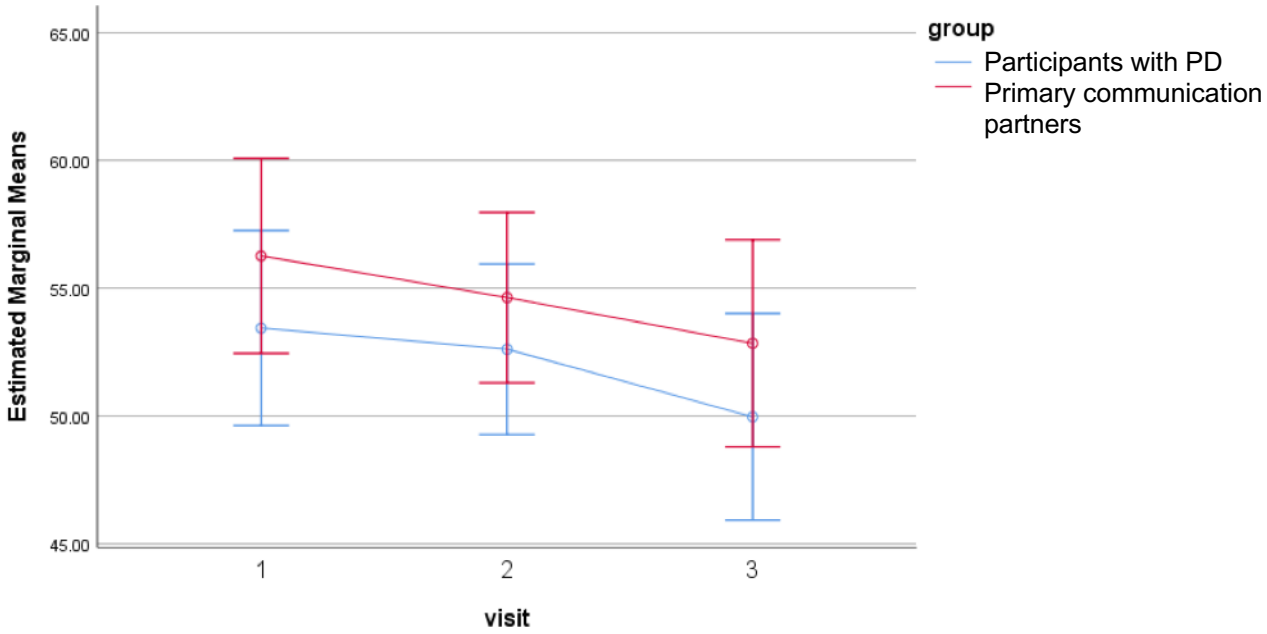

Figure 41. This figure demonstrates the changes in mean CPIB scores for participants with PD and their primary communication across visits. Error bars represent standard deviations.

3.7.3.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of CPIB scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4) pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and their primary communication partner.

3.7.3.1.1 Retest analysis of CPIB scores for participants with Parkinson's disease

See objective 4 for retest analyses of CPIB scores for participants with PD.

3.7.3.1.2 Retest analysis of CPIB scores for primary communication partners

Results for retest analyses of CPIB scores for primary communication partners are summarized in Table 43. The ICC values related to the repeated measurement of CPIB scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .78, $p < .001$, visit 1 versus visit 3 *ICC* = .76, $p = .001$, and visit 2 versus visit 3 *ICC* = .87, $p < .001$. These results suggest that CPIB scores as measured in the present study demonstrated good retest reliability in primary communication partners of IWPD because all of the ICC values across all comparisons were above our criterion of .75.

The mean difference between visit 1 and visit 2 for CPIB scores was 1.44 (*SD* = 7.47) and this difference was not significant $t(22) = 0.93$, $p = .364$. The mean differences in retest values ranged from -11.40 – 22.60. The mean difference between visit 1 and visit 3 for CPIB scores was 3.42 (*SD* = 8.48) and this difference was not significant $t(21) = 1.89$, $p = .072$. The mean differences in retest values ranged from -10.20 – 35.00 The mean difference between visit 2 and visit 3 for CPIB scores was 1.79 (*SD* = 6.54) and this difference was not significant $t(21) = 1.29$, $p = .213$. The mean differences in retest values ranged from -10.20 – 18.30. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 11.38 and *CR%* = 20.23%, for visit 1 vs visit 3, *CR* = 13.43 and *CR%* = 24.50%, and for visit 2 vs visit 3, *CR* = 9.72 and *CR%* = 18.40%. Based on the CR, an observed change in the CPIB scores of at least 9.72 – 13.43 would suggest the possibility of an acceptable amount of measurement error ranging from 18.40 – 24.50% variation in the CPIB scores of primary communication partners of IWPD. These results suggest that the measure of CPIB scores demonstrates marginal repeatability for primary communication partners of IWPD.

3.7.4 LSUS scores

LSUS scores were obtained based on whether participants indicated that the first, second, third, fourth, or fifth categories was indicative of their typical level of speech usage. Descriptive statistics for LSUS scores can be found in Table 44. The results of the two-way repeated measures ANOVA for the dependent variable of the LSUS score showed

that there was no significant main effect of "Group" $F(1,42) = 0.00$, $p = 1.00$. Additionally, there was no significant main effect of "Visit" $F(2,84) = 0.57$, $p = .568$ and no significant "Group" by "Visit" interaction $F(2,84) = 2.25$, $p = .112$ for the LSUS score. This interaction is illustrated in Figure 42. These results suggest that participants with PD and their primary communication partners are consistent in their perceptions of typical level of speech usage, and those ratings do not appear to be fluctuate over time.

Figure 42: Means of Level of Speech Usage Scale Scores for Participants with Parkinson's Disease and Primary Communication Partners of Participants with Parkinson's Disease Across Visits 1, 2, and 3

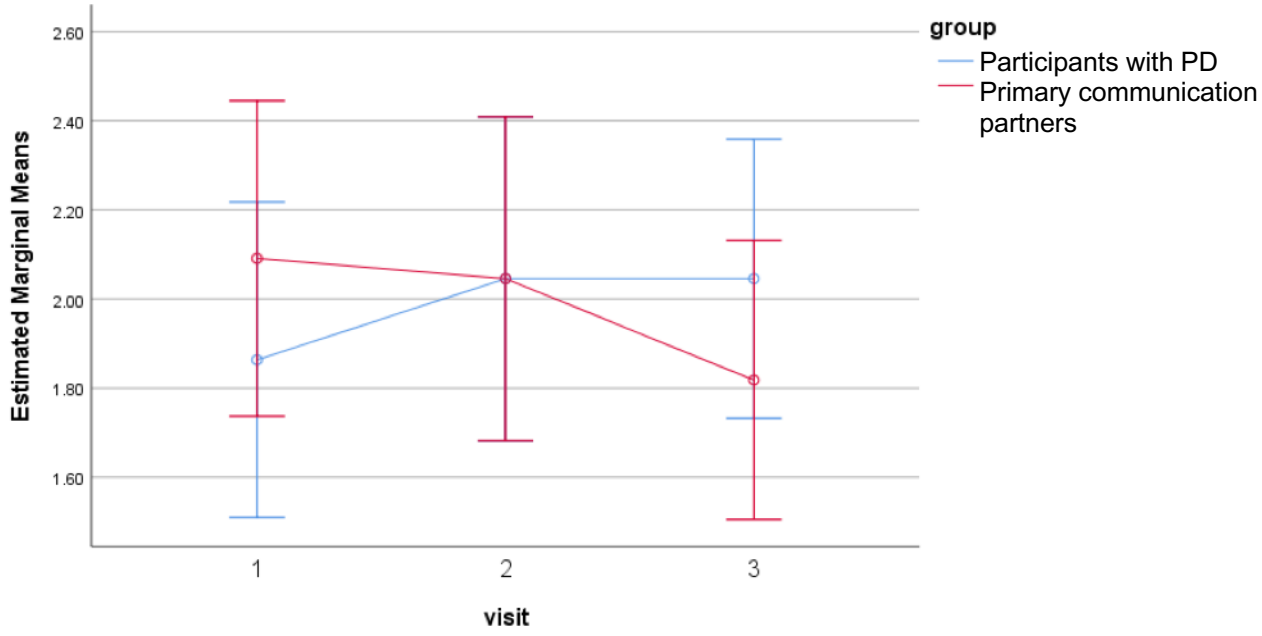

Figure 42. This figure demonstrates the changes in mean LSUS scores for participants with PD and their primary communication partners across visits. Error bars represent standard deviations.

3.7.4.1 Retest analyses

Following the primary statistical analyses described above, secondary retest analyses evaluated the retest reliability and repeatability of LSUS scores. These analyses were performed via: 1) correlations (ICC) between visits 1-2, 1-3 and 2-3, 2) SEM across visits 1-2, 1-3, and 2-3, 3) mean difference t-tests comparing visits 1-2, 1-3, and 2-3, 4)

pairwise CR involving visits 1-2, 1-3, and 2-3, and 5) CR% involving visits 1-2, 1-3, and 2-3. Separate retest analyses were performed on the participants with PD and their primary communication partner.

3.7.4.1.1 Retest analysis of LSUS scores for participants with Parkinson's disease

See objective 4 for retest analyses of LSUS scores for participants with PD.

3.7.4.1.2 Retest analysis of LSUS scores for primary communication partners

Results for retest analyses of LSUS scores for primary communication partners are summarized in Table 45. The ICC values related to the repeated measurement of LSUS scores across the three pairwise visits was found to be the following: visit 1 versus visit 2 *ICC* = .70, $p = .004$, visit 1 versus visit 3 *ICC* = .60, $p = .018$, and visit 2 versus visit 3 *ICC* = .80, $p < .001$. These results suggest that LSUS scores as measured in the present study did not demonstrate good retest reliability in primary communication partners of IWPD because the ICC values across 2 of 3 comparisons were below our criterion of .75.

The mean difference between visit 1 and visit 2 for LSUS scores was 0.09 (*SD* = 0.90) and this difference was not significant $t(22) = 0.46$, $p = .648$. The mean differences in retest values ranged from -2 – 2. The mean difference between visit 1 and visit 3 for LSUS scores was 0.27 (*SD* = 0.88) and this difference was not significant $t(21) = 1.45$, $p = .162$. The mean differences in retest values ranged from -1 – 2. The mean difference between visit 2 and visit 3 for LSUS scores was 0.23 (*SD* = 0.69) and this difference was not significant $t(21) = 1.56$, $p = .135$. The mean differences in retest values ranged from -1 – 2. Additionally, the following CR values were obtained for the following visit comparisons: for visit 1 vs visit 2, *CR* = 1.40 and *CR%* = 65.89%, for visit 1 vs visit 3, *CR* = 1.45 and *CR%* = 71.32%, and for visit 2 vs visit 3, *CR* = 1.04 and *CR%* = 56.90%. Based on the CR, an observed change in the LSUS scores of at least 1.04 – 1.45 would suggest a large amount of measurement error ranging from 56.90 – 71.32% variation in the LSUS scores of primary communication partners of IWPD. These results suggest that

the measure of LSUS scores demonstrates unacceptable repeatability for primary communication partners of IWPD.

3.8 Statistical Analysis for Objective 7: Inter-Relationships Among Variables in Participants with *Parkinson's Disease*

In order to answer the question '*Are measures of speech intensity, speech intelligibility, communicative participation, demographic factors, and non-speech factors related to one another in participants with PD?',* a matrix of inter-correlations was obtained via a series of three Pearson correlations ($p < .05$) applied to all possible pairwise combination of the experimental variables, demographic factors, and non-speech factors for participants with PD across each of the three experimental visits. The measures of speech intensity included in these analyses were: habitual speech intensity, maximum speech intensity, Lombard response function, magnitude production, and self-perception of typical speech loudness. The measures of speech intelligibility included in these analyses were: SIT transcription score, SIT VAS score, and conversational intelligibility VAS score. The measures of communicative participation included in these analyses were: CES questions 1 through 8, VAPP *self-perceived voice problem* score, VAPP *daily communication* score, VAPP *social communication* score, VAPP *emotion* score, VAPP *activity limitation* score, VAPP *participation restriction* score, CPIB score, VAPP *total* scores, and LSUS score. The measures of demographic factors included in these analyses were: age, gender, and disease duration. The measures of non-speech factors included in these analyses were: GDS scores, disease severity as measured by overall UPDRS scores, and MOCA scores. Only significant correlations, above $r \leq .50$, across the three visits were included and are presented below. This criterion value of $r \leq .50$ was chosen as the minimum threshold since correlations greater than .50 are indicative of a moderate to a very strong relationship between variables (Mukaka, 2012). Since a primary goal of this study is to investigate the consistency of various measures related to hypophonia and communication in IWPD, focusing on significant correlations that minimally had a moderate relationship across all three experimental visits permitted us to identify the most salient variables related to each other. Results of all analyzed correlations are summarized in Tables 50, 51, and 52 for visits 1, 2, and 3 respectively.

Table 50: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Participants with Parkinson's Disease for Visit 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. Habitual speech intensity	–	.61**	.17	-.16	-.07	-.24	-.47*	-.25	-.18	-.15	.08	-.02	.00	-.16	-.04	.10	-.24	.25	.24	.36	.29	. 31	. 33	.28	-.06	.14	-.41*	.06	-.18	-.02	-.21	-.25
2. Maximum speech intensity		–	.56**	.60**	-.26	.14	.01	.04	-.23	.00	-.04	.00	-.02	-.04	.17	-.09	-.04	.04	.05	.01	.04	.05	.10	. 01	.10	-.01	-.12	-.18	-.39t	-.15	.21	-.30
3. Lombard response function			–	.58**	-.08	.30	.27	.23	-.30	-.06	.09	.10	-.13	-.09	-.15	-.31	.04	-.26	-.14	-.20	-.29	-.22	-.10	-.25	.03	-.30	.29	-.29	-.10	-.03	.26	-.17
4. Magnitude production				–	-.21	.42	.48*	.34*	-.13	.14	-.05	.12	.08	.01	.09	-.16	.27	-.24	-.27	-.44	-.32	-.31	-.27	-.32*	.13	-.12	.33	-.27	-.26	-.12	.53	-.11**
5. Typical speech loudness					–	-.42*	-.22	.06	.20	.22	.30	.18	-.07	.02	.12	.16	.13	-.44*	-.32	-.17	-.35	-.33	-.39t	-.21	.00	.03	.37	.18	.38	.15	-.19	.03
6. SIT transcription						–	.85**	.28	-.02	.09	-.17	-.06	-.02	.07	-.02	.13	-.05	.02	.03	.01	.01	.03	.04	.03	.22	-.21	.35	-.41	-.03	-.24	.52*	.34
7. SIT VAS							–	.36	.15	.15	-.14	-.15	-.08	-.05	-.04	-.14	-.11	-.05	-.02	-.09	-.09	-.07	-.05	-.05	.24	-.23	.60**	-.39t	-.03	-.14	.65**	.31
8. Conversational intelligibility								–	-.22	-.17	-.25	.02	-.06	.15	.07	-.13	-.03	-.21	-.16	-.10	-.14	-.12	-.16	-.04	.13	-.39t	.29	-.41t	.25	-.03	.09	.01
9. CES Q1									–	.53**	.37	.10	.00	.34	.32	.12	.16	.00	.03	-.16	.01	-.01	-.03	.00	.51*	.60**	.16	-.05	-.47*	.06	.26	.12
10. CES Q2										–	.53*	.46*	.18	.59**	.52*	.22	.47*	-.50*	-.52*	-.53**	-.48*	-.51*	-.57**	-.43*	.23	.26	.16	.17	-.32	-.06	.03	.32
11. CES Q3											–	.59**	-.04	.46*	.24	-.04	.43*	-.49*	-.50*	-.55**	-.43*	-.46*	-.45*	-.45*	.11	.27	-.04	-.07	-.21	-.29	-.20	-.04
12. CES Q4												–	.49*	.56**	.30	.41t	.49*	-.43*	-.52*	-.58**	-.45*	-.49*	-.53**	-.43*	.02	.12	-.07	-.19	-.22	-.48*	-.04	.09
13. CES Q5													–	.40t	.27	.46*	.17	.19	.08	-.13	-.03	.01	-.09	.10	-.19	.20	-.02	.17	-.32	-.33	.07	.46*
14. CES Q6														–	.65**	.32	.49*	-.30	-.32	-.44*	-.29	-.32	-.39t	-.25	.15	.31	-.10	-.07	-.27	-.24	-.13	.40t
15. CES Q7															–	.32	.43*	-.36	-.37	-.31	-.43*	-.41t	-.46*	-.31	.04	.33	-.12	-.22	-.08	.03	-.17	.06
16. CES Q8																–	.39t	-.10	-.17	-.05	-.08	-.12	-.25	-.03	-.14	.23	-.17	.04	-.06	-.12	-.06	.27
17. CPIB																	–	-.69**	-.74**	-.83**	-.68**	-.76**	-.76**	-.76**	-.19	.32	-.01	.04	-.10	.21	-.26	.19
18. VAPP voice problem																		–	.94**	.76**	.93**	.95**	.92**	.90**	.10	.11	-.26	.11	-.32	-.19	.24	-.07
19. VAPP daily communication																			–	.82**	.92**	.96**	.97**	.92**	.20	.14	-.21	.04	-.28	-.04	.20	-.08
20. VAPP social communication																				–	.78**	.85**	.83**	.87**	-.02	-.06	-.26	-.03	.14	.05	.05	-.23
21. VAPP emotion																					–	.98**	.94**	.92**	.25	.08	-.35	.11	-.29	-.14	.15	-.13
22. VAPP total																						–	.97**	.97**	.24	.09	-.29	.05	-.24	-.12	.19	-.15

23. VAPP activity limitation																								–	.90**	.25	.10	-.28	-.01	-.24	-.05	.18	-.18
24. VAPP participation restriction																									–	.22	.07	-.22	.03	-.18	-.16	.20	-.13
25. LSUS																										–	.16	.08	-.18	-.47*	-.25	.39t	-.04
26. Age																											–	-.18	-.14	-.45*	.21	-.05	.03
27. Gender																												–	-.03	.15	-.03	.54**	.36
28. Disease duration																													–	-.06	.20	-.29	.37
29. GDS																														–	.22	-.37	-.03
30. UPDRS																															–	-.34	-.05
31. MOCA																																–	.00
32. Medication effectiveness																																	–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

Table 51: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Participants with Parkinson's Disease for Visit 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. Habitual speech intensity	–	.69**	-.22	.09	.09	.01	-.09	-.02	-.10	.05	-.18	-.19	-.04	-.43*	.20	-.21	.08	.30	.03	.25	.12	.15	.10	.15	-.02	.23	-.27	-.08	-.13	.00	-.04	-.08
2. Maximum speech intensity		–	-.09	.72**	.11	.09	-.05	.24	-.02	.29	.03	-.05	-.11	-.23	.19	-.18	.42*	.03	-.21	-.17	-.24	-.18	-.16	-.17	-.12	.12	-.09	-.34	.34	-.08	.23	-.29
3. Lombard response function			–	-.02	-.32	.22	.20	.20	.42*	.11	.38	.21	.01	-.02	.24	.01	-.04	-.27	-.13	-.13	-.21	-.18	-.14	-.15	-.04	.09	.50*	-.56**	.10	-.03	.39t	.05
4. Magnitude production				–	.04	.02	-.10	.21	.09	.51	.34*	.18	-.01	.08	.21	.01	.58	-.17**	-.29	-.44	-.38*	-.35	-.27	-.35	-.32	.06	.11	-.29	.35	-.01	.18	-.25
5. Typical speech loudness					–	-.34	-.37	.16	-.04	.01	-.27	-.41*	-.15	-.10	-.37	-.18	.12	.00	-.02	-.20	-.08	-.04	-.12	.05	.30	.19	.04	.24	-.24	-.06	.00	-.23
6. SIT transcription						–	.76**	.45*	.12	-.11	.31	.28	.55**	.14	.04	.16	.10	-.30	-.18	-.01	-.18	-.15	-.17	-.04	.18	-.20	.41t	-.30	.10	-.26	.52*	.26
7. SIT VAS							–	.52*	.08	-.24	.08	.24	.36	-.09	-.16	.16	-.03	-.45*	-.09	.02	-.06	-.09	-.03	-.09	.20	-.38	.35	-.08	.03	-.02	.44*	.39t
8. Conversational intelligibility								–	-.09	-.17	.01	.10	.22	-.08	-.12	.20	.24	-.66**	-.34	-.52*	-.37	-.38	-.37	-.30	.32	-.38	.28	-.17	-.03	-.30	.31	.07
9. CES Q1									–	.68**	.53**	.21	.31	.17	.50*	.09	.12	-.09	-.02	.02	-.04	-.02	.01	-.02	.15	.40t	.46*	-.17	-.45*	-.08	.54**	.26
10. CES Q2										–	.48*	.25	.16	.30	.58**	.09	.49*	-.10	-.24	-.24	-.28	-.22	-.13	-.26	.00	.45*	.08	-.16	-.49*	-.13	.21	.00
11. CES Q3											–	.47*	.60**	.30	.58**	.31	.37	-.21	-.27	-.22	-.30	-.32	-.34	-.29	-.36	.15	.54**	.01	-.16	-.12	.35	.40t
12. CES Q4												–	.50*	.28	.54**	.59**	.33	-.50*	-.54**	-.39t	-.37	-.50*	-.48*	-.53**	-.42*	-.10	.04	-.22	.15	-.25	.06	.03
13. CES Q5													–	.22	.29	.46*	.13	-.22	-.18	-.03	-.09	-.13	-.22	-.05	.00	.01	.29	.27	-.17	-.31	.27	.34
14. CES Q6														–	-.03	.39t	41t	-.12	-.26	-.27	-.15	-.25	-.34	-.23	.10	.25	-.17	-.04	-.17	-.08	-.10	.15
15. CES Q7															–	.23	.22	-.14	-.31	-.16	-.24	-.25	-.23	-.26	-.28	.12	.11	-.19	-.02	-.18	.13	.03
16. CES Q8																–	.00	-.25	-.15	-.22	.13	-.08	-.21	-.09	.10	-.09	-.21	-.04	-.18	-.45*	.07	.12
17. CPIB																	–	-.60**	-.79**	-.71**	-.75**	-.80**	-.77**	-.78**	-.25	.34	.09	-.08	-.16	.06	-.08	.11
18. VAPP voice problem																		–	.80**	.76**	.75**	.81**	.75**	.76**	.01	.23	-.36	.09	-.19	.07	-.14	-.12
19. VAPP daily communication																			–	.82**	.88**	.97**	.95**	.93**	.26	.00	-.12	.15	-.12	.16	.05	-.01
20. VAPP social communication																				–	.85**	.89**	.80**	.87**	.17	.10	-.13	.17	.01	.31	.08	.02
21. VAPP emotion																					–	.95**	.82**	.91**	.31	.03	-.29	.21	-.12	.09	.03	-.01
22. VAPP total																						–	.94	.97	.34	.04	-.22	.15	-.13	.11	.07	-.07

																								**	**							
23. VAPP activity limitation																								–	.86**	.26	-.02	-.17	.10	-.13	.10	.06	-.02
24. VAPP participation restriction																									–	.42*	.06	-.14	.13	-.11	.09	.14	-.13
25. LSUS																										–	.10	-.13	.02	-.37	-.19	.36	-.05
26. Age																											–	-.18	-.14	-.45*	.21	-.05	.03
27. Gender																												–	-.03	.15	-.03	.54**	.36
28. Disease duration																													–	-.06	.20	-.29	.37
29. GDS																														–	.22	-.37	-.03
30. UPDRS																															–	-.34	-.05
31. MOCA																																–	.00
32. Medication effectiveness																																	–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

Table 52: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Participants with Parkinson's Disease for Visit 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. Habitual speech intensity	–	.48 *	.12	-.16	-.27	.30	.13	.09	-.04	.03	-.22	-.11	-.25	-.44 *	.02	-.17	.10	-.15	-.13	-.08	-.10	-.11	-.12	-.09	-.12	.19	-.02	-.31	-.13	-.06	.13	.23
2. Maximum speech intensity		–	.38	.73 **	-.12	.60 **	.68 **	.43 *	-.13	-.16	-.11	.08	-.26	-.35	.02	-.25	.35	-.19	-.26	-.29	-.34	-.29	-.24	-.27	-.38	-.11	.10	-.46 *	-.32	-.21	.46 *	-.01
3. Lombard response function			–	28	-.03	.54 **	.60 **	.49 *	.13	-.03	.18	.05	.16	.02	.27	-.07	.25	-.21	-.32	-.30	-.34	-.36	-.32	-.39	-.13	-.30	.50 *	-.46 *	.19	-.02	.32	.09
4. Magnitude production				–	.02	.38	.53 *	.32	.01	-.21	.07	.11	-.08	-.10	-.01	-.21	.35	-.07	-.25	-.32	-.35	-.28	-.20	-.29	-.32	-.18	.03	-.18	-.36	-.06	.40 t	-.20
5. Typical speech loudness					–	-.03	.01	.05	.17	.37	.37	.30	.20	.54 **	.46 *	.43 *	.16	-.23	-.04	-.13	-.02	-.04	-.09	.01	.16	.15	.07	.10	-.29	-.13	.22	-.01
6. SIT transcription						–	.69 **	.27	.04	.06	.13	-.14	-.27	-.13	-.10	-.30	.09	-.04	.05	.06	.04	.05	.06	.07	-.06	-.21	.37	-.51 *	-.20	-.32	.64 **	-.04
7. SIT VAS							–	.58 **	-.03	-.08	.03	.09	.04	-.19	.03	-.02	.27	-.09	-.06	-.05	-.12	-.10	-.08	-.10	-.13	-.36	.46 *	-.41 t	-.04	-.24	.50 *	.17
8. Conversational intelligibility								–	-.07	.06	-.13	.05	.00	.04	.12	.12	.11	-.33	-.20	-.22	-.23	-.21	-.14	-.21	-.02	-.44 *	.09	-.50 *	.27	-.16	-.03	-.10
9. CES Q1									–	.72 **	.62 **	.31	.59 **	.69 **	.46 *	.51 *	.57 **	-.31	-.40 t	-.42 t	-.38	-.41 t	-.38	-.43 *	-.09	.13	.22	.20	-.02	.20	-.01	.18
10. CES Q2										–	.46 *	.38	.33	.65 **	.42 *	.55 **	.54 **	-.45 *	-.36	-.38	-.37	-.36	-.34	-.33	.05	.32	.07	.04	-.23	-.16	-.09	.26
11. CES Q3											–	.57 **	.61 **	.56 **	.53 *	.49 *	.30	-.17	-.24	-.22	-.10	-.20	-.25	-.22	-.15	-.07	.12	.19	-.04	-.20	.10	-.10
12. CES Q4												–	.68 **	.38	.52 *	.65 **	.63 **	-.39	-.52 *	-.46 *	-.37	-.48 *	-.53 *	-.48 *	-.31	.04	-.03	.05	-.14	-.45 *	-.06	.22
13. CES Q5													–	.51 *	.58 **	.70 **	.47 *	-.17	-.31	-.28	-.20	-.31	-.35	-.36	-.24	-.09	.16	.19	.18	.01	-.13	.20
14. CES Q6														–	.59 **	.65 **	.34	-.21	-.17	-.26	-.14	-.19	-.18	-.20	-.10	.04	.05	.07	.16	.08	-.15	-.02
15. CES Q7															–	.56 **	.47 *	-.31	-.44 *	-.46 *	-.38	-.47 *	-.50 *	-.48 *	-.40	.24	-.18	-.14	-.01	.12	-.17	.07
16. CES Q8																–	.41 t	-.21	-.25	-.27	-.16	-.23	-.27	-.22	-.01	.04	.08	.11	.13	-.28	-.07	.31
17. CPIB																	–	-.61 **	-.75 **	-.77 **	-.77 **	-.79 **	-.77 **	-.78 **	-.47 *	.29	.25	-.01	-.33	-.12	.08	.46 *
18. VAPP voice problem																		–	.83 **	.77 **	.70 **	.80 **	.83 **	.75 **	.08	-.02	-.11	-.10	.08	.19	.22	-.21
19. VAPP daily communication																			–	.94 **	.91 **	.98 **	.98 **	.98 **	.21	-.15	-.02	-.01	.17	.12	.13	-.16
20. VAPP social communication																				–	.95 **	.96 **	.93 **	.94 **	.21	-.16	-.06	-.02	.29	.07	.06	-.14
21. VAPP emotion																					–	.96 **	.90 **	.93 **	.23	-.27	-.13	.06	.31	-.04	.02	-.16
22. VAPP total																						–	.98 **	.99 **	.26	-.18	-.09	-.01	.19	.02	.12	-.21

23. VAPP activity limitation																								–	.96**	.25	-.15	-.09	-.07	.17	.08	.14	-.25
24. VAPP participation restriction																									–	.29	-.14	-.05	.00	.12	-.01	.16	-.19
25. LSUS																										–	-.24	-.04	.32	-.10	-.11	.04	.02
26. Age																											–	-.18	-.14	-.45*	.21	-.05	.03
27. Gender																												–	-.03	.15	-.03	.54**	.36
28. Disease duration																													–	-.06	.20	-.29	.37
29. GDS																														–	.22	-.37	-.03
30. UPDRS																															–	-.34	-.05
31. MOCA																																–	.00
32. Medication effectiveness																																	–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

3.8.1 Speech intensity measures

Three hundred ninety-five of 435 correlations involving speech intensity measures across all three visits were not significant. Twenty-four of the 40 significant correlations had correlation coefficients below our criterion threshold of .50, and were therefore, not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all 3 visits have been reported. Maximum speech intensity was significantly and highly correlated with magnitude productions for visits 1, 2, and 3 for participants with PD $r = .60$, $p = .003$, $r = .72$, $p < .001$, and $r = .73$, $p < .001$ respectively.

3.8.2 Speech intelligibility measures

Two hundred thirty-eight of 270 correlations involving speech intelligibility measures across all 3 visits were not significant. Eleven of the 32 significant correlations had correlation coefficients below our criterion threshold of .50, and were therefore, not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. SIT transcription scores and SIT VAS ratings were significantly and highly correlated across all three visits for participants with PD $r = .85$, $p < .001$, $r = .76$, $p < .001$, and $r = .69$, $p < .001$ respectively. These results suggest that VAS ratings and orthographic transcription percentage scores are moderately to strongly correlated as a means of evaluating sentence intelligibility in IWPD.

3.8.3 Communicative participation measures

Nine hundred fifty-two of 1173 correlations involving communicative participation measures across all three visits were not significant. Sixty-eight of the 221 significant correlations had correlation coefficients below our criterion threshold of .50, and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. The CPIB standardized score was significantly and highly correlated with the VAPP *self-perceived voice problem* score across all three visits in participants with PD $r = -.69$, $p < .001$, $r = -.60$, $p = .003$, and r

= -.61, p = .002 respectively. The CPIB standardized score was significantly and highly correlated with the VAPP *total* score in participants with PD across visits 1, 2, and 3 r = -.76, p < .001, r = -.80, p < .001, and r = -.79, p < .001 respectively. These correlations were negative, suggesting that as the CPIB score decreased, the VAPP scores increased, and vice versa. A lower score on the CPIB suggests a greater impact on one's communicative participation. A higher VAPP score suggests an increased impact that a voice problem has on one's activity and participation. Therefore, the negative correlations are the result of how the instruments are scaled. These results suggest that the CPIB and the VAPP *voice problem* and *total* scores are highly and significantly associated with each other.

Forty-seven of 84 correlations involving the pairwise comparisons of the 8 questions within the CES were significant and moderately correlated. Only 38% of these significant correlations were significantly correlated across all three visits. Three of the 18 correlations that were significant across all three visits fell above our criterion threshold of .5. The correlations of all pairwise comparisons of the eight questions within the CES ranged from r = -.04 – .66 for visit 1, r = -.03 – .68 for visit 2 and r = .31 – .72 for visit 3. The full correlation matrix related to these within test correlations for the CES at all three visits are provided in Tables 50, 51, and 52.

All 63 correlations involving the pairwise comparisons of the seven subsections within the VAPP were significant and highly correlated. These correlations ranged from r = .76 – .98 for visit 1, r = .75 – .97 for visit 2, and r = .75 – .99 for visit 3. The full correlation matrix related to these within test correlations for the VAPP at all three visits are provided in Tables 50, 51, and 52. These results suggest that, over time, IWPD consistently self-rate the impact of their voice problem on the VAPP.

3.8.4 Demographic and non-speech factors

Descriptive statistics for demographic and non-speech factors, including age, disease duration, GDS scores, UPDRS scores, MOCA scores, and medication effectiveness can be found in Table 53. Five hundred fifty of 588 correlations involving demographic and non-speech factors across all three visits were not significant. Twenty-one of the 38

significant correlations had correlation coefficients below our criterion threshold of .50, and were therefore, not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. Gender and MOCA scores were significantly and moderately correlated across visits 1, 2, and 3 r = .54, p = .008, r = .54, p = .008, and r = .54, p = .008. These results suggest a sex difference between IWPD, wherein women with PD were more likely to present with higher scores on the MOCA than men.

Table 53: Descriptive Statistics of Demographic and Non-Speech Factors for Participants with Parkinson's Disease and Control Participants

	Age (years) mean*(SD)*	Disease duration (years) mean*(SD)*	GDS scores mean*(SD)*	UPDRS scores mean*(SD)*	MOCA scores mean*(SD)*	MES scores mean*(SD)*
Participants with PD	69.48*(5.57)*	10.57*(5.84)*	2.48*(3.03)*	37.18*(12.27)*	24.70*(3.43)*	4.61*(1.27)*
Control participants	69.37*(7.69)*	N/A	.60*(1.00)*	N/A	28.10*(1.52)*	N/A

Note. This table illustrates the means and standard deviations for the different demographic and non-speech factors for participants with PD and control participants.

3.9 Statistical Analysis for Objective 8: Inter-Relationships Among Self- and Proxy-Measures in participants with *Parkinson's Disease*

In order to answer the question '*Are measures of speech intensity, speech intelligibility, proxy ratings of communicative participation, demographic factors, and non-speech factors related to one another in participants with PD?',* a matrix of inter-correlations was obtained via a series of three Pearson correlations ($p < .05$) applied to all possible pairwise combination of the experimental variables, demographic factors, and non-speech factors for participants with PD across each of the three experimental visits. The measures of speech intensity included in these analyses were: habitual speech intensity, maximum speech intensity, Lombard response function, magnitude production, and proxy-perception of typical speech loudness. The measures of speech intelligibility included in these analyses were: SIT transcription score, SIT VAS score, and

conversational intelligibility VAS score. The proxy measures of communicative participation included in these analyses were: CES questions 1 through 8, VAPP *self-perceived voice problem* score, VAPP *daily communication* score, VAPP *social communication* score, VAPP *emotion* score, VAPP *activity limitation* score, VAPP *participation restriction* score, CPIB score, VAPP *total* scores, and LSUS score as rated by the primary communication partners of participants with PD. The measures of demographic factors included in these analyses were: age, gender, and disease duration. The measures of non-speech factors included in these analyses were: GDS scores, disease severity as measured by overall UPDRS scores, and MOCA scores. Only significant correlations, above $r \leq .50$, across the three visits were included and are presented below. This criterion value of $r \leq .50$ was chosen as the minimum threshold since correlations greater than .50 are indicative of a moderate to a very strong relationship between variables (Mukaka, 2012). Since a primary goal of this study is to investigate the consistency of various measures related to hypophonia and communication in IWPD, focusing on significant correlations that minimally had a moderate relationship across all three experimental visits permitted us to identify the most salient variables related to each other. Results of all analyzed correlations are summarized in Tables 54, 55, and 56 for visits 1, 2, and 3 respectively.

Table 54: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Proxy-Rated Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Participants with Parkinson's Disease for Visit 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. Habitual speech intensity	–	.61**	.17	-.16	.06	-.24	-.47*	-.25	-.28	-.18	-.39t	-.31	-.31	-.34	.08	-.27	-.28	.05	.22	.19	.30	.27	.20	.30	-.07	.14	-.41*	.06	-.18	-.02	-.21	-.25
2. Maximum speech intensity		–	.56**	.60**	-.04	.14	.01	.04	-.65**	.08	-.21	-.02	-.01	-.24	.08	.05	-.23	.03	.10	.09	.10	.12	.15	.10	-.23	-.01	-.12	-.18	-.39	-.15	.21	-.30
3. Lombard response function			–	.58**	-.06	.30	.27	.23	-.31	-.13	.09	.13	-.07	-.20	.04	-.08	-.07	.00	-.06	-.02	-.31	-.11	.02	-.09	-.11	-.30	.29	-.29	-.10	-.03	.26	-.17
4. Magnitude production				–	-.07	.42*	.48*	.34	-.48*	.21	.16	.25	.30	.05	.19	.24	.03	.03	-.16	-.17	-.28	-.18	-.05	-.22	-.18	-.12	.33	-.27	-.26	-.12	.53**	-.11
5. Typical speech loudness					–	.01	-.04	.14	.23	.06	.39t	.41*	.40t	.36	.51*	.50*	.59**	-.66**	-.55**	-.55**	-.44*	-.58**	-.63**	-.48*	.15	-.24	.11	.13	.21	-.17	-.12	.17
6. SIT transcription						–	.85**	.28	-.01	.26	.40t	.29	.04	-.03	.45*	.23	.20	-.02	-.18	-.13	-.30	-.21	-.11	-.23	.30	-.21	.35	-.41t	-.03	-.24	.52*	.34
7. SIT VAS							–	.36	-.03	.12	.40t	.33	.05	.05	.41t	.20	.12	-.04	-.09	-.13	-.26	-.15	-.06	-.14	.27	-.23	.60**	-.39t	-.03	-.14	.65**	.31
8. Conversational intelligibility								–	-.01	.06	.24	.18	.20	.15	-.02	.17	.14	-.22	-.33	-.41t	-.41t	-.36	-.24	-.39t	-.03	-.39t	.29	-.41t	.25	-.03	.09	.01
9. CES Q1									–	.20	.31	.02	-.17	.00	-.03	-.13	-.18	.02	-.09	-.05	-.26	-.13	-.10	-.06	.26	.17	-.02	.01	.19	.06	-.19	.41t
10. CES Q2										–	.54**	.51*	.32	-.11	.22	.55**	.41*	-.16	-.20	-.13	-.18	-.20	-.21	-.16	-.26	-.05	-.02	.02	-.09	-.38	-.02	.37
11. CES Q3											–	.74**	.39t	.15	.48*	.48*	.66**	-.44*	-.35	-.44*	-.64**	-.46*	-.39t	-.30	.24	-.19	.38	-.07	.23	-.18	-.05	.38
12. CES Q4												–	.59**	.46*	.36	.75**	.68**	-.52**	-.49*	-.55**	-.47*	-.53**	-.54**	-.45*	.15	-.13	.37	-.16	.00	-.35	.08	.23
13. CES Q5													–	.61**	.43*	.83**	.52**	-.46*	-.56**	-.53**	-.35	-.53**	-.53**	-.55**	-.05	-.17	.27	-.20	.32	-.22	.11	-.13
14. CES Q6														–	.18	.42t	.28	-.27	-.36	-.45*	-.13	-.34	-.35	-.41t	.42	.17	.23	-.27	.18	-.13	.16	-.21
15. CES Q7															–	.51*	.35	-.37	-.22	-.27	-.16	-.23	-.29	-.14	.15	-.34	.41t	-.01	.20	-.35	.44*	.13
16. CES Q8																–	.61**	-.60**	-.59**	-.53**	-.28	-.57**	-.63**	-.56**	-.20	-.28	.22	-.13	.13	-.40t	.10	.12
17. CPIB																	–	-.77**	-.80**	-.72**	-.72**	-.83**	-.83**	-.73**	.09	-.17	.42*	.22	.16	-.27	-.10	.53*
18. VAPP voice problem																		–	.75**	.78**	.56**	.79**	.85**	.67**	.04	.31	-.38	-.09	-.28	.30	.17	-.18
19. VAPP daily communication																			–	.86**	.73**	.98**	.97**	.98**	.00	.15	-.36	.09	-.29	.26	.08	-.34
20. VAPP social communication																				–	.76**	.88**	.84**	.82**	-.21	.17	-.31	.19	-.21	.48*	.03	-.13
21. VAPP emotion																					–	.83**	.66**	.73**	-.23	.21	-.38	.13	-.26	.19	.09	-.35
22. VAPP total																						–	.96**	.96**	-.06	.18	-.39t	.09	-.30	.24	.11	-.37
23. VAPP activity limitation																							–	.92**	.02	.14	-.39t	-.10	-.28	.25	.13	-.38

24. VAPP participation restriction																									–	-.01	.15	-.32	.16	-.28	.21	.06	-.30
25. LSUS																										–	.30	.11	-.37	.04	-.16	.20	-.03
26. Age																											–	-.18	-.14	-.45 *	.21	-.05	.03
27. Gender																												–	-.03	.15	-.03	.54 **	.36
28. Disease duration																													–	-.06	.20	-.29	.37
29. GDS																														–	.22	-.37	-.03
30. UPDRS																															–	-.34	-.05
31. MOCA																																–	.00
32. Medication effectiveness																																	–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3 for proxy-rated dependent variables. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

Table 55: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Proxy-Rated Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Participants with Parkinson's Disease for Visit 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. Habitual speech intensity	–	.69**	-.22	.09	-.01	.01	-.09	-.02	-.32	-.08	-.22	-.13	-.09	.02	-.20	-.22	.10	.01	-.07	-.01	-.18	-.06	.00	-.03	.03	.23	-.27	-.08	-.13	.00	-.04	-.08
2. Maximum speech intensity		–	-.09	.72**	.16	.09	-.05	.24	-.57**	-.27	-.41t	-.11	-.13	-.12	-.50*	-.33	.16	.05	.00	.04	-.25	-.04	.08	-.01	-.06	.12	-.09	-.34	-.34	-.08	.23	-.29
3. Lombard response function			–	-.02	.00	.22	.20	.20	.09	.03	.14	.13	.19	.08	.10	.19	-.02	.01	.01	-.05	.03	.01	.02	-.02	-.08	.09	.50*	-.56**	.10	-.03	.39t	.05
4. Magnitude production				–	.22	.02	-.10	.21	-.54**	-.30	-.43*	-.12	-.09	-.10	-.52*	-.23	.22	.07	-.02	.01	-.22	-.07	.02	-.05	-.11	.06	.11	-.29	-.35	-.09	.18	-.25
5. Typical speech loudness					–	-.11	-.18	.27	.26	.56**	.40t	49*	.44*	.21	.44*	.48*	.48*	-.66**	-.59**	-.64**	-.39t	-.58**	-.64**	-.55**	.51*	.04	.25	-.13	-.12	-.36	.20	-.03
6. SIT transcription						–	.76**	.45*	-.07	-.01	-.03	-.05	-.02	.14	-.06	.13	.16	.02	-.16	-.17	-.24	-.17	-.09	-.19	.24	-.20	.41t	-.30	.10	-.26	.52*	.26
7. SIT VAS							–	.52*	-.13	-.12	-.11	-.20	-.12	.02	.05	.03	.04	.12	-.03	-.10	-.14	-.06	.01	-.10	.11	-.38	.35	-.08	.03	-.02	.44*	.39t
8. Conversational intelligibility								–	-.03	.08	.21	.09	.08	.00	.24	.24	.24	-.34	-.24	-.32	-.40t	-.30	-.27	-.22	.32	-.38	.28	-.17	-.03	-.30	.31	.07
9. CES Q1									–	.66**	.89**	.61**	.33	.21	.81**	.55**	.17	-.38	-.29	-.32	-.15	-.30	-.38	-.26	.46*	.20	-.02	.02	.21	-.12	-.24	.34
10. CES Q2										–	.75**	.80**	.64**	.56**	.61**	.72**	.54**	-.64**	-.61**	-.56**	-.42*	-.58**	-.62**	-.56**	.73**	.28	.14	-.01	-.05	.12	.02	.32
11. CES Q3											–	.75**	.51*	.36	.86**	.65**	.34	-.61**	-.49*	-.52*	-.43*	-.51*	-.54**	-.45*	.65**	.16	.01	-.17	.31	-.05	-.25	.14
12. CES Q4												–	.69**	.61**	.54**	.72**	.62**	-.66**	-.64**	-.60**	-.62**	-.64**	-.60**	-.60**	.61**	.31	.15	-.26	.06	-.10	-.01	.03
13. CES Q5													–	.88**	.37	.84**	.71**	-.59**	-.72**	-.68**	-.63**	-.73**	-.70**	-.73**	.61**	.18	.34	-.14	.38	.15	-.02	.17
14. CES Q6														–	.28	.72**	.67**	-.48*	-.67**	-.63**	-.65**	-.66**	-.60**	-.65**	.54**	.11	.25	-.12	.35	.11	-.08	.23
15. CES Q7															–	.64**	.23	-.54**	-.47*	-.56**	-.30	-.46*	-.55**	-.40t	.50*	-.07	-.04	-.04	.33	-.23	-.25	.22
16. CES Q8																–	.69**	-.64**	-.77**	-.75**	-.56**	-.74**	-.80**	-.73**	.68**	.07	.38t	-.07	.34	-.17	.08	.32
17. CPIB																	–	-.73**	-.91**	-.83**	-.87**	-.93**	-.87**	-.92**	.59**	.10	.41*	-.08	.19	-.07	-.04	.28
18. VAPP voice problem																		–	.81**	.78**	.76**	.84**	.84**	.74**	-.56**	.10	-.30	-.07	-.32	.10	.18	-.06
19. VAPP daily communication																			–	.95**	.84**	.98**	.97**	.97**	-.62**	-.05	-.37	.08	-.31	.11	.11	-.23
20. VAPP social communication																				–	.79**	.95**	.94**	.94**	-.58**	-.01	-.29	.19	-.29	.20	.10	-.16
21. VAPP emotion																					–	.90**	.77**	.84**	-.56**	-.03	-.27	.18	-.30	-.08	.16	-.07
22. VAPP total																						–	.96**	.98**	-.61**	-.03	-.35	.09	-.33	.03	.15	-.22
23. VAPP activity limitation																							–	.94**	-.60**	.00	-.35	-.01	-.32	.13	.16	-.29

24. VAPP participation restriction																								–	-.59**	-.08	-.35	.13	-.31	.01	.10	-.25
25. LSUS																									–	.08	.28	-.16	.17	-.10	.18	.17
26. Age																										–	-.18	-.14	-.45*	.21	-.05	.03
27. Gender																											–	-.03	.15	-.03	.54**	.36
28. Disease duration																												–	.06	.20	-.29	.37
29. GDS																													–	.22	-.37	-.03
30. UPDRS																														–	-.34	-.05
31. MOCA																															–	.00
32. Medication effectiveness																																–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3 for proxy-rated dependent variables. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

Table 56: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Proxy-Rated Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Participants with Parkinson's Disease for Visit 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. Habitual speech intensity	–	.23	-.33	.01	.20	-.08	-.11	.32	.20	.22	.14	.16	-.04	-.10	.39	-.01	.19	-.09	-.18	-.19	-.12	-.13	-.09	-.18	.05	.05	-.52 *	-.09	-.04	-.10	-.26	-.13
2. Maximum speech intensity		–	.38	.73 **	-.11	.60 **	.68 **	.43 *	-.18	.13	-.03	.14	.08	.03	.05	-.08	.10	.13	-.06	-.16	-.09	-.08	.05	-.27	-.31	-.11	.10	-.46 *	-.32	-.21	.46 *	.00
3. Lombard response function			–	.28	.03	.54 **	.60 **	.49 *	.19	.09	.06	-.05	.19	-.01	.11	.02	.07	-.13	-.15	-.15	-.21	-.21	-.18	-.20	-.15	-.30	.50 *	-.46 *	.19	-.02	.32	.09
4. Magnitude production				–	02	.38	.53	.32 *	-.12	.04	-.13	-.01	.14	.03	.21	-.05	-.10	.18	.11	-.02	.08	.09	.20	-.09	-.24	-.18	.03	-.18	-.36	-.06	.40 t	-.20
5. Typical speech loudness					–	-.15	-.02	.18	.50 *	.59 **	.58 **	.37	.45 *	.34	.76 **	.54 **	.42 t	-.66 **	-.61 **	-.57 **	-.48 *	-.57 **	-.62 **	-.45 *	.33	.16	.08	.11	-.23	-.29	.08	.16
6. SIT transcription						–	.69 **	.27	.07	.25	.02	-.04	-.12	-.08	-.09	-.26	-.13	.17	.17	.04	.03	.12	.18	.09	-.02	-.21	.37	-.51 *	-.20	-.32	.64 **	-.04
7. SIT VAS							–	.58 **	.11	.27	-.04	-.17	-.12	-.15	.09	-.24	-.08	.02	-.03	-.24	-.28	-.15	-.05	-.17	-.17	-.36	.46 *	-.41 t	-.04	-.24	.50 **	.17
8. Conversational intelligibility								–	.33	.33	.19	.03	.04	-.08	.30	-.15	.12	-.19	-.30	-.51 *	-.53 *	-.39	-.23	-.41 t	.00	-.44 *	.09	-.50 *	.27	-.16	-.03	-.10
9. CES Q1									–	.61 **	.55 **	.24	.30	.06	.60 **	.31	.40 t	-.65 **	-.49 *	-.55 **	-.61 **	-.57 **	-.56 **	-.40 t	.31	-.21	.20	.03	.25	-.21	-.20	.34
10. CES Q2										–	.66 **	.50 **	.49 *	.41 t	.55 *	.40 t	.60 **	-.49 *	-.52 *	-.56 **	-.56 **	-.55 **	-.54 **	-.43 *	.60 **	.03	.40 t	-.13	-.06	-.23	.24	.10
11. CES Q3											–	.74 **	.68 **	.59 **	.47 *	.59 **	.56 **	-.71 **	-.67 **	-.58 **	-.69 **	-.69 **	-.70 **	-.52 *	.46 *	.09	.26	.13	-.10	-.30	-.04	.42 t
12. CES Q4												–	.85 **	.73 **	.36	.68 **	.75 **	-.49 *	-.71 **	-.56 **	-.54 **	-.64 **	-.60 **	-.65 **	.44 *	.20	.22	.01	-.01	-.24	.09	.26
13. CES Q5													–	.73 **	.48 *	.86 **	.76 **	-.55 **	-.68 **	-.50 **	-.51 *	-.63 **	-.64 **	-.60 **	.47 *	.09	.27	.09	.08	-.06	.11	.19
14. CES Q6														–	.42 t	.67 **	.63 **	-.45 *	-.55 **	-.36	-.42 t	-.50 *	-.52 *	-.43 *	.48 *	-.04	.30	.32	-.05	.03	.08	.10
15. CES Q7															–	.54 **	.61 **	-.69 **	-.61 **	-.56 **	-.51 *	-.60 **	-.61 **	-.51 *	.23	-.16	.05	.23	-.03	-.25	-.06	.09
16. CES Q8																–	.59 **	-.68 **	-.63 **	-.42 t	-.41 t	-.57 **	-.64 **	-.45 *	.43 *	-.05	.19	.34	.07	-.16	.06	.25
17. CPIB																	–	-.58 **	-.74 **	-.61 **	-.56 **	-.69 **	-.67 **	-.70 **	.38	.13	.24	.06	.05	-.08	-.02	.18
18. VAPP voice problem																		–	.81 **	.68 **	.72 **	.82 **	.90 **	.58 **	-.19	.19	-.25	-.27	-.17	.36	.21	-.47 *
19. VAPP daily communication																			–	.89 **	.86 **	.98 **	.95 **	.91 **	-.25	.01	-.24	.02	-.27	.28	.14	-.33
20. VAPP social communication																				–	.92 **	.93 **	.79 **	.94 **	-.26	.11	-.21	.18	-.24	.43 *	.08	-.21
21. VAPP emotion																					–	.94 **	.84 **	.85 **	-.31	.17	-.28	.10	-.31	.30	.17	-.41 t
22. VAPP total																						–	.95 **	.91 **	-.24	.08	-.27	.04	-.30	.28	.17	-.39 t

23. VAPP activity limitation																								–	.77**	-.23	.05	-.30	-.09	-.31	.21	.19	-.45*
24. VAPP participation restriction																									–	-.15	.01	-.18	.20	-.23	.29	.09	-.20
25. LSUS																										–	-.01	.31	.01	.11	-.03	.28	-.12
26. Age																											–	-.18	-.14	-.45*	.21	-.05	.03
27. Gender																												–	-.03	.15	-.03	.54**	.36
28. Disease duration																													–	-.06	.20	-.29	.37
29. GDS																														–	.22	-.37	-.03
30. UPDRS																															–	-.34	-.05
31. MOCA																																–	.00
32. Medication effectiveness																																	–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3 for proxy-rated dependent variables. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

3.9.1 Speech intensity measures

Three hundred sixty-five of 435 correlations involving speech intensity measures across all three visits were not significant. Twenty-six of the 70 significant correlations had correlation coefficients below our criterion threshold of .50 and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. Proxy-rated typical speech loudness was significant and moderately correlated with the VAPP *voice problem* score r = -.66, p = .001, r = -.66, p = .001, and r = -.66, p = .001 respectively, VAPP *daily communication* score r = -.55, p = .006, r = -.59, p = .003, and r = -.61, p = .003 respectively, VAPP *social communication* score r = -.55, p = .006, r = -.64, p = .001, and r = -.57, p = .006 respectively, VAPP *total* score r = -.58, p = .004, r = -.58, p = .004, and r = -.57, p = .005 respectively, and VAPP *activity limitation* score r = -.63, p = .001, r = -.64, p = .001, and r = -.62, p = .002 respectively. These correlations were negative, suggesting that as the perceived typical speech loudness score decreased, the VAPP scores increased, and vice versa. A lower perceived typical speech loudness score suggests that the individual is perceived as communicating with a quieter or softer voice. A higher VAPP score suggests an increased impact that a voice problem has on one's activity and participation. Therefore, the negative correlations are the result of how the instruments are scaled. These results suggest that one's perceived reduced speech loudness is significantly and moderately associated with VAPP *voice problem, daily communication, social communication, total* and *activity limitation* scores. See Objective 7 for variables other than proxy-rated communicative participation that were significantly correlated with speech intensity measures.

3.9.2 Speech intelligibility measures

Two hundred forty of 270 correlations involving speech intelligibility measures across all three visits were not significant. Eleven of the 30 significant correlations had correlation coefficients below our criterion threshold of .50 and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits

have been reported. There was no consistent pattern of results that emerged for correlations between proxy-rated communicative participation and speech intelligibility measures. See Objective 7 for variables other than proxy-rated communicative participation that were significantly correlated with speech intelligibility measures.

3.9.3 Communicative participation measures

Eight hundred seventeen of 1173 correlations involving communicative participation measures across all three visits were not significant. Fifty-seven of the 356 significant correlations had correlation coefficients below our criterion threshold of .50 and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. The 299 significant correlations that fell above our criterion threshold of .50 were scattered across various other variables and no consistent pattern emerged. Proxy-rated communicative participation evaluated via the CPIB was significantly and moderately correlated across all three visits with proxy-ratings of CES question 4: $r = .68$, $p < .001$, $r = .62$, $p = .002$, and $r = .75$, $p < .001$ respectively, CES question 5: $r = .52$, $p = .010$, $r = .71$, $p < .001$, and $r = .76$, $p < .001$ respectively, and CES question 8: $r = .61$, $p = .002$, $r = .69$, $p < .001$, and $r = .59$, $p = .004$ respectively. Question 4 of the CES corresponds to communicating with a stranger over the telephone, question 5 of the CES relates to communicating in a noisy environment, and question 8 of the CES corresponds with communicating with someone at a distance. These results suggest that overall communicative participation as measured by the CPIB is moderately associated with the communicative contexts of communicating with an unfamiliar speaker over the telephone, speaking in noise, and communicating across a distance. The latter two communicative situations are challenging since they both require the use of adequate speech intensity. Furthermore, the primary communication partners of participants with PD rated these communicative contexts consistently over the three visits, suggesting relative stability in perception of communicative effectiveness over time.

The CPIB standardized score was also significantly and highly correlated with proxy-ratings of the VAPP *self-perceived voice problem* score across all three visits by the

primary communication partners of the participants with PD: r = -.77, p < .001, r = -.73, p < .001, and r = -.58, p = .005 respectively. The CPIB standardized score was significantly and highly correlated with proxy-ratings of the VAPP *total* score across all 3 visits by the primary communication partners of the participants with PD: r = -.83, p < .001, r = -.93, p < .001, and r = -.69, p < .001, respectively. These correlations were negative, suggesting that as the CPIB score decreased, the VAPP scores increased, and vice versa. A lower score on the CPIB suggests a greater impact on one's communicative participation. A higher VAPP score suggests an increased impact that a voice problem has on one's activity and participation. Therefore, the negative correlations are the result of how the instruments are scaled. These results suggest that the CPIB and the VAPP *self-perceived voice problem* and *total* scores are highly and significantly associated with each other.

Fifty-six of 84 correlations involving the pairwise comparisons of the eight questions within the CES were significantly correlated. A total of 64% of these significant correlations were significantly correlated across all three visits. Thirty-two of the 35 correlations that were significant across all three visits fell above our criterion threshold of .5. The correlations of all pairwise comparisons of the eight questions within the CES ranged from r = -.17 - .86 for visit 1, r = .06 - .85 for visit 2, and r = .21 - .89 for visit 3. The full correlation matrix related to these within test correlations for the CES at all three visits are provided in Tables 54, 55, and 56. These results suggest that the primary communication partners of our participants with PD rate their partners' communicative effectiveness fairly consistently across select communicative situations/contests and time.

All 63 correlations involving the pairwise comparisons of the seven subsections within the VAPP were significant and highly correlated. These correlations ranged from r = .56 - .98 for visit 1, r = .58 - .98 for visit 2, and r = .74 - .98. The full correlation matrix related to these within test correlations for the VAPP at all three visits are provided in Tables 54, 55, and 56. These results suggest that, over time, the primary communication partners of IWPD consistently rate the impart of their partners' voice problem on the VAPP

3.9.4 Demographic and non-speech factors

Five hundred fifty-three of 588 correlations involving demographic and non-speech factors across all three visits were not significant. Eighteen of the 35 significant correlations had correlation coefficients below our criterion threshold of .50 and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. There was no consistent pattern of results that emerged for correlations between proxy-rated communicative participation and demographic and non-speech factors. See Objective 7for variables other than proxy-rated communicative participation that were significantly correlated with demographic and non-speech factors. These results suggest that proxy-rated communicative participation measures are not consistently correlated with demographic and non-speech factors for participants with PD.

3.10 Statistical Analysis for Objective 9: Inter-Relationships Among Variables in Control Participants

In order to answer the question '*Are measures of speech intensity, speech intelligibility, communicative participation, demographic factors, and non-speech factors related to one another in control participants?',* a matrix of inter-correlations was obtained via a series of three Pearson correlations ($p < .05$) applied to all possible pairwise combination of the experimental variables, demographic factors, and non-speech factors for control participants across each of the three experimental visits. The measures of speech intensity included in these analyses were: habitual speech intensity, maximum speech intensity, Lombard response function, magnitude production, and self-perceptions of typical speech loudness. The measures of speech intelligibility included in these analyses were: SIT transcription score, SIT VAS score, and conversational intelligibility VAS score. The measures of communicative participation included in these analyses were: CES questions 1 through 8, VAPP *self-perceived voice problem* score, VAPP *daily communication* score, VAPP *social communication* score, VAPP *emotion* score, VAPP *activity limitation* score, VAPP *participation restriction* score, VAPP *total* score, CPIB score, and LSUS score. The measures of demographic factors included in these analyses were: age and gender. The measures of non-speech factors included in these analyses were: GDS and

MOCA scores. Only significant correlations above $r \leq .50$ across the three visits were included and are presented below. This criterion value of $r \leq .50$ was chosen as the minimum threshold since correlations greater than .50 are indicative of a moderate to a very strong relationship between variables (Mukaka, 2012). Since a primary goal of this study is to investigate the consistency of various measures related to hypophonia and communication in IWPD, focusing on significant correlations that minimally had a moderate relationship across all three experimental visits permitted us to identify the most salient variables related to each other. Results of all analyzed correlations are summarized in Tables 57, 58, and 59 for visits 1, 2, and 3 respectively.

Table 57: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Control Participants for Visit 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. Habitual speech intensity	–	.74**	.11	.08	-.11	-.13	-.11	-.12	.08	.04	.09	-.00	.09	-.03	-.07	.26	-.07	-.14	-.15	-.08	.05	-.13	-.19	-.18	.20	.22	-.12	.07	.03
2. Maximum speech intensity		–	.18	.57**	.03	.08	.11	.17	.17	.14	.08	.04	.24	.18	.28	.34t	-.03	-.32	-.31	-.20	-.11	-.28	-.39*	-.20	.22	.00	.01	-.04	.06
3. Lombard response function			–	.02	-.08	.28	.20	.04	.05	-.02	-.07	-.08	-.22	.06	.02	-.02	.02	.03	.20	.16	.40*	.28	.04	.46*	.02	-.23	.03	-.14	-.17
4. Magnitude production				–	.03	.14	.18	.35t	.12	.11	.13	.18	.19	.21	.27	.22	.01	-.28	-.36*	-.19	-.30	-.34t	-.37*	-.25	.11	-.05	.24	-.39*	.06
5. Typical speech loudness					–	.40*	.36t	.36*	.12	-.07	.15	-.23	-.18	-.25	.06	-.12	.21	-.30	-.12	-.03	-.20	-.15	-.14	-.06	.36*	-.08	.24	.16	.06
6. SIT transcription						–	.57**	.45*	.33	.26	.20	.00	-.12	.15	.23	-.08	.06	-.23	-.13	-.39*	-.07	-.21	-.28	-.16	.16	-.04	.44*	.08	.01
7. SIT VAS							–	.51**	.11	.05	-.06	-.01	-.03	-.14	.12	-.15	.10	-.14	-.05	-.12	.07	-.06	-.16	-.01	.14	-.43*	.47**	.11	.21
8. Conversational intelligibility								–	.19	.10	.20	.08	.11	.26	.30	.09	.19	-.31	-.12	-.33	-.03	-.17	-.26	-.13	.32	-.34t	.47**	-.09	.30
9. CES Q1									–	.83**	.71**	.46*	.56**	.56**	.74**	.64**	.60**	-.36t	-.49**	-.60**	-.34t	-.47**	-.64**	-.20	.18	-.19	-.02	-.14	-.23
10. CES Q2										–	.62**	.56**	.52**	.58**	.75**	.57**	.37*	-.38*	-.57**	-.66**	-.42*	-.54**	-.66**	-.32	-.03	-.15	.00	.00	-.04
11. CES Q3											–	.69**	.32	.53**	.45*	.64**	.68**	-.44*	-.46**	-.66**	-.38*	-.48**	-.61**	-.29	.22	.09	.04	-.20	.03
12. CES Q4												–	.46**	.44*	.38*	.68**	.48**	-.40*	-.48**	-.62**	-.37*	-.47**	-.55**	-.31	.16	.08	-.08	-.11	.09
13. CES Q5													–	.48**	.57**	.73**	.27	.00	-.25	-.38*	-.15	-.22	-.34t	-.07	.11	-.19	-.06	-.14	-.09
14. CES Q6														–	.59**	.52**	.41*	-.12	-.20	-.54**	-.12	-.21	-.39*	-.01	.00	.02	-.19	-.22	-.12
15. CES Q7															–	.64**	.46*	-.33	-.36t	-.48**	-.29	-.36t	-.49**	-.11	.03	-.43*	.01	.06	-.11
16. CES Q8																–	.57**	-.14	-.25	-.49**	-.14	-.25	-.41*	-.03	.23	-.01	-.06	-.13	.10
17. CPIB																	–	-.29	-.28*	-.36	-.24	-.28	-.44*	-.01	.18	-.19	.09	-.20	-.03
18. VAPP voice problem																		–	.77**	.61**	.69**	.80**	.78**	.64**	-.29	.15	-.12	.03	-.08
19. VAPP daily communication																			–	.65**	.90**	.97**	.93**	.83**	-.14	.02	.00	.06	.04
20. VAPP social communication																				–	.55**	.74**	.81**	.58**	-.11	.00	-.16	-.06	-.22
21. VAPP emotion																					–	.93**	.76**	.85**	-.13	-.12	.01	.08	.04
22. VAPP total																						–	.92**	.89**	-.17	-.04	-.07	.04	-.03
23. VAPP activity limitation																							–	.67**	-.17	.15	-.13	.06	.00

24. VAPP participation restriction																								–	-.12	-.23	-.04	-.04	-.14
25. LSUS																									–	.13	.00	-.18	.09
26. Age																										–	-.37 *	-.08	-.04
27. Gender																											–	.10	.34 t
28. GDS																												–	.14
29. MOCA																													–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

Table 58: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Control Participants for Visit 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. Habitual speech intensity	–	.72**	-.06	.04	.23	.30	.11	-.02	.12	.00	.00	.05	.20	.14	.34t	.27	.43*	-.25	-.27	-.27	-.18	-.28	-.33	-.22	-.02	-.04	.14	-.07	.13
2. Maximum speech intensity		–	-.05	.58**	.00	.40*	.34	.21	.18	.10	.10	.07	.08	-.01	.38*	.22	.32	-.28	-.26	-.35t	-.35t	-.31	-.31	-.23	.04	-.19	.11	-.13	.18
3. Lombard response function			–	-.07	.12	-.16	.09	.12	.09	.19	.19	.15	-.12	.01	.17	-.08	.11	-.07	.06	-.18	-.06	.00	-.01	.06	.18	-.17	-.16	-.28	-.29
4. Magnitude production				–	-.22	.26	.42*	.12	.05	.04	.04	.05	-.13	-.11	-.01	.06	-.08	-.25	-.13	-.16	-.31	-.18	-.10	-.16	.01	-.23	.18	-.10	.06
5. Typical speech loudness					–	.41*	.26	.15	.26	.15	.15	.09	.30	.11	.14	.33	.10	-.26	-.22	-.33	-.28	-.26	-.29	-.16	.26	.05	.19	.14	.24
6. SIT transcription						–	.61**	.48**	.18	.27	.27	-.02	.12	.15	.20	.27	.02	.07	-.05	-.06	.04	-.03	-.05	-.05	-.06	-.13	.23	.02	.60**
7. SIT VAS							–	.48**	.31	.21	.21	.05	.14	.06	.39*	.27	.24	-.23	-.01	.01	-.20	-.05	-.05	.08	-.07	-.54**	.41*	-.03	.32
8. Conversational intelligibility								–	.41*	.50**	.50**	.20	.15	.15	.37t	.12	.05	-.14	-.08	-.18	-.09	-.09	-.11	-.03	.16	-.57**	.18	-.01	.27
9. CES Q1									–	.85**	.85**	.42*	.55**	.52**	.65**	.68**	.55**	-.55**	-.55**	-.35t	-.41*	-.55**	-.61**	-.45*	.09	-.12	.17	-.36t	-.11
10. CES Q2										–	1.00**	.52**	.51**	.66**	.55**	.49**	.27	-.17	-.44*	-.31	-.15	-.40*	-.45*	-.46*	.17	-.10	.02	-.36*	-.05
11. CES Q3											–	.52**	.51**	.66**	.53**	.49**	.27	-.17	-.44*	-.31	-.15	-.40*	-.45*	-.46*	.17	-.10	.02	-.36*	-.05
12. CES Q4												–	.55**	.34t	.43*	.52**	.19	-.10	-.17	-.10	.05	-.16	-.17	-.24	-.08	-.06	.17	-.16	-.04
13. CES Q5													–	.54**	.57**	.77**	.53**	-.31	-.59**	-.41*	-.30	-.55**	-.59**	-.57**	.28	-.14	.10	-.28	-.06
14. CES Q6														–	.52**	.53**	.49**	-.00	-.48**	-.22	.00	-.37*	-.42*	-.51**	.07	.06	-.24	-.49**	-.22
15. CES Q7															–	.68**	.74**	-.25	-.33	-.26	-.22	-.32	-.38*	-.25	-.12	-.35t	.10	-.17	-.06
16. CES Q8																–	.71**	-.46*	-.58**	-.30	-.30	-.53**	-.59**	-.49**	-.07	.01	.13	-.29	-.06
17. CPIB																	–	-.46*	-.43*	-.26	-.34t	-.42*	-.51**	-.27	-.10	-.11	-.07	-.47**	-.27
18. VAPP voice problem																		–	.62**	.59**	.82**	.71**	.70**	.47**	-.20	.09	-.27	.14	.13
19. VAPP daily communication																			–	.81**	.75**	.98**	.97**	.96**	-.33	-.08	.01	.28	.20
20. VAPP social communication																				–	.84**	.89**	.85**	.81**	-.42*	.02	-.01	.06	.15
21. VAPP emotion																					–	.86**	.80**	.65**	-.32	.14	-.13	.01	.18
22. VAPP total																						–	.98**	.93**	-.35t	-.04	-.05	.22	.20
23. VAPP activity limitation																							–	.89**	-.33	-.05	-.05	.30	.17

24. VAPP participation restriction																								–	-.34 t	-.15	.05	.22	.21
25. LSUS																									–	-.05	-.28	-.25	-.03
26. Age																										–	-.37 *	-.08	-.04
27. Gender																											–	.10	.34 t
28. GDS																												–	.14
29. MOCA																													–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

Table 59: Correlations of Speech Intensity Measures, Speech Intelligibility Measures, Communicative Participation Measures, Demographic Factors, and Non-Speech Factors for Control Participants for Visit 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. Habitual speech intensity	–	.74**	.28	.15	.12	-.10	.21	-.01	-.12	-.12	.16	.08	.23	.15	-.09	.19	-.03	.15	-.26	-.23	-.22	-.24	-.24	-.26	.15	-.02	.06	.18	.04
2. Maximum speech intensity		–	.36t	.54**	.05	.02	.18	.14	-.20	-.20	-.03	-.08	.01	.01	-.27	.02	-.18	.09	-.24	-.25	-.27	-.23	-.21	-.23	.15	-.06	-.05	.04	.08
3. Lombard response function			–	.07	-.03	.12	.21	.07	-.03	-.03	.07	.15	-.10	.18	.12	.14	.00	.03	-.21	-.24	-.07	-.18	-.21	-.22	.15	-.10	-.01	.13	-.12
4. Magnitude production				–	-.16	.09	.28	.34t	-.09	-.09	.02	-.21	-.03	-.08	-.20	-.05	-.19	-.12	-.14	-.12	-.21	-.16	-.11	-.14	-.17	-.01	.05	-.26	.05
5. Typical speech loudness					–	-.05	-.14	-.13	.11	.11	.22	.55**	.40*	.42*	.05	.42*	.14	-.33	-.51**	-.44*	-.38*	-.49**	-.54**	-.43*	.11	.17	-.11	.08	-.04
6. SIT transcription						–	.34t	.27	.06	.06	.08	.14	.00	.17	-.07	-.12	-.24	-.23	-.37*	-.42*	-.38*	-.40*	-.38*	-.38*	.13	.01	.03	.08	.19
7. SIT VAS							–	.73**	.24	.24	.08	-.06	.06	-.04	.15	-.12	-.03	-.31	-.06	-.06	-.05	-.09	-.16	.01	.18	-.43*	.64**	.19	.17
8. Conversational intelligibility								–	.23	.23	-.02	-.12	.01	-.06	.05	-.20	-.03	-.30	.07	.08	.08	.05	-.01	.13	.25	-.41*	.55**	-.01	.39*
9. CES Q1									–	1.00**	.70**	.42*	.60**	.34t	.70**	.34t	.43*	-.85**	-.08	.10	.18	-.10	-.30	.05	-.16	-.21	.28	.11	.26
10. CES Q2										–	.70**	.42*	.60**	.34t	.70**	.34t	.43*	-.85**	-.08	.10	.18	-.10	-.30	.05	-.16	-.21	.28	.11	.26
11. CES Q3											–	.60**	.61**	.48**	.45**	.48**	.26	-.58**	-.35t	-.16	-.01	-.35t	-.50**	-.33	-.04	.10	.12	-.24	.02
12. CES Q4												–	.60**	.81**	.56**	.60**	.43*	-.45*	-.57**	-.44*	-.24	.53**	-.64**	-.53**	.13	.02	-.10	.00	-.03
13. CES Q5													–	.73**	.69**	.73**	.64**	-.56**	-.50**	-.31	-.25	-.48**	-.59**	-.41*	-.04	-.14	.04	-.10	.04
14. CES Q6														–	.60**	.63**	.48**	-.37*	-.62**	-.54**	-.34t	-.59**	-.64**	-.62**	.20	-.19	-.19	-.14	-.07
15. CES Q7															–	.60**	.79**	-.55**	-.09	.05	.19	-.06	-.21	-.01	-.16	-.27	.18	.11	.08
16. CES Q8																–	.65**	-.25	-.36*	-.21	-.05	-.31	-.42*	-.31	-.13	.08	-.02	-.06	.14
17. CPIB																	–	-.27	.05	.18	.28	.09	-.02	.11	-.16	-.14	.08	-.07	.10
18. VAPP voice problem																		–	.35t	.20	.18	.38*	.52**	.21	.00	.21	-.25	-.12	-.17
19. VAPP daily communication																			–	.96**	.90**	.99**	.97**	.98**	-.21	.09	.16	.10	.21
20. VAPP social communication																				–	.94**	.96**	.89**	.96**	-.30	.07	.20	.10	.25
21. VAPP emotion																					–	.92**	.81**	.89**	-.28	.05	.20	.05	.24
22. VAPP total																						–	.97**	.97**	-.24	.08	.14	.10	.21

23. VAPP activity limitation																							–	.91**	-.18	.12	.05	.08	.14
24. VAPP participation restriction																								–	-.26	.02	.24	.18	.28
25. LSUS																									–	-.23	.27	-.06	.21
26. Age																										–	-.37*	-.08	-.04
27. Gender																											–	.10	.34t
28. GDS																												–	.14
29. MOCA																													–

Note. t indicates correlations that trended towards significance. * indicates correlations that were statistically significant at the $p < .005$ level. ** indicates correlations that were statistically significant at the $p < .001$ level. Correlations highlighted in yellow were greater than or equal to .50 and statistically significant across visits 1, 2, and 3. Correlations highlighted in blue were statistically significant and greater than or equal to .50.

3.10.1 Speech intensity measures

Three hundred sixty of 390 correlations involving speech intensity measures across all three visits were not significant. Twenty-one of the 30 significant correlations had correlation coefficients below our criterion threshold of .50 and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. Maximum speech intensity was significantly and highly correlated across all three visits for control participants with habitual speech intensity $r = .74$, $p < .001$, $r = .72$, $p < .001$, and $r = .74$, $p < .001$, and magnitude production $r = .57$, $p = .001$, $r = .58$, $p = .001$, and $r = .54$, $p = .002$ respectively.

3.10.2 Speech intelligibility measures

Two hundred eight of 243 correlations involving speech intelligibility measures across all three visits were not significant. Seventeen of the 35 significant correlations had correlation coefficients below our criterion threshold of .50 and were therefore not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. The 18 significant correlations that were above .50 were scattered across various other variables, and no consistent pattern emerged.

3.10.3 Communicative participation measures

Seven hundred forty-two of 1020 correlations involving communicative participation measures across all three visits were not significant. One hundred two of the 278 significant correlations had correlation coefficients below our criterion threshold of .50, and therefore, not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. Self-perceived communicative participation evaluated via the CPIB was significantly and highly correlated across all three visits with self-ratings of CES question 8: $r = .57$, $p = .001$, $r = .71$, $p < .001$, and $r = .65$, $p < .001$. Question 8 of the CES relates to communicating with someone at a distance. These results suggest that overall communicative participation as measured by

the CPIB is strongly associated with the communicative context of communicating across a distance. This communicative situation is challenging since it requires adequate speech intensity. Furthermore, the control participants rated this communicative context consistently over the three visits suggesting relative stability in perception of effectiveness over time.

Seventy-eight of 84 correlations involving the pairwise comparisons of the 8 questions of the CES were significantly correlated. A total of 85% of these significant correlations were significantly correlated across all three visits. Fifty-three of the 66 correlations that were significant across all three visits fell above our threshold criterion of .50. The correlations of all pairwise comparisons of the eight questions within the CES ranged from r = .38 - .83 for visit 1, r = .42 – 1.00 for visit 2, and r = .42 – 1.00 for visit 3. The full correlation matrix related to these within test correlations for the CES at all three visits are provided in Tables 57, 58, and 59. These results suggest that control participants rate their communicative effectiveness fairly consistently across communicative situations/contexts and time.

Fifty-nine of 63 correlations involving the pairwise comparisons of the seven subsections within the VAPP were significantly correlated. A total of 86% of these significant correlations were correlated across all three visits. Forty-eight of the 51 correlations that were significant across all three visits fell above our threshold criterion of .50. These correlations of all pairwise comparisons of the 7 subsections within the VAPP ranged from r = .55 – .97 for visit 1, r = .47 – .98 for visit 2, and r = .38 – .99 for visit 3. The full correlation matrix related to these within test correlations for the VAPP at all three visits are provided in Tables 57, 58, and 59. These results suggest that with the exception of the VAPP *self-perceived voice problem* score, control participants consistently rate the impact of their voice problem on the VAPP.

3.10.4 Demographic and non-speech factors

Descriptive statistics for demographic and non-speech factors, including age, disease duration, GDS scores, UPDRS scores, MOCA scores, and medication effectiveness can be found in Table 53. Two hundred ninety-six of 318 of correlations involving

demographic and non-speech factors across all three visits were not significant. Seventeen of the 22 significant correlations significant correlations had correlation coefficients below our criterion threshold of .50, and therefore, not included in our analyses since the strength of the relationship was considered relatively weak. However only those correlations that were significantly and highly correlated across all three visits have been reported. For the five significant correlations that were above our threshold criterion of .50, there was no consistent pattern of results that emerged.

Chapter 4

4 Discussion

This study examined the temporal variability of speech intensity measures, speech intelligibility measures, and communicative participation measures in individuals with hypophonia and PD. This study also explored the relationships among speech intensity measures, speech intelligibility measures, self- and proxy-rated communicative participation, demographic factors, and non-speech factors. The first objective of this study addressed the temporal variability of speech intensity measures for participants with PD and control participants. The second objective of this study addressed the variability of perceived typical speech loudness for participants with PD, their primary communication partners, and control participants over three time points. The third objective of this study addressed the variability speech intelligibility measures for participants with PD and control participants. The fourth objective of this study addressed the variability of self-rated communicative participation measures for participants with PD and control participants. The fifth objective of this study addressed the variability of self- and proxy-rated communicative participation measures of participants with PD and their primary communication partners. The sixth objective of this study addressed the retest reliability and repeatability of measures of speech intensity, speech intelligibility, and communicative participation in participants with PD, their primary communication partners, and control participants. The seventh objective of this study addressed the relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD. The eighth objective of this study addressed the relationships among speech intensity measures, speech intelligibility measures, proxy-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD. Finally, the ninth objective of this study addressed the relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for control participants.

Several hypotheses were examined in relation to the variability of speech intensity measures, speech intelligibility measures, and communicative participation measures. It was predicted that IWPD would demonstrate greater temporal variability of speech intensity measures, speech intelligibility measures, and self-rated communicative participation measures compared to control participants. Greater variability in measures of speech intensity for IWPD was predicted in the present study as a previous study by Schalling and colleagues (2013) suggested the possibility of increased day-to-day variability and individual differences in IWPD. Greater variability in measures of speech intelligibility and self-rated communicative participation for IWPD was predicted in the current study, as a study by Fox and Ramig (1997) suggested that IWPD may demonstrate increased variability and individual differences in self-rated speech intelligibility and participation compared to control participants.

Additionally, it was hypothesized that IWPD would have reduced measures of speech intensity, speech intelligibility, and self-rated communicative participation compared to control participants. It was further predicted that self-rated typical speech loudness and self-rated communicative participation would be reduced for IWPD compared to proxy ratings made by their primary communication partners. Reduced measures of speech intensity were predicted in the present study as previous studies have reported that IWPD and hypophonia demonstrate reduced habitual speech intensity ranging from 2 – 5 dB SPL in IWPD and hypophonia (Adams et al., 2010; Dykstra et al., 2012a; Dykstra et al., 2015; Fox & Ramig, 1997; Tjaden et al., 2013), reduced maximum intensity (Dykstra et al., 2012a), parallel but attenuated Lombard response (Adams et al., 2006a; Adams et al., 2005; Dykstra et al., 2012a), a reduced magnitude production function (slope; Clark et al., 2014), and reduced ratings of self-loudness level (Andreetta et al., 2016; Clark et al., 2014; Fox & Ramig, 1997; Ho et al., 2000) compared to control participants. Differences between self- and proxy-rated typical speech loudness for IWPD was predicted as previous studies have suggested that IWPD may not be aware of their speech intensity deficits (Clark et al., 2014; Ho et al., 2000). Reduced measures of speech intelligibility were predicted in the present study as previous studies have found reduced speech intelligibility in IWPD and hypophonia compared to healthy speakers (Adams et al.,

2008; Andreetta et al., 2016; Feenaughty et al., 2014; Stipancic, Tjaden, & Wilding, 2016; Sussman and Tjaden 2012; Tjaden, Sussman, & Wilding, 2014). Reduced measures of self-rated communicative participation were predicted in the present study as previous studies have found that IWPD rate their communicative participation lower in comparison to control participants (Donovan et al., 2008; Dykstra et al., 2015; Fox & Ramig, 1997;). Differences between self- and proxy-rated communicative participation was predicted as several previous studies reported lower proxy-rated PROs compared to self-rated PROs for IWPD (Donovan et al., 2008; Simberg et al., 2012).

With regard to exploring relationships among the variables of interest in this study, it was hypothesized that measures of speech intensity, speech intelligibility, and self- and proxy-rated communicative participation would all be correlated with one another. It was also predicted that increased disease duration, decreased perceived medication effectiveness, and reduced cognitive abilities would be correlated with measures of speech intensity, speech intelligibility, and self- and proxy-rated communicative participation in IWPD. Correlations among measures of self- and proxy-rated communicative participation were predicted as previous studies have identified relationships between the CPIB and other functional PROs measuring communicative participation (McAuliffe et al., 2016). Correlations among demographic and non-speech factors and measures of speech intensity, speech intelligibility, and self- and proxy-rated communicative participation were predicted as previous studies have identified associations between self-rated communicative participation and speech severity, fatigue, cognitive difficulties, and emotional difficulties (McAuliffe et al., 2016), disease severity and self-rated communicative participation (Donovan et al., 2008), self-rated communicative participation, age and gender (Yorkston et al., 2008), as well as gender and speech intensity (Sapir et al., 2007) in IWPD.

The following sections will discuss the primary results of the present study and will relate these results to those of previous published studies. Subsequent sections will discuss limitations of the present study, recommendations for future studies, and the implications for clinical work and research.

4.1 Objective 1: Speech Intensity

This objective examined the temporal variability of habitual speech intensity, maximum speech intensity, Lombard response function, and magnitude production in order to determine whether speech intensity measures differ over three time points between and within participants with PD and control participants.

4.1.1 Habitual speech intensity

The comparison of habitual speech intensity scores revealed no significant differences between the marginal means of habitual speech intensity of participants with PD and control participants. These findings are consistent with other studies that found similar measures of habitual speech intensity in IWPD (Matheron, Stathopoulos, Huber, & Sussman, 2017; Tjaden et al., 2013). Tjaden and colleagues (2013) examined speech acoustics and speech intelligibility during "habitual", "clear", and "loud" speech in 13 IWPD and 15 healthy control participants. During the "habitual" speech condition, participants were asked to use their typical speech. During the "clear" speech condition, participants were asked to speak using a voice that was two times clearer than their typical speech. During the "loud" speech condition, participants were asked to speak using a voice that was two times louder than their typical speech. The authors reported that the speech intensity of IWPD was significantly reduced in the "clear" and "loud" speech conditions, but not in the "habitual" speech condition, compared to control participants.

Additionally, Matheron and colleagues (2017) asked 42 IWPD and hypophonia and 20 control participants to perform a speech task in no added noise and in the presence of multi-talker background noise in order to investigate laryngeal performance. The authors adjusted the intensity of background noise for each participant until their speech intensity was 3 – 5 dB SPL greater than their habitual speech intensity (Matheron et al., 2017). Their study revealed no significant differences between IWPD and hypophonia and control participants in no noise conditions (Matheron et al., 2017).

However, a significant effect of "Visit" and a significant "Group" by "Visit" interaction was found, where the marginal means of habitual speech intensity for participants with

PD were significantly lower than the marginal means of habitual speech intensity for control participants at visit 1. No significant differences were found at visit 2 or visit 3. The difference in speech intensity during visit 1 was approximately 3 dB SPL less for participants with PD compared to control participants. This difference is consistent with previous literature, which reported a 2 – 5 dB SPL reduction in the speech intensity of IWPD and hypophonia compared to healthy speakers (Adams et al., 2006a; Adams et al., 2005; Adams et al., 2006b; Adams, Winnell, & Jog, 2010; Dykstra et al., 2012a; Fox & Ramig, 1997; Matheron, et al., 2017).

Fox and Ramig (1997) examined the speech intensity of 30 IWPD, and 14 control participants. Participants attended three visits during a four-day period, wherein they completed four speech tasks at every visit. The results demonstrated that the speech intensity of IWPD was 2-4 dB SPL lower than that of control participants.

Additionally, Adams and colleagues (2010) explored the impact of interlocutor distance, background noise, and a concurrent task on speech intensity in 10 participants with PD and hypophonia and 14 control participants. In their study, participants engaged in conversation with an experimenter in 12 different experimental conditions – including: conversing in no noise, conversing in 50 and 65 dB SPL of background noise, conversing at an interlocutor distance of 1 m and 6 m, and conversing while performing a concurrent manual task (Adams et al., 2010). Adams and colleagues (2010) reported a 3 – 4 dB SPL reduction in the speech intensity of participants with PD compared to control participants.

Finally, Dykstra and colleagues (2012a) examined the habitual conversational speech intensity of 30 individuals with hypophonia and 15 control participants. The authors found that participants with PD demonstrated reduced habitual intensity of approximately 5 dB SPL compared to control participants (Dykstra et al., 2012a). However, although similar findings were reported, it is important to note that Dykstra and colleagues (2012a) explored conversational speech, while the present study asked participants to perform a reading task.

The results described above (Adams et al., 2005; Adams, Winnell, & Jog, 2010; Dykstra et al., 2012a; Fox & Ramig, 1997) are consistent with the significant group difference found in visit 1 of the present study, but not with the non-significant group difference found for the overall, grand mean, involving all three visits. The results of the current study suggest that following the completion of the speech tasks and communication questionnaires in visit 1, participants with PD may have been more aware of their speech production and may have focused on increasing their overall speech intensity or used greater effort during speech tasks during visits 2 and 3. It may also be possible that a performance effect occurred in participants with PD following visit 1. Performance effects have been previously observed in IWPD (Goberman & Elmer, 2005; Keintz, Bunton, & Joit, 2007).

A study by Keintz and colleagues (2007) explored the impact of visual information on the speech intelligibility of eight IWPD and hypokinetic dysarthria. Participants were recorded while reading 120 sentences in a formal research setting. Listeners later orthographically transcribed each sentence in order to obtain a measure of sentence intelligibility. The conversational speech intelligibility of four of these eight participants was also assessed outside of the formal research testing situation. Ratings of speech intelligibility were not consistent between these two evaluations, with greater speech intelligibility observed in a research testing scenario compared to ratings obtained during typical conversational intelligibility performance. The authors suggested that IWPD exhibit a performance effect, wherein IWPD work to be more intelligible in clinical or research settings compared to when engaged in casual conversation (Keintz et al., 2007). It is possible that the improvement in speech intelligibility in testing situations is related to the deliberate use of clear speech by IWPD in testing contexts (Goberman & Elmer, 2005). It is possible that following visit 1, participants were aware that the loudness of their speech was being studied. Therefore, participants may have been more focused on deliberately using clearer or louder speech during subsequent visits. This increased focus and awareness on their speech may have resulted in the performance effect described above.

This variability across visits for participants with PD is consistent with the poor ratings of retest reliability calculated via ICC for both participants with PD and control participants. However, habitual speech intensity measures were found to have good repeatability, as measured by CR and CR%, for both participants with PD and control participants. Differences in a measure greater than the CR reflect the value under which the difference between any two repeat measurements on the same individual obtained under identical conditions should fall within 95% probability rather than detecting a difference that may be the result of measurement error (Vaz, et al., 2013). Previous studies have proposed that CR% values less than or equal to 10% are indicative of good repeatability (Lu et al., 2007; Smidt, et al., 2002; Steffen & Seney, 2008). However, the studies that proposed CR% values indicative of good repeatability pertain to the pain, strength, mobility, and activities of daily living literature relating to populations with lateral epicondylitis, Parkinsonism, and chronic stroke (Lu et al., 2007; Smidt et al., 2002; Steffen & Seney, 2008). The current study has adopted the proposed CR% values less than or equal to 10% as indicative of good repeatability. The current study also proposes CR% values between 11 – 20 % to be indicative of marginal repeatability, and CR% values greater than 20% to be indicative of unacceptable repeatability. However, these values should be viewed as a starting point for developing appropriate CR and CR% within speech and communication measures and contexts for IWPD and hypophonia.

As previously stated, results from the current study suggest habitual speech intensity measures were found to have good repeatability. These findings suggest that observed changes greater than approximately 5 – 7 dB over time would be beyond the estimated error of measurement and therefore reflect "real" changes in habitual intensity (i.e. related to treatment or experimental conditions) for both participants in PD and control participants. It should be noted that, while a change of 5 – 7 dB falls below the CR% value of 10%, such a change appears to be quite large within the context of previously reported treatment effects for habitual speech intensity in IWPD. For example, multiple studies have revealed that LSVT resulted in an increase of approximately 4 dB in the speech intensity of IWPD and hypophonia (Gustafsson, Södersten, Ternström, & Schalling, 2019; Ramig, Countryman, O'Brien, Hoehn, & Thompson, 1996; Ramig,

Countryman, Thompson, & Horii, 1995; Ramig, Fox, & Sapir, 2004; Ramig et al., 2001). Based on the previous treatment studies, it appears that a clinically meaningful difference in habitual intensity is lower than the minimum CR% of 10% that was tentatively adopted in the present study. A 4 dB difference is only slightly more than a CR% of about 5% (i.e., 70 dB mean / 4 dB change = 5.7%). As a result, it may be more beneficial to designate a CR% value of 5% as a measure of appropriate and acceptable repeatability in the measurement of treatment outcome in hypophonia secondary to PD. Therefore, based on the results of previous outcome studies (Gustafsson et al., 2019; Ramig et al., 1996; Ramig et al., 1995; Ramig et al., 2004; Ramig et al., 2001), CR% values of 5% may be more appropriate for measure of habitual speech intensity. If a CR% value of 5% were to be applied to the measure of habitual speech intensity in the present study, then habitual speech intensity would be deemed to demonstrate poor or marginal repeatability for both participants with PD and control participants. In general, it appears that additional retest studies are required to verify the estimated CR% of 10% for habitual intensity. In addition, if a CR% of 10% is verified, it will be important to determine if there are refinements that can be introduced into the measurement of habitual intensity that will reduce the measurement error to a value that is less than the clinically meaningful change in habitual intensity, that is estimated to be slightly above a CR% of 5%.

4.1.2 Maximum speech intensity

The comparison of maximum speech intensity scores revealed a significant difference between the maximum speech intensity of participants with PD and control participants, wherein the overall marginal means of participants with PD was 4.68 dB SPL lower than that of control participants. These results suggest that control participants have the ability to voluntarily increase their speech intensity to a greater level than participants with PD. A previous study by Adams and colleagues (2006b) revealed a similar but slightly greater difference between the maximum speech intensity of IWPD and control participants than what is reported in the current study. Adams and colleagues (2006b) asked 10 IWPD and 10 control participants to repeat the phrase “I owe you a yo-yo, I owe you a yo-yo” using their maximum speech intensity. The authors reported that the maximum speech intensity of IWPD was reduced by 6.4 dB compared to control participants (Adams et al., 2006b).

Additionally, the observed difference of 4.68 dB between groups of participants in the present study is smaller than the difference that was reported in a 2012a study by Dykstra and others. Dykstra and colleagues (2012a) examined maximum conversational speech intensity of 30 individuals with hypophonia and 15 control participants. The authors found that participants with PD demonstrated reduced maximum intensity of approximately 10 dB SPL as compared to control participants (Dykstra et al., 2012a). A closer examination of the mean maximum speech intensity levels of participants with PD and control participants revealed a difference between the participants in the study by Dykstra and colleagues (2012a) compared to the current study. The mean maximum speech intensity levels of participants with PD were similar between the two studies, with intensity scores of 76.59 and 78.20 dB SPL in the study by Dykstra and colleagues (2012a) and the present study, respectively. The mean maximum speech intensity levels of control participants PD were more dissimilar between the two studies, with intensity scores of 87.07 and 82.88 dB SPL in the study by Dykstra and colleagues (2012a) and the present study, respectively. Thus, while participants with PD were found to produce similar maximum speech intensity levels in both studies, the control participants in the study by Dykstra and colleagues (2012a) produced greater levels of maximum speech intensity compared to the control participants in the present study.

Furthermore, measures of maximum speech intensity were consistent over time for both participants with PD and control participants. These findings are consistent with the finding that maximum speech intensity demonstrated good retest reliability and good repeatability for participants with PD and control participants. Thus, maximum speech intensity is a reliable measure to evaluate changes over time in IWPD and healthy speakers, wherein observed changes (i.e. those associated with a treatment or experimental condition) of greater than approximately 2 – 6 dB and 3 – 5 dB would be above the estimated error of measurement (95% confidence range) for maximum speech intensity and would therefore reflect a true change in the IWPD and control participants, respectively. Although the present study found that a change of 2 – 6 dB (CR% = 3-7%) falls below the tentatively defined acceptable CR% value of 10%, it is reasonable to further evaluate this CR% in terms of a minimum clinically important difference within

the context of treatment effects for maximum speech intensity in IWPD and hypophonia. A 2016 study by Tanner, Rammage, and Liu examined the impact of a singing-based treatment program for IWPD. The authors reported that participants demonstrated a significant increase of 7 dB (estimated to be about a 9% increase) in their maximum speech intensity following the completion of the treatment program. The CR% results of the current study suggest that the 7 dB change reported by Tanner and colleagues (2016) would reflect about a 9% change and that this value falls above the greatest CR value of 6 dB (CR% = 7%) obtained in the present study. Thus, it appears that the tentative interpretation of a CR% of 10% as an acceptable repeatability value for maximum intensity may be a bit too liberal. Since the results of the Tanner and colleagues (2016) study indicate a clinically meaningful difference of 9% for maximum speech intensity, it is suggested that a CR% of less than 9% should be used as the cutoff for a good repeatability related to maximum intensity in IWPD. Therefore, based on the CR% of 7% found in the present study it is concluded that maximum intensity is a measurement of IWPD that shows good reliability and repeatability.

4.1.3 Lombard response function

The comparison of Lombard response function revealed that participants with PD and control participants demonstrated similar slopes of their Lombard response when speaking in 60, 65, 70, and 75 dB SPL of multi-talker background noise. This similarity of Lombard response function was expected as previous studies found that the Lombard response function in IWPD was similar to the response of control participants (Adams et al., 2006a; Adams et al., 2005; Dykstra et al., 2012a).

Adams and colleagues (2005) examined the relationship between speech intensity and type of background noise. In their study, 10 IWPD and hypophonia and 10 control participants repeated sentences in five multi-talker background noise conditions. The authors found that both groups of participants demonstrated a Lombard response. Furthermore, Adams and colleagues (2005) reported a lack of “Group” by “Noise” interaction, suggesting that the slope of the Lombard response of participants with PD was parallel compared to control participants.

Additionally, Adams and colleagues (2006a) investigated the effect of different types of background noise on speech intensity in 23 IWPD and hypophonia and 15 control participants. Adams and colleagues (2006a) asked participants engage in two minutes of conversation while in the presence of three types of background noise (i.e., multi-talker noise, instrumental music, and pink noise) presented at five different intensity levels ranging from 50 – 70 dB SPL. Similar to the findings of the present study and the finding of Adams and colleagues (2008), the authors reported that the Lombard sign was elicited in both groups of participants. However, due to a lack of "Group" by "Noise" interaction, Adams and colleagues (2008) suggested that the Lombard response function of participants with PD ran parallel to that of control participants.

Finally, Dykstra and colleagues (2012a) explored the effect of various intensity levels of background noise (ranging in 5 dB increments from 50 – 70 dB SPL) on the conversational speech intensity of 30 IWPD and hypophonia and 15 control participants. The authors reported that both groups of participants demonstrated a Lombard response during the conversational speech task in background noise. Additionally, no significant "Group" by "Noise" interaction was found. Thus, Dykstra and colleagues (2012a) also suggested that the Lombard response pattern in IWPD was parallel compared to the response pattern of control participants.

These studies also reported that while the speech intensity of IWPD was less intense compared to control participants, the Lombard response slope of IWPD paralleled that of control participants (Adams et al., 2006a; Adams et al., 2005; Dykstra et al., 2012a). The results from the present study are consistent with the findings from the above mentioned studies and suggests that IWPD may increase their speech intensity in the presence of background noise in a manner than parallels the Lombard response of control participants.

Furthermore, a significant effect of visit was found in the present study, where the Lombard response function at visit 1 was significantly greater than the Lombard response function at visits 2 and 3. This pattern of results is similar to what was observed in the analysis of habitual speech intensity. However, since habitual speech intensity was not

used to calculate Lombard response function, the performance effect observed in measures of habitual speech intensity for participant with PD cannot explain the possible performance effect observed in Lombard response function. It is possible that a performance effect similar to the effect observed with the analysis of habitual speech intensity occurred across participants for the Lombard response. Participants with PD may have habituated to the background noise and produced the target sentence, "She saw Patty buy two poppies" with a smaller range in their speech intensity during the task, thereby resulting in a flatter Lombard response function by their third visit as compared to their first visit. As previously discussed, the empirical literature suggests that IWPD can exhibit a performance effect, wherein increased effort is used by IWPD increase their intelligibility in clinical or research setting, often by using clearer and louder speech (Goberman & Elmer, 2005; Keintz et al., 2007). Thus, it is possible that participants with PD paid greater attention to their speech loudness following visit 1, resulting in greater levels of speech intensity across noise conditions in subsequent visits. These findings appear to be consistent with the performed retest analysis, as it was found that the Lombard response function did not demonstrate good retest reliability and unacceptable repeatability in IWPD. It is possible that a portion of the variability observed in measures of speech intensity, such as habitual speech intensity and Lombard response function, may be related to a performance effect rather than temporal variability. Additionally, this is the first study to attempt to develop and apply a measure of true change compared to measurement error of the function of Lombard response in IWPD. The approach taken in the present study was to calculate the slope of the Lombard response in several multi-talker background noise conditions: 60, 65, 70, and 75 dB SPL. Further studies are required to continue to refine this slope method, and perhaps develop alternate methods of measuring the Lombard response function.

4.1.4 Magnitude production

The comparison of the magnitude production function revealed that participants with PD demonstrated a significantly smaller slope than did control participants. These findings are consistent with a previous empirical study (Clark et al., 2014).

Clark and colleagues (2014) investigated loudness perception in IWPD and hypophonia. Seventeen participants with PD and 25 control participants took part in a magnitude production task, wherein participants read 5-word target sentences at their habitual speaking volume. The speech intensity used for this initial reading was assigned a value of 100. Participants then reproduced the sentence in varying magnitudes (25, 50, 100, 200, and 400) compared to their initial performance. The authors reported that IWPD made smaller adjustments to their speech intensity compared to control participants across all levels of magnitude production. The results of the present study as well as those of Clark and colleagues (2014) suggest that the slope of the magnitude production function is less steep in IWPD compared to control participants. Additionally, the results of loudness perception of Clark and colleagues (2014) suggest that IWPD may demonstrate a deficit in their perception of loudness levels. It is possible that this deficit of loudness perception may in turn influence the ability of IWPD to plan and produce speech at greater intensities, thereby resulting in a shallower magnitude production slope.

Furthermore, a significant effect of visit was found in the present study, where the magnitude production function at visit 1 was significantly greater than the magnitude production function at visits 2 and 3. Furthermore, the mean magnitude production slope for participants with PD was significantly lower than that of controls by 0.81 at visit 1 only. This pattern of results is similar to what was observed in the analysis of habitual speech intensity and Lombard response function. Contrary to the calculations of Lombard response function, the measure of habitual speech intensity was used to calculate the slope of magnitude production. It is possible that visit differences in habitual speech intensity may have subsequently affected magnitude production function and accounted for a significant proportion of the variability observed in the measure. If a performance effect occurred in the measure of habitual speech intensity, one might have expected the performance effect to carry over to the other speech tasks involved in calculating magnitude production function. Additional research is needed in order to explore the role of habitual or reference speech intensity in the performance of magnitude production tasks and the calculation of magnitude production function.

Finally, the variability observed across visits for participants with PD is consistent with the unacceptable repeatability values found in the present study. However, magnitude production was found to have good retest reliability results for both participants with PD and control participants. In order to calculate the retest reliability of a measure, interclass correlations were used. These correlations provide information regarding the strength of the association between retests, and do not take into account any measurement error that may be present in the results. However, the CR is calculated using both the correlation and measurement error, resulting in a better estimate of the similarities of retest measures are likely to be for any given measure. Since the results varied for the different retest analyses performed in this study, it is unclear whether measures of magnitude production demonstrate significant variability over time in IWPD and control participants. Additional research is needed to refine or design alternative methods of measuring magnitude production speech tasks that are associated lower measurement error.

4.2 Objective 2: Typical Speech Loudness

This objective examined the temporal variability of perceived typical speech loudness, measured on a VAS. A score of 0% on the VAS indicated that the participant perceived their speech as “very quiet”, and a score of 100% indicated that they perceived their speech as “normal loudness”. This objective sought to determine whether participants with PD perceived their typical speech loudness differently over time, whether perceived ratings differed from ratings made by primary communicative partners, and if ratings differed as compared to control participants. The comparison of perceived typical speech loudness revealed that self-perceived typical speech loudness in participants with PD was significantly quieter compared to control participants. This difference between control participants and participants with PD is consistent with the findings of previous studies (Andreetta et al., 2016; Clark et al., 2014; Fox & Ramig, 1997; Ho et al., 2000).

As part of their work assessing the efficacy of seven different speech amplification devices, Andreetta and colleagues (2016) measured self-perceived typical speech volume of 11 IWPD and hypophonia and 10 control participants. Participants indicated on a VAS scale, very similar to the one use in the present study, to indicate their typical speech

volume. The authors reported that IWPD rated their perceived typical speech intensity to be significantly lower compared to control participants.

Additionally, in a 1997 study, Fox and Ramig explored the perceptions of speech in IWPD while using a rating scale very similar to the one used to measure perceived typical speech loudness in the present study. Fox and Ramig (1997) asked 30 IWPD and 14 healthy control participants to rate nine perceptual variables, including perceived loudness, using a VAS at three separate time points. The authors reported that self-perceived speech loudness ranged from 51.62 – 57.80 % across visits for IWPD. These values are similar to marginal mean score of 51.32 % for participants with PD in the current study. Additionally, significant differences were found between perceived speech loudness for participants with PD and control participants in both the current study and the work of Fox and Ramig (1997). These results suggest that IWPD may be aware of the differences in loudness of their speech compared to control participants or IWPD may be aware of their hypophonia.

Ho and colleagues (2000) explored the perceived loudness level of self-produced speech stimuli of 15 IWPD and hypophonia and 15 control participants. Participants completed a reading task and engaged in conversation while using their soft, normal, and loud voice. Participants were then asked to adjust a volume control knob indicating the loudness level at which they had just spoken. These volume adjustments were completed immediately following each speech production and following a replay of the recorded sample. The authors found that IWPD significantly over-estimated their spoken loudness levels during both reading and conversational tasks compared to control participants.

In a 2014 study, Clark and colleagues compared loudness perception in 17 IWPD and hypophonia and 25 control participants. Prior to beginning a magnitude estimation task, a 70 dB SPL presentation of the target sentence was assigned a value of 100. Participants then rated the loudness of the target sentence when presented at 60, 65, 70, 75, and 80 dB SPL. The authors found that IWPD rated stimuli presented at higher intensity levels (75 and 80 dB SPL) lower than did control participants. Participants with PD were also observed to rate stimuli presented at lower intensity levels (60 and 65 dB SPL) higher

than did control participants. Clark and colleagues (2014) suggested that IWPD have a flatter psychophysical loudness function and a more restricted range of intensity perception than control participants. As a result, the authors concluded that IWPD may demonstrate a deficit in the perception of their loudness levels, which may in turn influence their ability to produce speech at greater intensities. The reported differences between perceived and produced speech intensity levels in the studies by Clark and colleagues (2014) and Ho and colleagues (2000) may suggest impaired sensorimotor integration deficits in individuals with hypophonia and PD.

However, it is important to note that both Clark and colleagues (2014) and Ho and colleagues (2000) asked their participants to perform loudness perception tasks that incorporated speech and auditory stimuli. In contrast, participants in the current study were asked to indicate their typical speech loudness via a questionnaire. It is possible that the observed difference in loudness perception between participants with PD and control participants appears to remain constant, regardless of the modalities used to measure perceived speech loudness. However, the relationship between ratings of self-perceived speech loudness, loudness ratings of self-productions of speech stimuli, and loudness ratings of externally presented speech stimuli is unknown. Additional research is recommended to explore this relationship, as well as identify the sensorimotor processes involved in each type of task.

Additionally, it was interesting to note that self- and proxy-rated perceived typical speech loudness for participants with PD were quite similar, wherein the average scores of self- and proxy-rated typical speech loudness was found to be approximately 51% for IWPD. As previously mentioned, the findings for perceived speech loudness for IWPD in the present study were similar to the results of Fox and Ramig (1997). The findings of the current study lend support to the aforementioned suggestion that IWPD may be aware of the differences in their speech loudness (Fox & Ramig, 1997). These findings contradict the hypothesis that IWPD are unaware of their loudness deficits and may subsequently be unable to appropriate adjust their speech intensity during speech production (Clark et al., 2014; Ho et al., 2000). Further research is needed to determine whether IWPD accurate perceive their speech loudness or demonstrate a lack of awareness of their reduced speech

loudness. Furthermore, participants rated their overall perception of typical speech loudness in the present study. This measure did not address potential deficits in the perception of moment-to-moment speech loudness. Further studies are also recommended to consider exploring differences in the overall versus moment-to-moment perceptions of typical speech loudness.

Finally, while the lack of significant effect of visit might be indicative of consistency in the measure of perceived typical speech loudness, the results of the retest analyses do not support this interpretation. It was found that perceived typical speech loudness did not demonstrate good retest reliability or acceptable repeatability in participants with PD, control participants, or primary communication partners of IWPD. Additional research may be required in order to refine or design alternative methods of measuring self- and proxy-perceptions of typical speech loudness that are associated lower measurement error.

4.3 Objective 3: Speech Intelligibility

This objective examined the temporal variability of SIT transcription scores, SIT VAS scores, and VAS conversational intelligibility scores in order to determine whether speech intelligibility measures differ over time between and within participants with PD and control participants.

4.3.1 SIT transcription scores

The comparison of SIT transcription scores revealed that the marginal means of participants with PD demonstrated a significant reduction in their speech intelligibility compared to control participants. This finding is consistent previous findings in IWPD and hypophonia that demonstrate reduced speech intelligibility (Adams et al., 2008; Cannito et al., 2012; Sussman & Tjaden, 2012).

Adams and colleagues (2008) explored speech intelligibility via the SIT in 25 IWPD and hypophonia and 15 control participants. Two listeners then orthographically transcribed the SIT sentences for all participants. The authors reported that the speech intelligibility of participants with PD was significantly reduced compared to control participants. The orthographically transcribed SIT score reported by Adams and colleagues (2008), 92.2%

for participants with PD and 99.3% for control participants, were similar to those found in the present study, 94.20% for participants with PD and 98.12% for control participants.

Additionally, a study by Cannito and colleagues (2012) examined sentence intelligibility via orthographic transcription of the SIT in eight individuals with hypophonia before and after a loudness-training program. The authors reported average sentence intelligibility of participants with PD to be 81.11% pre-treatment. The sentence intelligibility values in the present study are somewhat higher than those reported in the previous study (Cannito et al. 2012). It appears that the study conducted by Cannito and colleagues (2012) included a greater number of IWPD who demonstrated severe dysarthria. Therefore, it is possible that the discrepancy between ratings of speech intelligibility reported by Cannito and colleagues (2012) and the present study may be attributed to our lack of participants with PD with severe dysarthria. Thus, future research may be required to examine the temporal variability of speech intelligibility in IWPD with more severe hypophonia and dysarthria.

Sussman and Tjaden (2012) explored sentence intelligibility in dysarthric speakers with PD, dysarthric speakers with multiple sclerosis, and control participants. The authors assessed the sentence intelligibility of their participants via orthographic transcription of the SIT. Sussman and Tjaden (2012) reported that the speech intelligibility of female participants with PD was significantly reduced compared to control participants. However, the intelligibility scores of male participants with PD were not significantly different from control participants. One third of the participants in the present study were female, and yet a reduction in SIT transcriptions scores of participants with PD compared to control participants was observed. It may be that the participants with PD in Sussman and Tjaden's (2012) study demonstrated a greater variety of speech features related to hypokinetic dysarthria compared to those IWPD recruited for the present study who exhibited hypophonia as their primary dysarthric feature.

Finally, the non- significant difference in the overall mean SIT transcription scores across the three visits, suggests that these scores remained fairly consistent over time for both participants with PD and control participants. The consistency of the SIT transcription scores is further supported by the retest results. The retest results indicated that the SIT

transcription scores demonstrated good reliability and good repeatability for participants with PD. However, for the control participants, the retest analyses indicated that SIT transcription scores demonstrated good repeatability but did not demonstrate good retest reliability. Thus, SIT transcription scores are a reliable and repeatable measure (indicating an acceptable measurement error) and would be potentially useful for examining the effects of treatments or experimental conditions in IWPD. In order to determine if a CR% of 10% represents a clinically meaningful difference in the SIT transcription score, this difference value can be compared to previous treatment outcome studies of IWPD. A 2015 study by Martens and colleagues examined the impact of a speech rate and intonation treatment program on sentence intelligibility in IWPD. The authors reported that participants demonstrated a significant increase of 17.6 % in their sentence intelligibility following the completion of the treatment program. The CR% results of the current study suggest that the 17.6% change reported by Martens and colleagues (2015) is well above the CR value of 10 % that was obtained in the present study. Thus, the measurement error, as reflected by the CR%, appears to be well below the minimal clinically important CR% value that has been reported for one treatment study in IWPD (Martens et al., 2015).

In contrast to the Martens and colleagues (2015) study, the 2012 study by Cannito and colleagues described above may suggest that a clinically meaningful difference may be less than a CR% of 10%. Cannito and colleagues (2012) reported a 4.71% increase in the sentence intelligibility in IWPD following LSVT. This change of 4.71% falls below the CR% range proposed in the current study for demonstrating a clinically important difference. However, it should be noted that the LSVT program is focused on improving speech intensity and does not directly focus on the treatment of intelligibility. Thus, the Cannito and colleagues (2012) study may not be suitable for estimating the clinically meaningful difference or the CR% for SIT transcription scores. Further treatment outcome studies of intelligibility of IWPD involving the SIT transcription score as an outcome measure are required in order to provide a better estimate of the clinically meaningful CR%. At the present time, the CR% of 10% found in the present study suggests that SIT transcription has good reliability and repeatability and has a

measurement error that is low enough to be used to detect clinically meaningful differences in treatment outcome studies.

4.3.2 SIT VAS scores

The comparison of SIT VAS scores revealed that participants with PD demonstrated a significant reduction in their speech intelligibility compared to control participants. This finding is consistent with other empirical studies that have used VAS scaling to explore speech intelligibility in IWPD (Feenaughty et al., 2014; Tjaden et al., 2014). Tjaden and colleagues (2014) explored the impact of reduced rate of speech, increased speech loudness and the use of clear speech on sentence intelligibility in IWPD, individuals with multiple sclerosis, and control participants. Measures of speech intelligibility were obtained via VAS evaluations of 25 sentences from the Harvard Psychoacoustic Sentences [The Institute of Electrical and Electronic Engineers, 1969]. The authors reported that the sentence intelligibility of participants with PD was significantly reduced compared to control participants.

Additionally, Feenaughty and colleagues (2014) examined various acoustic measures and VAS ratings of sentence intelligibility in 12 IWPD and 12 control participants. Similarly to Tjaden and colleagues (2014), Feenaughty and colleagues (2014) used sentences from the Harvard Psychoacoustic Sentences [The Institute of Electrical and Electronic Engineers, (1969)] in their research. The authors reported that ratings sentence intelligibility of participants with PD were significantly lower compared to control participants. While the sentences used by Feenaughty and colleagues (2014) and Tjaden and colleagues (2014) differed from the current study, their findings that participants with PD demonstrate reduced speech intelligibility compared to control participants are consistent with the results of the present study.

Furthermore, a difference of approximately 15% was observed between SIT transcription scores and SIT VAS scores, wherein the SIT transcription scores were greater than the SIT VAS scores for participants with PD. Differences between SIT transcription scores and SIT VAS scores have been investigated in previous literature (Abur, Enos, & Stepp, 2018; Adams et al., 2008; Stipancic et al., 2016).

Stipancic and colleagues (2016) compared two different measures of sentence intelligibility in 16 IWPD, 30 individuals with multiple sclerosis, and 32 control participants. Fifty listeners orthographically transcribed and provided VAS ratings for the 25 Harvard psychoacoustic sentences [Institute of Electrical and Electronics Engineers, 1969] produced by each speaker. Speech intelligibility scores of IWPD were found to be approximately 25% lower than those of control participants for both orthographic transcription and VAS ratings. Additionally, VAS ratings were consistently found to be lower than orthographic transcription scores by approximately 10 – 15% across groups of speakers. Stipancic and colleagues (2016) then investigated the relationship between orthographic transcription scores and VAS ratings of sentence intelligibility. The authors reported strong associations between transcription scores and VAS ratings, ranging from .94 - .99 for participants with PD and ranging from .83 - .89 for control participants. The authors suggested that although the magnitude of scores for the measures of sentence intelligibility may differ, the overall patterns observed in these ratings remains constant. Abur and colleagues (2018) also explored the relationship between orthographic transcription scores and VAS ratings as measures of sentence intelligibility. Sixty-six naïve listeners either orthographically transcribed or provided VAS ratings for a subset of SIT sentences produced by 20 IWPD and five healthy speakers. The authors reported a strong relationship of .89 between SIT orthographic transcription and VAS scores. This relationship was consistent with the findings of Stipancic and colleagues (2016).

Additionally, Adams and colleagues (2008) examined sentence intelligibility via the SIT in 25 IWPD and hypophonia and 15 control participants. Two listeners then measured speech intelligibility via orthographically transcription and VAS ratings of the SIT sentences for all participants. The authors reported that the speech intelligibility of participants with PD was significantly reduced compared to control participants for both SIT transcription and VAS scores. Adams and colleagues (2008) reported that SIT VAS ratings were 3.4 % and 2.6% lower than SIT transcription scores for participants with PD and control participants, respectively. These differences are smaller than those differences observed in the current study, where SIT VAS ratings were 14.38 % and

7.21% lower than SIT transcription scores for participants with PD and control participants, respectively.

One possible reason for the discrepancy between SIT transcription and VAS scores may be due to SIT VAS scores capturing variables other than speech intelligibility in individuals with dysarthria, i.e. articulation impairments, comprehensibility, prosody, etc. (De Bodt et al., 2002; Kent, Kent, Vorperian, & Duffy, 1999; Kent et al., 1990). A 2002 study by De Bodt and colleagues investigated how individual dimensions influence overall perceptions of speech intelligibility in 79 dysarthric speakers. Participants presented with the following types of dysarthria: flaccid, spastic, ataxic, hypokinetic, and mixed. Speech samples were collected from participants via the Grandfather passage or segments of spontaneous speech. Two experienced listeners then rated the speech samples on 4-point ratings scales for the following dimensions: voice quality, articulation, nasality, prosody, and overall speech intelligibility. The authors found that articulation was very strongly associated with perceived intelligibility. The authors also reported that all four dimensions contributed in some degree to ratings of speech intelligibility in dysarthric speakers. Thus, additional research is required to further explore whether SIT VAS ratings include perceptions of factors other than speech intelligibility in IWPD and hypophonia.

Finally, as a result of the lack of significant difference in the mean of the SIT VAS scores across the three visits, it appears that these scores remain constant over time for participants with PD and control participants. The result of consistent SIT VAS scores over time is further supported by the findings from the retest analysis. The retest results suggested that the SIT VAS scores demonstrated good reliability and marginal repeatability for participants with PD and good reliability and good repeatability for control participants. Therefore, SIT VAS scores appear to be a reliable and fairly repeatable measure (indicating the possibility of an acceptable measurement error) for IWPD and healthy speakers, and would be potentially useful for examining the effects of treatments or experimental conditions in IWPD.

4.3.3 VAS conversational intelligibility scores

The comparison of SIT VAS scores revealed that participants with PD demonstrated a significant reduction of approximately 12% in their conversational intelligibility compared to control participants. This difference between groups appears to be consistent with, but smaller than the differences reported by Adams and colleagues (2008). Adams and colleagues (2008) examined speech intelligibility in IWPD and hypophonia and healthy controls in the presence of varying intensity levels of background noise. Twenty-five participants with PD and 15 age-matched control participants engaged in conversational speech tasks in various background noise conditions, including no added background noise, 60, 65, and 70 dB SPL. The conversational intelligibility scores of IWPD were approximately 20 – 30% lower than those of control participants across all noise conditions. The authors reported that conversational speech intelligibility scores in noise conditions were on average 5 – 10% lower than conversational intelligibility scores assessed in a quiet environment. Once this 5 – 10% adjustment is taken into account, the conversational intelligibility scores from the work of Adams and colleagues (2008) are consistent with the differences between IWPD and control participants in the current study.

In contrast, a study by Dykstra and colleagues (2012b) explored conversational speech intelligibility in background noise in IWPD and healthy speakers. Two listeners rated conversational speech intelligibility 30 participants with hypophonia and 15 control participants via VAS ratings. All participants engaged in conversation in four background noise conditions: no added background noise, 60, 65, and 70 dB SPL. The authors reported lower conversational intelligibility in IWPD compared to control participants across all noise condition. However, only the differences observed in 60, 65, and 70 dB SPL were statistically significant. It should be noted that a larger difference in conversational intelligibility scores was observed between groups in the present study compared to the work of Dykstra and colleagues (2012b). This discrepancy may be explained by any number of factors, including the presence of additional dysarthric features in participants with PD that may have had a greater impact on speech intelligibility or performance may have been influenced by the research testing environment (Keintz et al., 2007).

Additionally, the average VAS conversational intelligibility scores of participants with PD in the current study was approximately 81%. This finding is fairly consistent with the rating of SIT VAS scores observed in the current study. However, both of these intelligibility scores are approximately 12 – 15% lower than SIT transcription intelligibility scores. These results are fairly consistent with those of Adams and colleagues (2008). Adams and colleagues (2008) compared SIT transcription scores, SIT VAS scores, and conversational intelligibility transcription scores in IWPD and hypophonia and control participants. The authors reported that conversational intelligibility transcription scores were 9.7 % and 5.3% lower than SIT transcription scores for participants with PD and control participants, respectively. These differences are fairly similar those observed in the current study, where VAS conversational intelligibility scores were 13.46 % and 5.24% lower than SIT transcription scores for participants with PD and control participants, respectively. While the scores are similar, it is important to note that Adams and colleagues (2008) examined transcribed scores whereas the current study explored VAS scores. However, the overall pattern of lower conversational intelligibility scores compared to SIT transcription scores is consistent across both studies.

As previously discussed, a possible explanation for the discrepancy between VAS conversational intelligibility scores and SIT transcription scores may be related to VAS rating capturing other speech features in addition to speech intelligibility in individuals with dysarthria (Kent et al., 1999; Kent et al., 1990; De Bodt et al., 2002). Perceptions of voice quality, articulation, nasality, and prosody were found to contribute to ratings of speech intelligibility in dysarthric speakers (De Bodt et al., 2002). While it may be quite likely that other speech and voice features contribute to overall VAS ratings of speech intelligibility in IWPD, it is important to note that the participants in the study by De Bodt and colleagues (2002) presented with different types of dysarthria secondary to a variety of medical conditions. Additional research is recommended in order to further examine what other speech and voice dimensions may be assessed during VAS ratings of speech intelligibility.

Finally, based on the lack of a significant difference in the mean of the VAS conversational intelligibility scores across the three visits, these scores appear to be fairly consistent over time for both participants with PD and control participants. However, the consistency of the VAS conversational intelligibility scores is not supported by the retest results. The findings from the retest analyses were inconsistent. The retest results indicated that the VAS conversational intelligibility scores demonstrated good reliability and good repeatability for participants with PD. For the control participants, the retest results indicated that VAS conversational intelligibility scores demonstrated good repeatability but did not demonstrate good retest reliability. Thus, given the good reliability and good repeatability for participants with PD, these findings suggest that there is an acceptable level of measurement error in the VAS conversational intelligibility scores for their use in treatment outcome studies for IWPD. However, it is recommended that this retest analysis be replicated in future studies in order to verify the measurement error that was obtained for VAS conversational intelligibility scores in the present study.

Moreover, in order to further examine CR% and clinically meaningful difference in the VAS conversational intelligibility score, the CR% value of the current study can be compared to previous treatment outcome studies of IWPD. A 2016 study by Andreetta and colleagues examined the efficacy of seven speech amplification devices (ADDvox, BoomVox, ChatterVox, Oticon Amigo, SoniVox, Spokeman, and Voicette) in IWPD and hypophonia. The authors reported that VAS conversational intelligibility scores when participants were using the BoomVox was significantly greater compared to the other speech amplification devices, as well as no device speech conditions. Andreetta and colleagues (2016) also reported a 39% improvement in conversational intelligibility scores of participants with PD following treatment with the BoomVox. This clinically meaningful change of 39% is much greater than the highest CR% of 2.55 – 5.21% found in the present study. This finding suggests that for certain treatment outcome studies, such as those related to the use of speech amplification devices in IWPD, the VAS conversational intelligibility score has a measurement error that is well below the clinically meaningful difference scores that are expected to be associated with treatment outcomes.

4.4 Objective 4: Self-Rated Communicative Participation

This objective examined the temporal variability of CES question scores, VAPP subsection scores, CPIB scores, and LSUS scores in order to determine whether self-rated communicative participation differs over time between and within participants with PD and control participants.

4.4.1 Communicative Effectiveness Survey

The comparison of CES question scores revealed that overall, participants with PD self-rated their communicative effectiveness significantly lower than control participants. These findings are consistent with previous studies that also reported IWPD self-rated their communicative effectiveness lower than control participants on the CES (Donovan et al., 2008; Dykstra et al., 2015).

Donovan and colleagues (2008) compared self-rated communicative effectiveness via the CES in 25 individuals with hypokinetic dysarthria and PD and 25 control participants. The authors reported that IWPD had lower CES scores than control participants. These findings are similar to those reported in a previous study (Dykstra et al., 2015). Dykstra and colleagues (2015) explored self-rated communicative effectiveness in 30 IWPD and hypophonia and 15 control participants. The authors found that IWPD and hypophonia reported reduced scores on the CES compared to control participants. Furthermore, Dykstra and colleagues (2015) reported that participants with PD had lower scores on all CES items compared to control participants. These results are consistent with the current study, as univariate analyses revealed that participants with PD consistently rated themselves approximately 1 point lower than control participants on each of the eight communication situations included on the CES. These results suggest that participants with PD possess an awareness that they experience reduced communication effectiveness across different communicative situations.

A closer look at the results of the current study revealed that larger differences between groups were present in CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*), followed by CES question 8 (*Having a conversation with someone at a distance (across a room)*). These results are consistent with the

findings of Dykstra and colleagues (2015), who reported that IWPD and hypophonia reported reduced scores of communicative effectiveness in communicative contexts and situations related to conversing over distances and conversing in background noise. The authors suggested that there may be a predictable hierarchy from more to less challenging communicative situations that are unique to individuals with hypophonia. It is likely that talking in noise and talking at a distance demands greater speech intensity than the other situations that are examined in the CES (i.e. talking on the phone). Thus, it is likely that both of these communicative situations present as particularly challenging for IWPD and hypophonia.

Additionally, significant "Group" by "Visit" interactions were reported in the present study for each of the individual CES question – with the exception of question 3 (*Conversing with a familiar person over the telephone*) and question 7 (*Having a conversation while traveling in a car*). For each interaction, it was observed that self-rated communicative effectiveness increased over time for control participants, and decreased over time for participants with PD. It is possible that over the course of the study, participants became more aware of their communicative effectiveness in different communicative situations. During visits 2 and 3, multiple participants with PD and their primary communication partners reported that they had noticed and discussed with their partner particular trends in their communicative participation. Participants noted that they had not noticed or paid attention to communicative behaviours prior to being asked to consider and think about them in order to complete the communicative participation questionnaires. Miller and colleagues (2006) reported that when IWPD and hypophonia experience changes in their communicative functioning, such as reduced communicative participation, reduced confidence in their speaking abilities, and feelings of embarrassment at the reaction of others, they may be more likely to describe their communicative abilities as being impaired. They may also be concerned regarding the impact that changes in their speech production may have on their communicative effectiveness (Kwan & Whitehill, 2011; Miller et al., 2006). Miller and colleagues (2006) also reported that the negative perceptions held by IWPD of their communication partners' reactions to communicative interactions, including embarrassment, impatience,

agitation, and lack of appreciation of speaking difficulties faced by IWPD can have adverse effects on the IWPD's communicative participation and emotional response. It may be that the experience of completing self-ratings of their communication during the first visit led to subsequent reflection and an increased awareness of their reduced communication effectiveness, and this led to a lowering of their self-ratings of communication effectiveness when retested in visits 2 and 3. It is also possible that the experience of completing self-ratings of communication effectiveness, in the context of a study related to the communication in IWPD, may have made the control participants more aware of the positive aspects of their communication and this led to an increase in the ratings of communicative effectiveness in this cohort of participants.

Finally, the variability observed in CES question scores over time is consistent with the results of the retest analyses. Each of the individual questions of the CES did not demonstrate good retest reliability for the participants with PD or control participants – with the exception of CES question 3 (*Conversing with a familiar person over the telephone*) which demonstrated good retest reliability for control participants only. Furthermore, the retest analyses indicated that each of the individual questions of the CES demonstrated unacceptable repeatability for participants with PD. The retest analyses also revealed that the individual questions of the CES demonstrated unacceptable repeatability for control participants – with the exceptions of CES question 1 (*Having a conversation with a family member or friends at home*), CES question 3 (*Conversing with a familiar person over the telephone*), and CES question 7 (*Having a conversation while traveling in a car*) which demonstrated marginal repeatability. Given the poor retest reliability and repeatability results, it becomes difficult to interpret the significant group by visit interaction obtained for several of the CES questions. While it was previously suggested that the IWPD showed a decrease in CES values from visit 1 to visits 2 or 3 (and the control participants showed an increase in CES values from visit 1 to visits 2 or 3), it is possible that these changes are the result of measurement error and not the result of a 'real' change over time in the participants self-rating of communication effectiveness. This result is surprising as the CETI, the original questionnaire from which the CETI-M and CES were modified, was designed to track changes over time (Lomas et

al., 1989). However, it is possible that the CES in its present form (i.e. 4-point scale) may have a measurement error that is too high and thus unacceptable for the evaluation of treatment effects or the effects of experimental conditions in IWPD and hypophonia. It is recommended that future research explore whether modifications of the CES, such as having participants rate their communicative effectiveness using a VAS as opposed to a 4-point Likert scale, will increase the reliability and repeatability of the CES. Thus, it is recommended that additional research be conducted to refine or design alternative methods of measuring self-rated communicative effectiveness that are associated with lower measurement error.

4.4.2 Voice Activity and Participation Profile

The comparison of self-rated VAPP scores revealed that participants with PD had greater scores across all VAPP subsections compared to control participants. These results suggest that our participants with PD rated themselves with significantly more activity limitations and participation restrictions in comparison to control participants. Additionally, an overall significant effect of visit was reported in the present study. Subsequent univariate analyses revealed that significant changes in VAPP scores were observed over time for VAPP *social communication* and VAPP *participation restriction* scores for both participants with PD and control participants. For these two subsections, the scores of both groups of participants were greater at visit 1 than their scores at visit 2 but lower than those at visit 3. These observed changes were more prevalent in the scores of participants with PD compared to control participants. It would appear that the *social communication* and *participation restriction* subsections of the VAPP demonstrated greater variability and this was associated with the development of more severe self-ratings of communicative participation and participation restrictions over time.

As previously discussed, many participants with PD shared with the researcher during visits 2 and/or 3 the impression that they had become more aware of their communicative activity and communicative participation following the completion of the participation-based PROs after visit 1. Subsequently, participants with PD reported becoming more aware of the changes that had occurred within their speech production and communicative participation over the course of their disease. These anecdotal reports

are consistent with previous studies (Miller et al., 2006). As previously discussed, Miller and colleagues (2006) reported that perceived communicative impairments were associated with reduced communicative participation, reduced confidence in speaking abilities, and increased feelings of embarrassment at the reaction of others. The authors also found that the negative perceptions of their communication partners' reactions to the communicative interactions of the IWPD may result in adverse effects on the communicative participation and emotional response of the IWPD. These responses may lead to withdrawal, and barriers and restrictions in the communicative and social interactions of IWPD.

Finally, the observed changes in VAPP subsection scores over time were not consistent with the retest analyses. The retest analyses of self-rated VAPP subsection scores demonstrated good retest reliability across many VAPP subsections for participants with PD and control participants, with several exceptions. The VAPP *self-perceived voice problem* scores for participants with PD did not demonstrate good retest reliability. Additionally, the VAPP *self-perceived voice problem, social communication* and *participation restriction* scores for control participants did not demonstrate good retest reliability. Additionally, all VAPP subsection scores demonstrated unacceptable repeatability for participants with PD and control participants. Measures of retest reliability provide information regarding the strength of the relationship between retests, and do not account for the presence of measurement error. Whereas measures of repeatability incorporate both correlations and measurement error and result in better estimates of the similarities of retest measures. Given the poor retest reliability and repeatability results, it becomes difficult to interpret the significant visit effects obtained for the VAPP subsections. While the previous interpretation of the ANOVA results suggested the IWPD showed an increase in VAPP subsection scores from visit 1 to visits 2 and/or 3, it is possible that these changes are the result of measurement error and not the result of a 'real' change over time in the participants' VAPP-related self-ratings of communicative activity and participation. These inconsistent findings do not support the results of Ma and Yiu (2001), who reported good test-retest reliability for the VAPP. However, it is important to note that Ma & Yiu (2001) assessed the test-retest reliability

of the VAPP in individuals with dysphonia. It is possible that higher levels of measurement error may be present in the VAPP when administered to IWPD and hypophonia. Ma & Yiu (2001) also used Pearson's correlation coefficient in order to assess test-retest reliability. While these findings are consistent with most of the retest reliability results in the present study, it must be noted that neither Pearson's *r* nor ICC take into account the degree of measurement error present in the measures. Additional research is recommended in order to refine or design alternative methods of measuring the impact of an individual's voice impairment on activity limitation and participation restriction scores that are associated with lower measurement error in IWPD.

4.4.3 Communicative Participation Item Bank

The comparison of self-rated CPIB scores revealed that overall, participants with PD perceived their communicative participation to be lower than control participants' self-rated CPIB scores. No published studies have compared the CPIB scores of IWPD to control participants. Additionally, a significant "Group" by "Visit" interaction was found for CPIB scores. CPIB scores were observed to increase over time for control participants, and were observed to decrease over time for participants with PD. It is possible that over the course of the study, participants became more aware of their communicative participation. It may be that such an increased awareness resulted in the observed changes in perceived communicative participation over time for participants with PD and control participants. As previously stated, several participants with PD became more aware of the impact of their voice on their activity and communicative participation over the course of the study. This observation is consistent with the work of Miller and colleagues (2006), who reported that greater perception of communicative impairments which may in turn result in negative emotions, withdrawal, and barriers and restrictions to the communicative and social interactions of IWPD.

Finally, self-rated CPIB scores demonstrated good retest reliability and marginal repeatability for participants with PD, and good retest reliability and good repeatability for control participants. Thus, self-rated CPIB scores may be a reliable measure to evaluate changes over time in IWPD, wherein observed changes (i.e. those associated with a treatment or experimental condition) of greater than approximately 5 – 6 points

would be above the estimated error of measurement (95% confidence range) for standardized CPIB scores and would therefore reflect a true change in the IWPD. It is important to note that the CPIB was not intended to be used as a pre- and post-treatment outcome measure (Baylor et al., 2013). However, the good retest reliability and marginal repeatability findings from the present study suggest the possibility that the CPIB may be an effective and reliability measure to detect change over time in IWPD. Furthermore, these retest analyses suggest that the main effect of "Visit" and the "Group" by "Visit" interaction obtained from the above described ANOVA results are more likely to be related to a real effect compared to an effect of measurement error. These results also suggest that the similar "Group" by "Visit" interaction found in the CES analysis may reflect a real visit effect, despite the concerns regarding measurement error in the CES.

4.4.4 Level of Speech Usage Scale

The comparison of self-rated LSUS scores revealed that overall, participants with PD rated their level of speech usage lower than did control participants. As part of their evaluation of the efficacy of seven different speech amplification devices in IWPD, Andreetta and colleagues (2016) asked participants with PD and control participants to complete the LSUS. The authors found that level of speech usage reported by participants with PD was fairly evenly divided between *Routine – frequent periods of talking on most days, Intermittent – quiet for long periods of time on many days,* and *Undemanding – quiet for long periods of time almost every day.* However, their control participants reported their level of speech usage to be *Intermittent – quiet for long periods of time on many days* or *Routine – frequent periods of talking on most days.* In the present study, participants with PD typically rated their level of speech usage to be *Undemanding – quiet for long periods of time almost every day* or *Intermittent – quiet for long periods of time on many days.* Control participants typically rated their level of speech usage to be *Intermittent – quiet for long periods of time on many days* or *Routine – frequent periods of talking on most days.* These results suggest that IWPD rate their speech usage to be less frequent on a day-to-day basis compared to healthy speakers. Thus, IWPD may be engaging less and have more limited opportunities to participate in situations that rely on communication. Measures of participation may be used to evaluate the functional

outcomes of impaired activities, such as speech intensity and speech intelligibility (Baylor et al., 2009). The reduced measures of speech intensity and speech intelligibility found in IWPD in the present study may be related to the difference in the levels of speech usage observed in participants with PD and control participants. Level of speech usage may be a functional measure of the impact of communication challenges resulting from impaired speech intelligibility and deficits in the regulation of speech intensity in IWPD. However, it is important to note that many other factors, such as mild cognitive impairment, alterations in mood, and reduced mobility, may also contribute to reduced levels of speech usage in IWPD (McAuliffe et al., 2016).

Finally, self-rated LSUS scores appear to have been fairly consistent over time for both participants with PD and control participants. However, other measures of communicative participation, specifically the CES, VAPP, and CPIB, were observed to change over the course of the study. It is possible that participants' perceptions of their communicative participation shifted while the manner and degree to which they used their speech remained constant. The above reported stability in self-rated LSUS scores are inconsistent with the retest analyses. The results revealed that self-rated LSUS scores demonstrated good retest reliability for participants with PD but did not demonstrate good retest reliability for control participants. These results suggest that participants with PD are more reliable in their ratings of speech usage. It is possible that this difference between groups is related to the nature of the adjacent categorical levels on the LSUS that control participants may have varied across testing visits. It may be that the higher usage categories are more difficult for control participants to judge consistently, where the lower usage categories are easier to consistently select. Therefore, this discrepancy between groups may be related to the nature of the level items more than the consistency of the participant selections. Future studies may wish to explore item analysis of the LSUS in order to attempt to refine the reliability and validity of the measure. Additionally, the retest analyses also indicated that self-rated LSUS scores demonstrated unacceptable repeatability for both participants with PD and control participants. Thus, it is unclear whether self-rated LSUS scores demonstrate significant variability over time in IWPD and control participants. Additional studies are needed in order to refine or create

alternative methods of measuring self-rated levels of speech usage that are associated with lower measurement error.

4.5 Objective 5: Self- and Proxy-Rated Communicative Participation

This objective examined the temporal variability of CES question scores, VAPP subsection scores, CPIB scores, and LSUS scores in order to determine whether self- and proxy-rated communicative participation differ over time between and within participants with PD and their primary communication partners.

4.5.1 Communicative Effectiveness Survey

The comparison of self- and proxy-rated CES question scores revealed similar overall ratings between participants with PD and their primary communication partners. These findings are similar to those of Dykstra and colleagues (2015), and contrary to those of Donovan and colleagues (2008). Dykstra and colleagues (2015) compared self- and proxy-rated communicative effectiveness using the CES in 30 IWPD and hypophonia and their primary communication partners. The authors reported similar ratings of communicative effectiveness between IWPD and their primary communication partners. The findings of Dykstra and colleagues (2015) and the present study are in contrast to those of Donovan and colleagues (2008). Donovan and others (2008) compared ratings of communicative effectiveness via the CES in 25 individuals with hypokinetic dysarthria and PD and their 25 significant others. The authors noted that IWPD self-rated their communicative effectiveness significantly higher than did their communicative partners. The difference in the results of Donovan and colleagues (2008) compared to those of Dykstra and colleagues (2015) as well as the present study may be related to factors such as severity, or salient dysarthric features. For instance, in their study, Donovan and colleagues (2008) included IWPD with a diagnosis of hypokinetic dysarthria. These participants likely presented with a range of dysarthric features associated with hypokinetic dysarthria (i.e., articulatory imprecision, prosodic abnormalities, impairments in speech rate; Donovan et al., 2008). Whereas the current study as well as the study of Dykstra and colleagued (2015) recruited IWPD with hypophonia as their primary dysarthric feature. While Donovan and colleagues suggested that IWPD demonstrate

reduced awareness with respect to their communicative effectiveness, it may be that this hypothesis is not applicable in IWPD with hypophonia as their primary dysarthric feature.

Additionally, a significant effect of visit was found in the present study. Subsequent univariate analyses of the individual CES questions revealed a gradual reduction in ratings of communicative effectiveness over time. These significant findings were found for CES question 2 (*Participating in conversation with strangers in a quiet place*), CES question 4 (*Conversing with a stranger over the telephone*), CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*), and CES question 6 (*Speaking to a friend when you are emotionally upset or you are angry*). These findings suggest that there may be an unknown factor that is causing the reduction in ratings over time. However, it is possible that this reduction is due to random variation related to the measurement error. The findings of the retest analysis indicate that there was a fairly high level of measurement error. It was found that self- and proxy-rated CES questions do not demonstrate good retest reliability for the participants with PD or their primary communication partner – with the exception of CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) which demonstrated good retest reliability for the primary communication partners of participants with PD. Furthermore, the retest analyses indicated that each of the individual questions of the CES demonstrated unacceptable repeatability for both participants with PD and their primary communication partners. The results obtained from the primary communication partner for CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) can provide a useful illustration of the interpretive challenges that arise in the context of high levels of measurement error (i.e. poor repeatability). The univariate and post-hoc analysis revealed that there was a significant reduction in CES scores from visit 1 to visit 3 (-0.36) for the combined scores of participants with PD and their primary communication partners. Similarly, the *t*-test results related to the visit 1 to visit 3 comparison of proxy-ratings showed a reduction (-0.32) that approached significance. Thus, it appears that there may be a real reduction in CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) scores from visit 1 to visit 3. This is supported by the good reliability value (*ICC* = .75) for the visit 1 to visit 3 comparison.

However, the visit 1 vs visit 3 showed poor repeatability (*CR* = 1.09 and *CR%* = 44.86%). This poor repeatability was also shown for the other two visit comparisons (visit 1 versus visit 2, *CR* = 1.54 and *CR%* = 58.11%, and for visit 2 versus visit 3, *CR* = 0.89 and *CR%* = 37.71%). These CR values (0.89 to 1.54) are at least two times the value of the visit 1 to visit 3 difference value (0.32). Thus, these repeatability results indicate that the measurement error was probably too high to detect a real difference between the visit 1 and visit 3 CES scores for question 5 (*Being part of a conversation in a noisy environment (social gathering)*).

Given that the results for CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) showed some of the largest differences for the visit comparisons examined across all of the eight CES questions, it is likely that the high measurement error makes the interpretation of all of the visit comparisons related to the CES a challenge. In general, the high measurement error obtained for all of the communication partner CES questions makes it difficult to evaluate the across visit temporal variability of these CES results. Since the results varied for the retest analyses performed on the different CES questions in the present study, it is unclear whether CES questions scores demonstrate significant variability over time in IWPD and their primary communication partners. These findings are consistent with the above retest analyses findings for self-rated CES question scores for participants with PD and control participants. As previously discussed, these findings are unexpected as the CETI was designed to track changes over time (Lomas et al., 1989). Once again, it is possible that the CES in its present form (i.e. 4-point scale) may have a measurement error that is too high and thus unacceptable for the evaluation of treatment effects or the effects of experimental conditions in IWPD and hypophonia. Thus, it is recommended that future research explore whether changes of the CES, such as using VAS estimation as opposed to a 4-point Likert sale, will increase the reliability and repeatability of the CES. Additional research is needed to refine or create alternative methods of measuring perceived communicative effectiveness that are associated with lower measurement error.

4.5.2 Voice Activity and Participation Profile

The comparison of self- and proxy-rated VAPP scores revealed that both participants with PD and their primary communication partners held similar perceptions of the IWPD's voice problem, their activity limitations, and restrictions to their participation. These findings contradict those of Simberg and colleagues (2012), who reported that proxy-ratings of their partner's voice functioning was less severe compared to the self-ratings of IWPD. It may be possible that the discrepancy between the findings of Simberg and colleagues (2012) and the present study are related to differences in severity and dysarthric features of the IWPD in each study. Simberg and colleagues (2012) recruited participants to take part in a two-day treatment session, and likely had a more heterogenous groups of dysarthria symptoms and severity compared to the present study. Additional research exploring self- and proxy-rated perceptions of voice impairment in IWPD separated according to different dysarthric features may be warranted.

Additionally, the results of the aforementioned multivariate testing would suggest that the self- and proxy-scores of each subsection of the VAPP are consistent over time for participants with PD and their primary communication partners. However, the results of the retest analysis make it difficult to interpret these temporal variability results for the VAPP. Each of the individual subsections of the VAPP did not demonstrate good retest reliability for the participants with PD. However, each of the individual subsections of the VAPP demonstrated good retest reliability for the primary communication partners of IWPD – with the exception of VAPP *daily communication* and *total* scores which did not demonstrate good retest reliability. Furthermore, the retest analyses indicated that each of the individual subsections of the VAPP demonstrated unacceptable repeatability for both participants with PD and their primary communication partners. As previously discussed, measures of retest reliability provide information with regards to the strength of the association between retests, and do not factor the presence of measurement error. However, measures of repeatability use both the correlation and measurement error, resulting in a better estimate of the similarities of retest measures. Since the results varied for the different retest analyses performed in this study, it becomes difficult to determine whether self- and proxy-rated VAPP subsection scores are a reliability measure of activity limitation and participation restrictions over time in IWPD. These inconsistent

findings do not support the results of Ma and Yiu (2001), who reported good test-retest reliability for the VAPP. However as previously mentioned, Ma & Yiu (2001) collected data from individuals with dysphonia and used a different means of calculating test-retest reliability compared to the current study. Additional research is recommended in order to refine or create alternative methods of measuring the impact of an individual's voice on their activity limitations and participation restrictions that are associated with lower measurement error in IWPD.

4.5.3 Communicative Participation Item Bank

The comparison of self- and proxy-rated CPIB scores revealed similar perceptions of communicative participation in participants with PD and their primary communication partners. These results suggest that IWPD may demonstrate an awareness of their deficits in communicative participation comparably to their primary communication partners.

Additionally, the comparison of self- and proxy-rated CPIB scores revealed significant differences among the three visits. The scores at visit 1 were higher than the scores at visits 2 and 3. The scores at visit 2 were higher than the scores at visit 3. These results suggest that there is significant variability in self- and proxy-perceptions of communicative participation, as CPIB scores decreased over time for both groups of participants. It is possible that participants became more attuned to their level of communicative participation over the course of the study. As previously mentioned, many participants with PD and their primary communication partners shared that they were more conscious of communicative behaviours and participation in different communicative situations in the days and weeks that followed visit 1. As a result, participants may have overestimated their degree of communicative participation prior to the study. These observations are consistent with the aforementioned work of Miller and colleagues (2006), who reported that perceived impairments to and IWPD's communication was associated with reduced communicative participation, reduced confidence in speaking abilities, and increased feelings of embarrassment at the reaction of others. The authors also suggested that these negative feelings and perceptions may result in adverse effects on the communicative participation and emotional response of the IWPD. The increased variability reported may be associated with increased awareness

on the part of the participants with PD and their primary communication partners. Further research exploring this variability over time may shed some light on any shifts in awareness that occurred in participants with PD and their primary communication partners over time.

Finally, retest analyses of self- and proxy-rated CPIB scores demonstrated good retest reliability for participants with PD and their primary communication partners. Additionally, these scores also demonstrated good repeatability for participants with PD and marginal repeatability for the primary communication partners of IWPD. Thus, self- and proxy-rated CPIB scores may be a fairly reliable measure to evaluate self- and proxy-perceptions of change over time in IWPD, wherein observed changes (i.e., those associated with a treatment or experimental condition) of greater than approximately 5 – 6 and 6 – 11 would be above the estimated error of measurement (95% confidence range) for standardized CPIB scores, and would therefore reflect a true change in the scores of IWPD and their primary communication partners, respectively. As previously mentioned, the CPIB was not intended to be used as a pre- and post-treatment outcome measure (Baylor et al., 2013). However, the good retest reliability and good repeatability findings from the present study suggest that the CPIB may be an effective and reliability measure to detect change over time in IWPD. Furthermore, these retest analyses suggest that the main effect of "Visit" obtained from the ANOVA results are more likely to be related to a 'real' effect compared to an effect of measurement error. These results also suggest that the similar main effect of "Visit" found in the CES analysis may reflect a real visit effect, despite the aforementioned concerns regarding measurement error in the CES.

4.5.4 Level of Speech Usage Scale

The comparison of self- and proxy-rated LSUS scores revealed that participants with PD perceive their level of speech usage to be similar to the perceptions of their primary communication partners. These results suggest that IWPD and their primary communication partner demonstrate consistency in their perceptions of the degree to which IWPD use their speech on a day-to-day basis.

Finally, while the lack of a significant effect of visit might be indicative of the consistency in self- and proxy-rated LSUS scores, the results of the retest analyses suggest the need for a cautious approach to this interpretation. It was found that LSUS scores demonstrated good retest reliability for participants with PD but did not demonstrate good retest reliability for their primary communication partners. The retest analyses also indicated that self- and proxy-rated LSUS scores demonstrated unacceptable repeatability for participants with PD and their primary communication partners. Thus, it is inclusive whether self- and proxy-rated LSUS scores are a reliable measure to evaluate changes over time in IWPD. Additional study is needed in order to further establish the test-retest reliability of the LSUS. Given the poor retest reliability and repeatability results, it becomes difficult to interpret the results obtained for the LSUS. It is possible that the consistency of self- and proxy-rated LSUS scores is the result of measurement error and not the result of a 'real' change over time in the participants' perceived level of speech usage. Thus, it is recommended that additional research be conducted to refine or design alternative methods of measuring self- and proxy-rated levels of speech usage that are associated with lower measurement error in IWPD,

4.6 Objective 7: Inter-Relationships Among Variables in Participants with Parkinson's Disease

This objective examined inter-relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD.

4.6.1 Speech intensity measures

The correlations involving speech intensity measures revealed a moderate association between maximum speech intensity and magnitude production in participants with PD. These results suggest that the greater an IWPD's maximum speech intensity, the steeper their function of magnitude production as a greater range in their speech intensity would result in a sharper slope. A study by Clark and colleagues (2014) reported similar findings to the present study, in that IWPD demonstrated a smaller magnitude production slope compared to control participants. Clark and colleagues (2014) also suggested that

IWPD demonstrate a flatter psychophysical loudness function and a more restricted range of intensity perception compared to control participants. However, there was a sizable difference in the magnitude production slopes in the present study compared to the magnitude production slope in reported by Clark and colleagues (2014), 2.09 and 0.025, respectively. It is possible that a larger magnitude production slope was observed in the present study as the participants with PD might have presented with a greater range in their speech intensity compared to the IWPD who took part in the study by Clark and colleagues (2014). If IWPD are able to generate greater maximum speech intensities, it may directly influence their ability to increase their speech intensity during magnitude production tasks, resulting in a greater slope of magnitude production.

It is also interesting to note that while the maximum speech intensity and function of magnitude production of IWPD was found to be reduced compared to controls in the present study, both groups of participants demonstrated a similar relationship between these maximum speech intensity and magnitude production. While Clark and colleagues (2014) have explored magnitude production in IWPD, no published studies have explored the relationship between magnitude production and maximum speech intensity. The results of the current investigation are consistent with some of the findings of Clark and colleagues (2014), yet there are some key differences between the two studies. Clark and colleagues (2014) had 17 IWPD and hypophonia and 25 control participants complete a magnitude production involving a five-word target sentence as part of their examination of loudness perception. During the magnitude production task, participants read the target sentence at their habitual speaking volume Participants were then asked to reproduce the target sentence in varying magnitudes (25, 50, 100, 200, and 400), where 100 was their habitual speaking volume. Similar to Clark and colleagues (2014), functions of magnitude production were calculated by asking participants to repeat a sentence at two and four times louder than their habitual speaking volume. However, the present study did not ask participants to reduce their speaking volume as part of a magnitude production task. Yet both the present study and the works of Clark and colleagues (2014) arrived at the same conclusion, suggesting that the slope of the magnitude production function is less steep in IWPD compared to control participants. Additional research is

recommended to further explore the relationship between maximum speech intensity and magnitude production.

Finally, it is interesting to note that speech intensity measures and self-rated communicative effectiveness, as measured by the CES, were not found to be significantly correlated with one another at any of the three visits. This finding is consistent with those of Dykstra and colleagues (2015) who examined the relationship between speech intensity and self-rated communicative effectiveness in IWPD and hypophonia. Thirty IWPD and hypophonia and 15 control participants engaged in various speech tasks and completed the CES (Dykstra et al., 2015). The authors identified a non-significant relationship between speech intensity measures and self-rated communicative effectiveness (Dykstra et al., 2015). Dykstra and colleagues (2015) suggested that communicative participation is a distinct construct that differs from perceptual or acoustic measures of hypophonia, such as reduced speech intensity or reduced loudness. The results of the current study lend support to the possibility that communicative participation is a separate and distinctive construct from measures of speech intensity in the measurement of hypophonia in IWPD.

Previous studies have hypothesized that the presentation of hypophonia in IWPD may be causally related to deficits in sensorimotor integration, as well as sensory and somatosensory deficits, as evidenced by reduced speech intensity and loudness perception (Clark et al., 2014; Conte et al., 2013; Ho et al., 2000). It has been suggested that the dopamine-modulated subcortical structures of the basal ganglia and cortico-striatal pathways connecting the ipsilateral caudate nucleus and dorsolateral prefrontal cortex, in addition to non-dopaminergic pathways and neural structures such as the striatum, are involved in the regulation of muscle tone and coordination, scaling the force, amplitude, and duration of movements, as well as learning, planning, and initiating movements (Arnold et al., 2014; Braak et al., 2004; Duffy, 2013; Kempler & Van Lanker, 2002; Sapir, 2014). All of which can be attributed to the presentation of the development and presentation of hypokinetic dysarthia features associated with PD, including hypophonia (Arnold et al., 2014; Duffy, 2013; Sapir, 2014). Thus, it is possible that sensory, somatosensory, and sensorimotor integration deficits may be related to a lack of self-

awareness of reduced speech intensity. However, participants with PD did not demonstrate a lack of self-awareness of communicative effectiveness. This discrepancy in self-awareness may explain why speech intensity and CES scores were not associated with one another in the present study. This lack of significant correlations further supports the earlier suggestion that speech intensity and communicative participation are different constructs.

4.6.2 Speech intelligibility measures

The correlations involving speech intelligibility measures revealed the SIT transcription and SIT VAS scores were strongly associated with one another across all three visits in participants with PD. These findings suggest that both orthographic transcription and VAS ratings may be valid ways of assessing sentence intelligibility for IWPD. Additionally, it is possible that orthographic transcription and VAS ratings of the SIT measure and capture similar aspects of speech intelligibility in IWPD.

4.6.3 Communicative participation

The correlations involving the CPIB revealed a moderate association with self-rated VAPP *self-perceived voice problem* scores, and a moderate to good association with self-rated VAPP *total* scores. These results suggest that as IWPD perceive a more negative impact of their voice with respect to their day-to-day functioning, the lower they will rate their overall communicative participation. While no published studies have examined the relationship between the CPIB and the VAPP, two studies have explored the relationship between the CPIB other functional questionnaires (Baylor et al., 2009; McAuliffe et al., 2016).

Baylor and colleagues (2009) conducted psychometric analyses of the CPIB on 208 participants with spasmodic dysphonia. Participants completed the 141-item CPIB and the Voice Handicap Index (VHI). The authors then conducted Rasch analysis in order to identify the most appropriate items to include in the CPIB short form and correlated the CPIB and VHI. Baylor and colleagues (2009) found that the CPIB was moderately correlated with the VHI total score. Additionally, McAuliffe and colleagues (2016) explored the relationship between health-related quality of life measures and

communicative participation as measured via the CPIB in 378 IWPD in the United States and New Zealand. The authors reported that the CPIB was moderately correlated with the *Physical, Mental,* and *Social Roles* subsections of the Patient Reported Outcomes Measurement Information System (PROMIS; McAuliffe et al., 2016). The CPIB was also found to be strongly correlated with the Parkinson's Disease Questionnaire – 8 (PDQ-8; McAuliffe et al., 2016). While relationships between the CPIB and subsections of the PROMIS, PDQ-8; and VHI have been explored (Baylor et al., 2009; McAuliffe et al., 2016), additional research is required to further examine any relationships between the CPIB and other measures of communicative participation, such as the VAPP.

The correlations amongst self-rated CES questions revealed weak associations between almost all of the individual CES question scores for participants with PD. A moderate relationship was found between CES question 1 (*Having a conversation with a family member or friends at home*) and CES question 2 (*Participating in conversation with strangers in a quiet place*). These communication contexts are both relatively unchallenging. Listeners who are familiar with dysarthric speech, such as close friends, family members, and spouses, have been found to better recognize and understand the speech of individuals with various types of dysarthria, including hypokinetic dysarthria (Depaul & Kent, 2000; Liss, Spitzer, Cavinesss, & Adler, 2002; Tjaden & Liss, 1995). Thus, it is plausible that IWPD and dysarthric speech might perceive their communicative effectiveness to be greater in situations wherein they are communicating with listeners who are familiar with them and their speech difficulties.

Additionally, the presence of background noise has been found to have adverse effect on the conversational speech intelligibility of IWPD and hypophonia (Adams et al., 2008; Dykstra et al., 2012b). Adams and colleagues (2008) examined the conversational speech intelligibility of 25 IWPD and hypophonia in the presence of various background noise conditions, including: no added background noise, 60, 65, and 70 dB SPL. The authors reported that conversational speech intelligibility scores in noise conditions were on average 5 – 10% lower than conversational intelligibility scores assessed in a quiet environment (Adams et al., 2008). Furthermore, Dykstra and colleagues (2012a) also sought to assess conversational speech intelligibility in background noise in 30

individuals with hypophonia and PD. All participants engaged in conversational speech tasks in different background noise conditions, including: no added background noise, 60, 65, and 70 dB SPL. Dykstra and colleagues (2012a) reported that the conversational speech intelligibility of IWPD decreased as the intensity of background noise increased, with intelligibility values of 89.63%, 77.47%, 68.98%, and 57.57% in no noise, 60 dB SPL, 65 dB SPL, and 70 dB SPL, respectively. Based on these findings, it was demonstrated that the conversational speech intelligibility of IWPD became further reduced with increasing levels of background noise (Adams et al., 2008; Dykstra et al., 2012a). Since IWPD are more become less intelligible when speaking in background noise, it is possible that IWPD might perceive their communicative effectiveness to be greater in communicative situations and contexts that are quieter and involve familiar conversational partners. It is possible that the less challenging nature of both of these communicative situations may provide some explanation as to why CES question 1 (*Having a conversation with a family member or friends at home*) and CES question 2 (*Participating in conversation with strangers in a quiet place*) were significantly correlated with one another across all three visits for participants with PD.

The correlations amongst self-rated VAPP subsection scores revealed good to excellent associations among all of the individual VAPP subsection scores for participants with PD. These findings suggest that IWPD are extremely consistent in their perception of the impact of their voice difficulties on their activity and participation over time. These results appear to be consistent with previous work exploring the test-retest reliability of the VAPP (Ma & Yiu, 2001). As part of its development, the VAPP was administered to 25% of its original testing population two weeks following participants' initial completion of the questionnaire (Ma & Yiu, 2001). The authors reported good test-retest reliability between the two administrations of the VAPP (Ma & Yiu, 2001). The VAPP was originally tested and validated for use on individuals with dysphonia (Ma & Yiu, 2001) and was later validated on IWPD (Simberg et al., 2012). However, it is important to note that the present study revealed inconsistent results for the retest reliability and repeatability analyses for the individual VAPP subsection scores. Therefore, caution is recommended in the interpretation of the perceived consistency of VAPP subsection

scores. Additional research is recommended to further examine the test-retest reliability of the VAPP in IWPD and hypophonia.

Finally, level of speech usage was not found to be related to other measures of communicative participation in participants with PD. These results contradict those reported by McAuliffe and colleagues (2016), who found that increased levels of speech usage were associated with increased communicative participation as measured by the CPIB in IWPD. It is possible that this difference is related to the degree of speech difficulties experienced by the IWPD. In the present study, participants with PD were selected due to the presence of hypophonia. However, McAuliffe and colleagues (2016) did not exclude participants on the basis of their presenting speech characteristics. It is possible that the IWPD who took part in McAuliffe and colleagues' (2016) study demonstrated a greater variety of speech features related to hypokinetic dysarthria compared to the participants with PD in the present study who exhibited hypophonia as their primary dysarthric feature.

4.6.4 Demographic and non-speech factors

The correlations involving demographic and non-speech factors revealed a moderate correlation between gender and MOCA scores for participants with PD. These findings suggest the presence of sex differences in cognitive abilities in IWPD. In the present study, the MOCA scores of female participants with PD were on average 3.48 points higher than those of male participants with PD. Both male and female participants with PD had been diagnosed with PD for a similar duration of time and presented with similar UPDRS scores, approximately 10 years and 36 points. Several published studies have explored sex differences on various measures of speech intensity and communicative participation in IWPD (Fox & Ramig, 1997; Sapir et al., 2007; Yorkston et al., 2008).

Sapir and colleagues (2007) explored the impact of LSVT on vowel intensity in 29 IWPD. Sapir and colleagues (2007) reported significant gender differences in intensity measures of the vowels, wherein men had higher vocal sound pressure level compared to women, and women demonstrated higher second formant values for /u/ as well as the ratio of second formants for /i/ and /u/ prior to receiving treatment. While the present

study did not specifically explore vowel intensity or formants, speech intensity was not found to be significantly correlated with gender over time.

Fox and Ramig (1997) examined the impact of LSVT on speech intensity and nine self-rated perceptual variables in 30 IWPD and 14 healthy controls. No gender differences were found in the speech intensity across all speech tasks (Fox & Ramig, 1997). These findings are consistent with those in the present study, as measures of speech intensity were not found to be significantly correlated with gender over time. Additionally, the work of Fox and Ramig (1997) did not reveal gender differences in the nine self-rated perceptual variables. These variables included self-perceived loudness, understandability, participation, and various voice quality variables (Fox & Ramig (1997). Similarly, results from the current study did not reveal significant correlations between gender and self-perceived typical speech loudness or self-rated communicative participation. However, it is important to note that self-rated perceptual measures differed significantly between the two studies. Fox and Ramig (1997) asked participants to rate each variable using a VAS. Whereas the currently study used a similar method for perceptions of typical speech loudness, but validated and standardized questionnaires for measures of self-rate communicative participation.

Yorkston and colleagues (2008) explored the relationships between self-rated participation and personal factors in 112 individuals with multiple sclerosis. Gender effects were observed, with men with multiple sclerosis demonstrating increased communicative participation compared to women with multiple sclerosis (Yorkston et al., 2008). While Yorkston and colleagues (2008) explored relationships between variables in individuals presenting with a different movement disorder. It is interesting to note that gender was not significantly correlated with measures of communicative participation over time in the present study. It is possible that the different presentation of dysarthric features present in PD and multiple sclerosis may influence participation differently.

Several studies have reported inconsistent results pertaining to MOCA scores and sex differences in healthy adults, with multiple studies demonstrating no effect of gender (Conti, Bonazzi, Laiacona, Masina, & Coralli, 2015; Kopecek, Stepankova, Lukavsky,

Ripova, Nikolai, & Bezdicek, 2017; Santangelo et al., 2015; Siciliano, Chiorri, Passaniti, Sant'Elia, Trojano, & Santangelo, 2019; Zheng, Teng, Varma, Mack, Mungas, Lu, & Chui, 2012), others showing females demonstrate lower MOCA scores (Liu, Luo, Tang, & Wong, 2018; Shaik et al., 2016), while some studies report males demonstrate lower MOCA scores (Pavlovic, Abel, Barlow, Farrell, Weiner, & DeFina, 2018; Thomann et al., 2018), and, finally, one study revealing gender differences on the attention and delayed recall subsections of the MOCA (Ojeda, del Pino, Ibarretxe-Bilbao, Schretlen, & Peña, 2016). Additionally, prior to undergoing bariatric surgery, older men demonstrated greater MOCA scores compared to older women (Mohun, Spitznagel, Gunstad, Rochette, & Heinberg, 2018). A study by Niwald, Redlicka, and Miller (2018) revealed no gender differences in the cognitive status, measured via the MOCA, in individuals with multiple sclerosis. To date, no published studies have explored whether there exists a relationship between gender differences and cognitive abilities in IWPD and hypophonia. Additional research is recommended to develop a greater understanding of the effect of gender on cognition in IWPD suggested by the results of the present study.

4.7 Objective 8: Inter-Relationships Among Proxy-Measures in Participants with Parkinson's Disease

This objective examined inter-relationships among proxy-rated speech loudness and communicative participation measures for participants with PD.

4.7.1 Proxy-rated speech loudness measures

The correlations revealed that proxy-rated typical speech loudness was significantly correlated with all proxy-rated VAPP subsection scores for participants with PD – with the exception of VAPP *participation restriction* scores. These findings suggest that a lower proxy-rating of typical speech loudness of the IWPD is significantly and moderately associated with higher proxy-rated VAPP *voice problem, daily communication, social communication, total* and *activity limitation* scores. While few studies have explored communicative participation in IWPD (Donovan 2008; Dykstra et al., 2015), there are no published studies that have explored proxy-rated speech loudness in IWPD. While a previous study by McAuliffe and others (2016) also found communicative participation to be significantly associated with perceived speech severity

(standardized beta coefficient = 0.28), it is important to note that the methodology used by McAuliffe and colleagues (2016) was quite different than that of the present study. McAuliffe and others (2016) explored whether a variety of variables was associated with communicative participation as measured by the CPIB in 378 IWPD in the United States and New Zealand. However, self-rated speech severity and communicative participation were tested at a single time point (McAuliffe et al., 2016). Secondly, proxy-measures were not taken into account as the CPIB was used as a measure of self-rated communicative participation (McAuliffe et al., 2016). Finally, self-perceived speech severity was rated on the basis of "understandability" in the study by McAuliffe and colleagues (2016). While the results of these two studies cannot be directly compared, it is interesting to note that similar relationships were observed in measures of communicative participation and perceived speech difficulties.

4.7.2 Speech intelligibility measures

The correlations involving speech intelligibility measures and any proxy-rated speech loudness and proxy-rated communicative participation measures yielded no consistent patterns over the three visits. No published studies have investigated the relationship among measures of speech intelligibility and measures of proxy-rated speech loudness and proxy-rated communicative participation. However, the results from the present study suggest that proxy-rated measures of speech loudness and proxy-rated measures of communicative participation are not associated with measures of speech intelligibility in participants with PD. Thus, while measures of proxy-rated speech loudness and proxy-rated communicative participation may provide a broader perspective an and individual's experience with PD, ratings made by primary communication partners do not appear to be useful in assessing the functional outcomes of impaired activities, such as speech intelligibility (Baylor et al., 2009).

4.7.3 Proxy-rated communicative participation

The correlations involving proxy-rated CPIB scores revealed a moderate association with proxy-rated CES question 4 (*Conversing with a stranger over the telephone*) scores, proxy-rated CES question 5 (*Being part of a conversation in a noisy environment (social gathering)*) scores, and proxy-rated CES question 8 (*Having a conversation with*

someone at a distance (across a room)) scores across all three visits for participants with PD. These findings suggest that proxy-measures of communicative participation, as measured by proxy-rated CPIB scores, is moderately associated with proxy-rated items on the CES related to effectiveness while conversing with an unfamiliar speaker over the telephone, speaking in noise, and communicating across a distance. It may be possible that these communicative contexts are perceived by the primary communication partner of the IWPD to be particularly challenging. Such a conclusion would be consistent with a previous study by Dykstra and colleagues (2015), who compared self- and proxy-ratings of the CETI-M in 30 IWPD and hypophonia. The authors reported that primary communication partners perceived communicating in noise and over a distance to be the most challenging communicative contexts for their partners with PD.

Additionally, correlations involving proxy-rated CPIB revealed a moderate to good association with proxy-rated VAPP *self-perceived voice problem* scores, and a moderate to excellent association with proxy-rated VAPP *total* scores. These results suggest that as primary communication partners of IWPD perceive a more negative impact of their partner's voice on their day-to-day functioning, the lower they will rate their partner's overall communicative participation. Simberg and colleagues (2012) evaluated a 15-day treatment on the speech and voice of 6 IWPD. Prior to beginning treatment, IWPD and their spouses completed the VAPP in order to obtain self- and proxy-ratings of voice function. Six months and one year post-treatment onset, self- and proxy-ratings of voice function were evaluated using the VAPP. The authors reported that proxy-ratings of voice functioning were less severe than the self-ratings of IWPD. Simberg and colleagues (2012) also found that the spouses of all participants with perceived positive changes in the speech and voice of their partner with PD following the treatment. The authors suggested that the proxy-rated VAPP provided valuable insight to the functional perspective of individuals with communication disorders. As mentioned above, the relationships between the CPIB and other functional patient reported outcome measures, including the PROMIS and VHI, have been explored in previous studies (Baylor et al., 2009; McAuliffe et al., 2016), additional research is required to further examine contrasts

and similarities between the CPIB and other measures of communicative participation, such as the CES and VAPP.

The correlations amongst proxy-rated CES questions revealed moderate to good associations among many of the individual CES question scores. These findings suggest that the primary communication partners of IWPD perceive their partner's communicative effectiveness to be consistent across different communicative situations over time relative to one another. The current study found the individual questions of the CES to demonstrate poor retest reliability and unacceptable repeatability for participants with PD, their primary communication partners, and control participants. While proxy-rated CES question scores may have varied over time, the differences between scores may have remained relatively constant over time. It is possible that many of the scores demonstrated a similar pattern of temporal variability across all three visits. A more thorough investigation of the variability and degrees of change in proxy-rated CES scores over time is recommended for subsequent research.

The correlations amongst proxy-rated VAPP subsection scores revealed moderate to excellent associations between all of the individual VAPP subsection scores. These findings suggest that the primary communication partners of IWPD may be extremely consistent in their perception of the impact of their partner's voice difficulties on their activity and participation over time. It is important to note that retest reliability and repeatability analyses for the individual VAPP subsection scores yielded inconsistent results in the present study. The correlations performed in the present analyses and in the retest reliability analysis provide information with regard to the strength of the relationship under examination. These correlations do not incorporate measurement error. Since the retest analyses yielded inconsistent results, additional research is recommended to explore the test-retest reliability of the VAPP in IWPD and hypophonia to explore the influence of measurement error.

4.7.4 Demographic and non-speech factors

The correlations involving demographic non-speech factors, and any proxy-ratings yielded no consistent patterns over the three visits for participants with PD. These

findings suggest that the demographic and non-speech factors under examination in this study are not related to proxy-measures of speech loudness or proxy-measures of communicative participation related to participants with PD.

4.8 Objective 9: Inter-Relationships Among Variables in Control Participants

This objective examined inter-relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for control participants.

4.8.1 Speech intensity measures

The correlations involving speech intensity measures revealed a moderate to good association between maximum speech intensity and habitual speech intensity, as well as maximum speech intensity and magnitude production in control participants across all three visits. These results suggest that the greater an individual's habitual and maximum speech intensity, the steeper their function of magnitude production.

4.8.2 Speech intelligibility measures

The correlations involving speech intelligibility measures yielded no consistent patterns over the three visits for control participants. These findings suggest that the measures of speech intelligibility under examination in this study are not related to measures of speech intensity, measures of communicative participation, demographic and non-speech factors, or other measures of speech intelligibility in control participants. It is unclear whether the different measures of speech intelligibility are not associated with one another in control participants due to significant variability in control participants or if the presence of unknown confounding variables were present. Additionally, it may be that the small range of speech intelligibility scores observed in control participants may have contributed to the lack of consistent correlations between measures of speech intelligibility and other variables under examination in the present study.

4.8.3 Communicative participation

The correlations involving self-rated CPIB revealed a moderate association with CES question 8 (*Having a conversation with someone at a distance (across a room)*) across all

three visits for control participants. These findings suggest that communicative participation as measured by the CPIB is moderately associated with perceived communicative effectiveness while conversing at a distance. As previously discussed, the relationships between the CPIB and other functional questionnaires, including the PROMIS and VHI, in IWPD have been explored by Baylor and colleagues (2009) and McAuliffe and colleagues (2016). However, no published studies have explored the relationships between measures of communicative participation in control participants. Additional research is recommended to further examine associations between measures of communicative participation, including the CES, CPIB, and VAPP.

The correlations amongst self-rated CES questions revealed moderate to excellent associations between many of the individual CES question scores for control participants. Similar to the results of correlations of proxy-rated CES questions scores for participants with PD, these findings suggest that control participants perceive their communicative effectiveness to be fairly consistent across different communicative situations relative to one another over time. The individual questions of the CES were found to demonstrate poor retest reliability and unacceptable repeatability for all participants. It is possible that self-rated CES question scores for control participant may have varied over time, while differences between scores may have remained relatively constant. An increase in one score across time points may have been consistent across other scores, resulting in CES question scores that were significantly correlated with one another across visits. Additional research examining the variability of individual CES scores over time is recommended.

The correlations amongst self-rated VAPP subsection scores revealed moderate to excellent associations between many of the individual VAPP subsection scores. These findings suggest that control participants may be highly consistent and relatively uniform in their perception of the impact of their voice on their activity and participation over time. However, these results must be interpreted with caution as retest reliability and repeatability analyses for the individual VAPP subsection scores yielded inconsistent results in the current study. Thus, further exploration of the test-retest reliability of the VAPP is recommended.

4.8.4 Demographic and non-speech factors

The correlations involving demographic and non-speech factors yielded no consistent patterns over the three visits for control participants. These findings suggest that the demographic and non-speech factors under examination in this study are not related to measures of speech intensity, speech intelligibility, communicative participation, or other demographic and non-speech factor in control participants.

4.9 Study Limitations

While this exploratory study yielded many interesting results, it is important to note that several limitations were present. The first limitation of this study is related to participant characteristics. The IWPD who took part in the present study presented with mild to moderate hypophonia and as a result, did not have dysarthria in the severe range. Furthermore, the current study included only those IWPD for whom hypophonia was their primary dysarthric feature. However, IWPD and hypokinetic dysarthria may present with a wide variety of distinctive speech features. As a result, the findings from the current study may not be generalizable to a wider variety of IWPD who may be experiencing more severe hypophonia or other features of hypokinetic dysarthria (i.e., imprecise consonant production or prosodic disturbances). Further study of the variability of speech intensity, speech intelligibility, and communicative participation in IWPD with a wide range of severity and speech symptoms is required in order to further understand and explore the impact of hypokinetic dysarthria secondary to PD.

It would also be prudent to recruit an equal number of male and female speakers with PD and hypophonia. Such participant proportions may lend support the presence of sex differences in the variability of and relationships among measures of speech intensity, speech intelligibility and communicative participation.

A second limitation of the study is related to the study's overall design. Since participants were required to attend three experimental sessions and perform the same speech tasks at each visit, it is possible that performance effects occurred. The occurrence of performance effects may account for increased speech intensity during speech tasks at visits 2 and 3 as compared to visit 1. Furthermore, while precautions were taken in order

to deter the possibility of recall, such as the randomization of the order of questionnaires across participants and visits, participants were required to complete five questionnaires during each visit. It may be possible that completing these questionnaires primed our participants to pay greater attention to their speech loudness and communicative participation between visits. This greater focus on their speech loudness and communicative participation may account for some of the changes observed over time in self-rated communicative participation. Additional investigation into these changes via qualitative interviews is recommended as qualitative interviews would allow researchers to further explore and develop a deeper understanding of how IWPD perceived changes to their communicative participation.

Another potential limitation of the present study is related to the sample of listeners. Despite a small sample size of three listeners, reliability between and within listeners was moderate to excellent for the majority of the speech intelligibility ratings. Similarly, multiple studies exploring sentence and conversational intelligibility in IWPD have recruited small samples of listeners (sample sizes ranging from 2 – 5 listeners) and reported good interrater and intra-rater reliability (e.g., Abur et al., 2018; Adams et al., 2008; Dromey, 2003; Dykstra et al., 2015). Furthermore, Abur and colleagues (2018) compared the relationship between VAS and transcription scores of speech intelligibility between three listeners to 66 listeners. The authors reported a strong relationship ($R^2 >$ .82) between VAS and transcription intelligibility scores by three listeners. Abur and colleagues (2018) also suggested that the strength of the relationship between VAS and transcription estimates of speech intelligibility do not increase with additional listeners. Additionally, the listeners in the current study were speech language pathology students that were familiar listening to dysarthric speech. Two published studies have found that speech language pathologists are more likely to provide higher ratings of speech intelligibility compared to naïve listeners (Dagenais, Garcia, & Watts, 1998; Dagenais, Watts, Turnage, & Kennedy, 1999). However, a more recent study has suggested that various factors, including familiarity, may not significantly impact ratings of speech intelligibility when rated under standardized listening conditions (Pennington & Miller, 2007). One final possible limitation of this study is that listeners may have experienced

fatigue as a result of listening to and rating speech samples. While listeners did take part in three separate listening sessions, each session was approximately two hours long. It is possible that listeners became fatigued from orthographic transcription and VAS rating tasks, which may have negatively impacted their transcriptions and ratings.

4.10 Future Directions

The results from the current study suggest the need for additional studies. Possible avenues for future directions include replication of the study while including some modifications to the methodology, and further exploration of the main findings.

As previously mentioned, it is recommended that the participant population be widened to include IWPD who present with severe hypophonia, as well as other key features associated with hypokinetic dysarthria such as imprecise articulation and short rushes of speech. It is possible that different speech symptoms associated with PD may influence measures such as speech intelligibility and communicative participation differently. The use of such a methodology may yield results that are more generalizable to the overall population of IWPD. Furthermore, including IWPD with lower cognitive scores as well as including equal numbers of male and female participants might further assist with the generalizability of the results.

As previously mentioned, it is possible that following the completion of each visit, participants may have become more attuned to their speech loudness and day-to-day communicative participation. This increased focus may have resulted in changes to their performance throughout the study. If this study were to be replicated with the addition of qualitative interviews further exploring participant views on their speech loudness and communicative participation, a more developed rationale for some of the observed changes could be identified. Finally, it is also possible that self-rated communicative participation questionnaires which use equal appearing intervals, such as the CES, may not capture all degrees of change. It is possible that measures which make use of a VAS, may demonstrate greater sensitivity to changes over time. Additional research may be warranted in order to explore whether modification of these PROs to use VAS may result in an increased sensitivity to the detection of changes over time. Such research would

necessitate validation studies exploring different scaling procedures in order to identify which one(s) would be most effective in identifying significant changes over time in questionnaires such as the CES and CPIB.

4.11 Clinical Implications

Several clinical implications may be drawn from the results of the current study regarding the assessment and management of individuals with hypophonia and PD. The consistency in self- and proxy-ratings in the present study suggest that both IWPD and their primary communication partner provide relatively consistent ratings of perceived speech loudness and communicative participation. Proxy-measures of perceived speech loudness and communicative participation as measured in the current study could be used in the event that the individual with PD is not able to provide this information, however, it should be underscored that it is best clinical practice to gather this information from the actual person versus his/her communication partner.

Furthermore, clinicians may wish to consider, when evaluating treatment outcomes, that measures of speech intensity may be variable over time. Clinicians should be cognizant that a certain amount of temporal variability and possible measurement error may be present in some measures of speech intensity such as habitual speech intensity, Lombard response function, magnitude production, and perceived typical speech loudness; speech intelligibility (VAS conversational intelligibility scores), and communicative participation (CES, VAPP, and LSUS). However, maximum speech intensity, speech intelligibility, measured by SIT transcription and SIT VAS scores, and the CPIB appear to demonstrate less temporal variability and good reliability and repeatability.

The CR% cutoff value of 10% used in the present study to denote measures with good repeatability and measures that would reflect a true change was adopted from the empirical literature outside of PD (Lu et al., 2007; Smidt, et al., 2002; Steffen & Seney, 2008). However, the CR% value of 10% adopted by the present study may be too liberal for certain outcome measures, such as habitual speech intensity and maximum speech intensity. The present study suggests that CR% values slightly greater than 5% for habitual speech intensity and less than 9% for maximum speech intensity would be

indicative of a clinically meaningful change. Additionally, the present study appears to support the use of SIT transcription and VAS scores, VAS conversational intelligibility scores, and CPIB scores as treatment outcome measures in order to detect clinically meaningful differences. Whereas measures, such as the Lombard response function, magnitude production, perceived speech loudness, individual CES question scores, VAPP subsection scores, and LSUS scores, may need to be refined or alternative measures designed that are associated with lower measurement error. These proposed changes to the aforementioned measures may be useful in assessing meaningful clinical differences in order the measure to be effective treatment outcomes. Continued research in the area of variability, reliability, and repeatability in speech intensity and communicative participation measures in IWPD and hypophonia may better inform clinical practice and in time result in additional effective and efficient options for assessing an individual over the course of their treatment.

4.12 Summary and Conclusion

The current study explored the variability of acoustic and perceptual measures of speech intensity, speech intelligibility, and communicative participation measures in IWPD and hypophonia. This study also examined the relationships among speech intensity measures, speech intelligibility measures, self- and proxy ratings of communicative participation, demographic factors, and non-speech factors.

The first objective addressed the variability of speech intensity measures for participants with PD and control participants. The results revealed a significant reduction in speech intensity for participants with PD compared to control participants at visit 1, but not at visits 2 and 3. These findings suggest that there is significant variability in habitual speech intensity over time. Significant differences were also found in maximum speech intensity for participants with PD and control participants. These results suggest that IWPD demonstrate a narrower range of speech intensity compared to control participants. Furthermore, these differences in measures of maximum speech intensity were consistent over time. Similar Lombard response slopes between participants with PD and control participants were also reported in the current study. These findings suggest that participants with PD may demonstrate a parallel but attenuated Lombard response

function compared to control participants. The results also revealed a significant difference in Lombard response function for participants with PD and to control participants at visit 1, but not at visits 2 and 3. Finally, participants with PD demonstrated significantly smaller magnitude production slope compared to control participants. These results suggest that IWPD demonstrate a less steep magnitude production function compared to control participants, and thus exhibit a narrower range of speech intensity. The results also revealed a significant difference in magnitude production for participants with PD compared to control participants at visit 1, but not at visits 2 and 3. The variability reported in three of the four speech intensity measures may be related to an increased awareness and focus on overall speech intensity and great subsequent speech effort for participants with PD. These results lend support to the hypothesis that IWPD would demonstrate greater variability of speech intensity measures compared to control participants.

The second objective of this study addressed the variability of perceived typical speech loudness for participants with PD, their primary communication partners, and control participants. The results revealed similar ratings by participants with PD and their primary communication partners. These perceptions were significantly lower than the self-perceived typical speech loudness of control participants. These findings suggest that participants with PD and their primary communication partners perceived the speech loudness of IWPD to be flatter compared to control participants. While these findings support the hypothesis that self-rated typical speech loudness would be reduced for participants with PD compared to control participants, it does not support the prediction that self- and proxy-rated typical speech loudness would differ for participants with PD. Additionally, it was unclear whether there was variability in the perceived typical speech loudness over time.

The third objective of this study addressed the variability speech intelligibility measures for participants with PD and control participants. This study found a significant reduction in all measures of speech intelligibility for participants with PD compared to control participants. A fairly large difference was observed in the SIT transcription and VAS scores for participants with PD. These differences are consistent with findings from

previous empirical studies (Abur, Enos, & Stepp, 2018; Adams et al., 2008; Stipancic et al., 2016). Orthographic transcription captures a listener's accuracy of their transcription of each word. However, other aspects of the speech signal may be captured during VAS estimation, such as prosodic disturbances and perceived acceptability (Kent et al., 1999; Kent et al., 1990; De Bodt et al., 2002). These differences may explain the discrepancy between SIT transcription and VAS scores. Furthermore, SIT transcription and VAS scores were both found to be fairly consistent over time for both participants with PD and control participants. These findings do not support the prediction that participants with PD would demonstrate greater variability in speech intelligibility measures compared to control participants. Finally, it was unclear whether there was variability in conversational intelligibility scores over time.

The fourth objective of this study addressed the variability self-rated communicative participation measures for participants with PD and control participants. The results revealed a significant difference between participants with PD and control participants across all self-rated measures of communicative participation. Participants with PD exhibited lower scores on the CES, CPIB and LSUS, and greater scores on the VAPP compared to control participants. These findings suggest that participants with PD may be aware of their level of speech usage and their speech and communication difficulties. Changes over time were observed for six of the eight CES questions, as well as the CPIB and the VAPP *social communication* and *participation restriction* subsections. The findings of observed changes in the CES and VAPP support the hypothesis that participants with PD would demonstrate greater variability in measures of communicative participation over time. However, it was unclear whether there was significant variability in VAPP subsection scores and LSUS scores.

The fifth objective of this study addressed the variability of self- and proxy-rated communicative participation measures of participants with PD and their primary communication partners. This study revealed comparable self- and proxy-ratings across all measures of communicative participation for participants with PD. These results suggest that participants with PD demonstrate an awareness of their communication

difficulties. These findings do not support the prediction that self- and proxy-rated communicative participation measures would differ for participants with PD.

The sixth objective of this study addressed the retest reliability and repeatability of measures of speech intensity, speech intelligibility, and communicative participation in participants with PD, their primary communication partners, and control participants. Habitual speech intensity was not found to demonstrate good retest reliability but did demonstrate fairly good repeatability for participants with PD and control participants. Maximum speech intensity was found to demonstrate good retest reliability and good repeatability for both participants with PD and control participants. SIT VAS scores were found to demonstrate good retest reliability and marginal repeatability for participants with PD and good retest reliability and good repeatability for control participants. Conversational intelligibility scores demonstrated good retest reliability and good repeatability for participants with PD and good repeatability but not good retest reliability for control participants. Measures of communicative participation typically did not demonstrate good retest reliability and unacceptable repeatability in participants with PD, their primary communication partners, and control participants. However, the CPIB demonstrate good retest reliability and marginal repeatability for participants with PD and their primary communication partners, and good retest reliability and good repeatability. These results suggest the possibility that select measures of speech intensity, speech intelligibility, and communicative participation demonstrate acceptable levels of measurement error and may be used to observe clinically meaningful change.

The seventh objective of this study addressed the relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD. Maximum speech intensity and magnitude production were consistently correlated with one another, suggesting that the greater an individual's maximum speech intensity, the greater their intensity range. Sentence intelligibility scores were consistently associated with one another. Additionally, the CPIB was consistently correlated with the VAPP *self-perceived voice problem* and *total* subsection scores, and the VAPP subsections were

consistently correlated with one another. These results suggest good internal consistency among these different measures of speech intelligibility and communicative participation. Finally, the findings did not support the prediction that disease duration, medication effectiveness, and cognition would be correlated with measures of speech intensity, speech intelligibility, and self-rated communicative participation in IWPD.

The eighth objective of this study addressed the relationships among speech intensity measures, speech intelligibility measures, proxy-rated communicative participation measures, demographic factors, and non-speech factors for participants with PD. Perceived typical speech loudness was associated with all but one VAPP subsection, suggesting that a greater perceived reduction in an IWPD's loudness is associated with increased perceived impairment in the day-to-day functioning of their activity and participation. The CPIB was consistently correlated with select CES questions and VAPP subsections. These results suggest good internal consistency among many proxy-measures of speech intensity and communicative participation. Finally, the findings did not support the prediction that disease duration, medication effectiveness, and cognition would be correlated with measures of proxy-rated communicative participation in IWPD.

Finally, the ninth objective of this study addressed the relationships among speech intensity measures, speech intelligibility measures, self-rated communicative participation measures, demographic factors, and non-speech factors for control participants. Maximum speech intensity was consistently correlated with habitual speech intensity and magnitude production. These results suggest that most measures of speech intensity are related to one another in control participants. The CPIB was only consistently associated with CES question 8 (*Having a conversation with someone at a distance (across a room)*). These results suggest that the CPIB, CES, VAPP, and LSUS may be measuring different components of communicative participation in control participants.

This study has contributed to the knowledge of the variability of speech intensity measures, speech intelligibility measures, and communicative participation measures in IWPD and hypophonia. The findings from the current study will contribute to the

growing understanding of hypophonia in PD. This study may also provide valuable information with regard to more comprehensive assessments and improved interpretation of treatment outcomes of IWPD in clinical practice.

References

Abur, D., Enos, N.M., & Stepp, C.E. (2019). Visual analogue scale ratings and orthographic transcription measures of sentence intelligibility in Parkinson's disease with variable listener exposure. *American Journal of Speech Language Pathology, 11,* 1 – 11.

Adams, S.G., & Dykstra, A. (2009). Hypokinetic dysarthria. In McNeil, M.R. (Ed.), *Clinical Management of Sensorimotor Speech Disorders, 2nd Edition*. New York, NY: Thieme Medical Publishers, Inc.

Adams, S., Dykstra, A., Abrams, K., Winnell, J., Jenkins, M., & Jog, M. (2006a). Conversational speech intensity under different noise conditions in hypophonia and Parkinson's disease. *Canadian Acoustics, 34*(3), 96 – 97.

Adams, S.G., Dykstra, A., Jenkins, M., & Jog, M. (2008). Speech-to-noise levels and conversational intelligibility in hypophonia and Parkinson's disease. *Journal of Medical Speech-Language Pathology, 16*(4), 165 – 172.

Adams, S., Haralabous, O., Dykstra, A., Abrams, K., & Jog, M. (2005). Effects of multi-talker background noise on the intensity of spoken sentences in Parkinson's disease. *Canadian Acoustics, 33*(3), 94 – 95.

Adams, S.G., & Jog, M. (2009). Parkinson's disease. In McNeil, M.R. (Ed.), *Clinical Management of Sensorimotor Speech Disorders, 2nd Edition*. New York, NY: Thieme Medical Publishers, Inc.

Adams, S.G., & Lang, A.E. (1992). Can the Lombard effect be used to improve low voice intensity in Parkinson's disease? *European Journal of Disorders of Communication, 27,* 121 – 127.

Adams, S., Moon, B.-H., Dykstra, A., Abrams, K., Jenkins, M., & Jog, M. (2006b). Effects of multitalker noise on conversational speech intensity in Parkinson's disease. *Journal of Medical Speech-Language Pathology, 14*(4), 221 – 228.

Adams, S.G., Winnell, J., & Jog, M. (2010). Effects of interlocutor distance, multi-talker background noise, and a concurrent manual task on speech intensity in Parkinson's disease. *Journal of Medical Speech-Language Pathology, 18*(4), 1 – 8.

Andreetta, M.D., Adams, S.G., Dykstra, A.D., & Jog, M. (2016). Evaluation of speech amplification devices in Parkinson's disease. *American Journal of Speech Language Pathology, 25*(1), 29 – 45.

Arnold, C., Gehrig, J., Gispert, S., Seifried, C., & Kell, C.A. (2014). Pathomechanisms and compensatory efforts related to Parkinsonian speech. *NeuroImage: Clinical, 4,* 82 – 97.

Astolfi, A., Bottalico, P., Accornero, A., Garzaro, M., Nadalin, J., Giordano, C. (2012). *Proceedings – European Conference on Noise Control,* 55 – 60.

Astolfi, A., Puglisi, G.E., Pavese, L., & Carullo, A. (2014). Long-term vocal parameters of primary school teachers and classroom acoustics with and without an acoustical treatment. *Proceedings of Forum Acusticum, 2014,* code 117181.

Ball, L.J., Beukelman, D.R., & Pattee, G.L. (2004). Communicative effectiveness of individuals with amyotrophic lateral sclerosis. *Journal of Communication Disorders, 37*(3), 197 – 215.

Bassi, I.B., Assunção, A.A., De Medeiros, A.M., De Menezes, L.N., Teixeira, L.C., & Côrtes Gama, A.C. (2011). Quality of life, self-perceived dysphonia, and diagnosed dysphonia through clinical tests in teachers. *Journal of Voice, 25*(2), 192 – 201.

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., McAuliffe, M.J., Hughes, L.E., Yorkston, K., Anderson, T., Kim, J., & Amtmanna, D. (2014). A differential item functioning (DIF) analysis of the communicative participation item bank (CPIB): Comparing individuals with Parkinson's disease from the United States and New Zealand. *Journal of Speech, Language, and Hearing Research, 57*(1), 90 – 95.

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, 56,* 1190-1208.

Baylor, C., Yorkston, K.M., Eadie, T., Miller, R., & Amtmann, D. (2008). Levels of speech usage: A self-report scale for describing how people use speech. *Journal of Medical Speech-Language Pathology, 16*(4), 191 – 198.

Baylor, C.R., Yorkston, K.M., Eadie, T.L., Miller, R.M., & Amtmann, D. (2009). Developing the communicative participation item bank: Rasch analysis results from a spasmodic dysphonia sample. *Journal of Speech, Language, and Hearing Research, 52*(5), 1302 – 1320.

Beckerman, H. Roebroeck, M.E., Lankhorst, G.J., Becher, J.G., Bezermer, P.D., & Verbeek, A.L.M. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research, 10*(7), 571 – 578.

Behlau, M., Oliveira, G., dos Santos, L.M.A., & Ricarte, A. (2009). Validation in Brazil of self-assessment protocols for dysphonia impact. *Pro-Fono, 21*(4), 326 – 332.

Berke, G.S., Gerratt, B., Kreiman, J., & Jackson, K. (1999). Treatment of Parkinson hypophonia with percutaneous collagen augmentation. *The Laryngoscope, 109,* 1295-1299.

Boersma, P., & Weenink, D. (2018). Praat, software for speech analysis and synthesis.

Bottalico, P., & Astolfi, A. (2012). Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *Journal of the Acoustical Society of America, 131*(4), 2817 – 2827.

Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., & Del Tredici, K. (2004). Stages in the development of Parkinson's disease-related pathology. *Cell and Tissue Research, 318,* 121 – 134.

Brumm, H., & Zollinger, S.A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour, 148,* 1173 – 1198.

Cannito, M.P., Suiter, D.M., Berverly, D., Chorna, L., Wolf, T., & Pfeiffer, R.M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice, 26*(2), 214 – 219.

Clark, J.P., Adams, S.G., Dykstra, A.D., Moodie, S., & Jog, D. (2014). Loudness perception and speech intensity control in Parkinson's disease. *Journal of Communication Disorders, 51,* 1-12.

Conte, A., Khan, N., Defazio, G., Rothwell, J.C., & Berardelli, A. (2013). Pathophysiology of somatosensory abnormalities in Parkinson disease. *Nature Reviews Neurology, 9*(12), 687 – 697.

Conti, S., Bonazzi, S., Laiacona, M., Masina, M., & Coralli, M.V. (2015). Montreal Cognitive Assessment (MoCA) – Italian version: Regression based norms and equivalent scores. *Neurological Sciences, 36*(2), 209 – 214.

Costa, A.S., Fimm, B., Friesen, P., Soundjock, H., Rottschy, C., Gross, T., … Reetz, K. (2012). Alternate-form reliability of the Montreal Cognitive Assessment screening test in a clinical setting. *Dementia and Geriatric Cognitive Disorders, 33*(6), 379 – 384.

Dagenais, P.A., Garia, J.M., & Watts, C.R. (1998). Acceptability and intelligibility of mildly dysarthric speech by different listeners. In Cannito, M.P., Yorkston, K.M., & Beukelman, D.R. (Eds). *Neuromotor speech disorders: Nature, assessment and treatment,* (pp. 229 – 239). Baltimore, MD: Brookes.

Dagenais, P.A., Watts, C.R., Turnage, L.M., & Kennedy, S. (1999). Intelligibility and acceptability of moderately dysarthric speech by three types of listeners. *Journal of Medical Speech-Language Pathology, 7*(2), 91 – 95.

Dalrymple-Alford, J.C., MacAskill, M.R., Nakas, C.T., Livingston, L., Graham, C., Crucian, C.P., … Anderson, T.J. (2010). The MoCA: Well-suited screen for cognitive impairment in Parkinson disease. *Neurology, 75*(19), 1717 – 1725.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969a). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246 – 269.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969b). Cluster of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*(3), 462 – 496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). Hypokinetic dysarthria: disorders of extrapyramidal system. *Motor Speech Disorders*, 171-197.

De Bodt, M.S., Hernández-Díaz Huici, M.E., & Van De Heyning, P.H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders, 35,* 283 – 292.

Depaul, R, & Kent, R.D. (2000). A longitudinal case study of ALS: Effects of listener familiarity and proficiency on intelligibility judgements. *American Journal of Speech-Language Pathology, 9*(3), 230 – 240.

Donovan, N.J., Velozo, C.A. & Rosenbek, J.C. (2007). The communicative effectiveness survey: Investigating its item-level psychometric properties. *Journal of Medical Speech-Language Pathology, 15*(4), 433-447.

Donovan, N.J., Kendall, D.J., 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 – 47.

Dragone, M.L.S. (2011). Voice activation and participation profile presenting coordinate for readjustment of preventive action of educators. *Folia Phoniatrica et Logopaedica, 63*(1), 49 – 54.

Dromey, C. (2003). Spectral measures and perceptual ratings of hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology, 11*(2), 85 – 94.

Duarte De Almeida, L., Santos, L.R., Bassi, I.B., Teixeira, L.C., & Côrtes Gama, A.C. (2013). Relationship between adherence to speech therapy in patients with dysphonia and quality of life. *Journal of Voice, 27*(5), 617 – 621.

Duffy, J.R. (2013). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (3rd ed.). St. Louis, MO: Elsevier Mosby.

Dykstra, A.D., Adams, S.G., & Jog, M. (2007). The effects of botulinum toxin type A on speech intelligibility in lingual dystonia. *Journal of Medical Speech-Language Pathology, 15*(2), 173 – 186.

Dykstra, A.D., Adams, S.G., & Jog, M. (2012a). The effect of background noise on the speech intensity of individuals with hypophonia associated with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20*(3), 19 – 30.

Dykstra, A.D., Adams, S.G., & Jog, M. (2012b). Examining the conversational speech intelligibility of individuals with hypophonia associated with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20*(4), 53 – 57.

Dykstra, A.D., Adams, S.G., & Jog, M. (2015). Examining the relationship between speech intensity and self-rated communicative effectiveness in individuals with

Parkinson's disease and hypophonia. *Journal of Communication Disorders, 56,* 103 – 112.

Eadie, T.L., Yorkston, K.L., Klasner, E.R., Dudgeon, B.J., Deitz, J.C., Baylor, C.R., … Amtmann, D. (2006). Measuring communicative participation: A review of self-report instruments in speech-language pathology. *American Journal of Speech-Language Pathology, 15*(4), 307-320.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods, 39*(2), 175 – 191.

Fava, G., Paolillo, N.P., Oliveira, G., & Behlau, M. (2014). Cross-cultural adaptation of the Italian version of the Voice Activity Participation Profile. *Communication Disorders, Audiology and Swallowing, 26*(3), 252 – 255.

Feenaughty, L., Tjaden, K., & Sussman, J. (2014). Relationship between acoustic measures and judgements of intelligibility in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics, 28*(11), 857 – 878.

Fox, C.M., & Ramig, L.O. (1997). Vocal sound pressure level and self-perception of speech and voice in men and women with idiopathic Parkinson's disease. *American Journal of Speech-Language Pathology, 6,* 85-94.

Franca, M.C. (2013) A comparison of vocal demands with vocal performance among classroom student teachers. *Journal of Communication Disorders, 46*(1), 111 – 123.

Furlan, L. & Sterr, A. (2018). The applicability of standard error of measurement and minimal detectable change to motor learning research – A behavioral study. *Frontiers in Human Neuroscience, 12.*

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech and Language, 28,* 580 – 597.

Gill, D.J., Freshman, A., Blender, J.A., & Ravina, B. (2008). The Montreal Cognitive Assessment as a screening tool for cognitive impairment in Parkinson's disease. *Movement Disorders, 23*(7), 1043 – 1046.

Goberman, A.M., & Elmer, L.W. (2005). Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease. *Journal of Communication Disorders, 38,* 215 – 230.

Goetz, C.G., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stebbins, G.T., … LaPelle, N. (2007). Movement Disorders Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Movement Disorders, 22*(1), 41 – 47.

Gustafsson, J.K., Södersten, M. Ternström, S., Schalling, E. (2019). Long-term effects of Lee Silverman Voice Treatment on daily voice use in Parkinson's disease as measured with a portable voice accumulator. *Logopedics, Phoniatrics Vocology, 44*(3), 124 – 133.

Ho, A.K., Bradshaw, J.L., & Iansek, R. (2000). Volume perception in parkinsonian speech. *Movement Disorders, 15*(6), 1125 – 1131.

Ho., A.K., Bradshaw, J.L., Iansek, R., & Alfredson, R. (1999a). Speech volume regulation in Parkinson's disease: effects of implicit cues and explicit instructions. *Neuropsychologia, 37,* 1453 – 1460.

Ho, A.K., Iansek, R., & Bradshaw, J.L. (1999b). Regulation of Parkinsonian speech volume: The effect of interlocutor distance. *Journal of Neurology, Neurosurgery, and Psychiatry, 67,* 199 – 202.

Ho, A.K., Iansek, R., & Bradshaw, J.L. (2001). Motor instability in parkinsonian speech intensity. *Neuropsychiatry, Neuropsychology, and Behavioural Neurology, 14*(2), 109 – 116.

Hoops, S. Nazem, S., Siderowf, A.D., Duda, J.E., Xie, S.X., Stern, M.B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology, 73*(21), 1738 – 1745.

Huber, J.E., Stathopoulos, E.T., Ramig, L.O., & Lancaster, S.L. (2003). Respiratory function and variability in individuals with Parkinson disease: Pre- and post-Lee Silverman Voice Treatment. *Journal of Medical Speech-Language Pathology, 11*(4), 185 – 201.

Institute of Electrical and Electronics Engineers. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 225 – 246.

Jiang, J., O'Mara, T., Chen, H.-J., Stern, J.I., Vlagos, D., & Hanson, D. (1999). Aerodynamic measurements of patients with Parkinson's disease. *Journal of Voice, 13*(4), 583 – 591.

Keintz, K., Bunton, K., & Joit, J.D. (2007). Influence of visual information on the intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 16,* 222 – 234.

Kempler, D., & Van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease. *Brain and Language, 80,* 449 – 464.

Kent, R.D. (2004). The uniqueness of speech among motor systems. *Clinical Linguistics & Phonetics, 18*(6-8), 495 – 505.

Kent, R.D., Kent, J.F., Vorperian, H.K., & Duffy, J.R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal on Communication Disorders, 32,* 141 – 186.

Kent, R.D., Kent, J.F., Weismer, G., Suffit R.L., Rosenbek, J.C., Martin, R.E., & Brooks, B.R. (1990). Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders, 55,* 721 – 728.

Kent, R.D., Weismer, G., & Kent, J.F. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482 – 499.

Keppel, G. (1991). *Design and analysis: A researcher's handbook (3rd edition).* Englewood Cliff, NJ, US: Prentice-Hall, Inc.

Kleemola, L., Helminen, M., Rorarius, E., Isotalo, E., & Sihvo, M. (2011a). Voice activity and participation profile in assessing the effects of voice disorders on quality of life: Estimation of the validity, reliability and responsiveness of the Finnish version. *Folia Phoniatrica et Logopaedica, 63*(3), 113 – 121.

Kleemola, L., Helminen, M., Rorarius, E., Sihvo, M., & Isotalo, E. (2011b). Twelve-month clinical follow-up study of voice patients' recovery using the voice activity and participation profile (VAPP). *Journal of Voice, 25*(5), e245 – e245.

Kopecek, M., Stepankova, H., Lukavsky, J. Ripova, D., Nikolai, T., & Bezdicek, O. (2017). Montreal Cognitive Assessment (MoCA): Normative data for old and very old Czech adults. *Applied Neuropsychology: Adult, 24*(1), 23 – 29.

Kwan, L.C., & Whitehill, T.L. (2011). Perception of speech by individuals with Parkinson's disease: A review. *Parkinson's Disease, 11,* 1 – 11.

Liss, J.M., Spitzer, S.M., Caviness, J.N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 112*(6), 3022 – 3030.

Liu, T., Luo, H., Tang, J.Y., & Wong, G.H. (2018). Does lifestyle matter? Individual lifestyle factors and their additive effects associated with cognitive function in older men and women. *Aging and Mental Health, 6,* 1 – 8.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Developmental and psychometric evaluation of a functional measure for adult aphasia. *Journal of Speech and Hearing Disorders, 54*(1), 113 – 124.

Lu, W.-S., Chen, C.C., Huang, S.-L., & Hsieh, C.-L. (2007). Smallest real difference of 2 instrumental activities of daily living measures in patients with chronic stroke. *Archives of Physical Medicine and Rehabilitation, 93,* 1097 – 1100.

Ma, E.P.-M., & Yiu, E.M.-L. (2001). Voice activity and participation profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language, and Hearing Research, 44,* 511 – 524.

Martinello, J.G., Lauris, J.R.P., & Brasolotto, A.G. (2011). Psychometric assessments of life quality and voice for teachers within the municipal system, in Bauru, SP, Brazil. *Journal of Applied Oral Science,* 19(6), 573 – 578.

Matheron, D., Stathopoulos, E.T., Huber, J.E., & Sussman, J.E. (2017). Laryngeal aerodynamics in healthy older adults and adults with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 60*, 507 – 524.

McAuliffe, M.J., Baylor, C.R., & Yorkston, K.M. (2016). Variables associated with communicative participation in Parkinson's disease and its relationship to measures of health-related quality-of-life. *International Journal of Speech-Language Pathology,* 1 – 11.

McCaig, C.M., Adams, S.G., Dykstra, A.D., & Jog, M. (2016). Effect of concurrent walking and interlocutor distance on conversational speech intensity and rate in Parkinson's disease. *Gait and Posture, 43,* 132 - 136.

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders, 48*(6), 601 – 612.

Miller, N., Noble, E., Jones, D., & Burn, D. (2006). Life with communication changes in Parkinson's disease. *Age and Ageing, 35,* 235-239.

Mohun, S.H., Spitznagel, M.B., Gunstad, J.,Rochette, A., & Heinberg, L.J. (2018). Performance on the Montreal Cognitive Assessment (MoCA) in older adults presenting for bariatric surgery. *Obesity Surgery, 28*(9), 2700 – 2704.

Morse, J.M., & Johnson, J.L. (1991). *The Illness Experience – Dimensions of Suffering.* Sage: Newbury Park, California.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695 – 69.

Niwald, M., Redlicka, J., & Miller, E. (2018). Aerobic training enhances cognition and functional status in depressive multiple sclerosis patients. A case control pilot study. *Fizjoterapia Polska, 18*(2), 96 – 102.

Ojeda, N., del Pino, R., Ibarretxe – Bilbao, N., Schretlen, D.J., & Peña, J. (2016). Montreal Cognitive Assessment test: Normalization and standardization for Spanish population. *Revista de Neurologia, 63*(11), 488 – 496.

Pavlovic, A., Abel, K., Barlow, C.E., Farrell, S.W., Weiner, M., & DeFina, L.F. (2018). The association between serum vitamin d level and cognitive function in older adults: Cooper Center longitudinal study. *Preventive Medicine, 113,* 57 – 61.

Pennington, L. & Miller, N. (2007). Influence of listening conditions and listener characteristics on intelligibility of dysarthric speech. *Clinical Linguistics & Phonetics, 21*(5), 393 – 403.

Piwowarczyk, T.C., Oliveira, G., Loureno, L., & Behlau, M. (2012). Voice symptoms, voice activity, and participation profile and professional performance of call center operators. *Journal of Voice, 26*(2), 194 – 200.

Portney, L.G. & Watkins, M.P. (2000). *Foundations of clinical research: Applications to practice (2nd edition).* New Jersey, US: Prentice Hall Health.

Ramig, L., Countryman, S., O'Brien, C., Hoehn, M., Thompson, L. (1996). Intensive speech treatment for individuals with Parkinson's disease: Short- and long-term comparison of two techniques. *American Academy of Neurology, 47,* 1496 – 1504.

Ramig, L., Countryman, S., Thompson, L., & Horii, Y. (1995). A comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research, 38,* 1232 – 1251.

Ramig, L.O., Fox., C., & Sapir, S., (2004). Parkinson's disease: Speech and voice disorders and their treatment with the Lee Silverman Voice Treatment. *Seminars in Speech and Language, 25*(2), 169 – 180.

Ramig, L.O., Fox, C., & Sapir, S. (2008). Speech treatment for Parkinson's disease. *Expert Review of Neurotherapeutics, 8*(2), 297-309.

Ramig, L.O., Sapir, S., Fox, C., & Countryman, S. (2001). Changes in vocal loudness following intensive voice treatment (LSVT®) in individuals with Parkinson's disease: A comparison with untreated patients and normal age-matched controls. *Movement Disorders, 16*(1), 79 – 83.

Ricarte, A., Oliveira, G., & Behlau, M. (2013). Validation of the voice activity and participation profile protocol in Brazil. *Communication Disorders, Audiology and Swallowing, 25*(3), 242 – 249.

Rosen, K.M., Kent, R.D., Delaney, A.L., & Duffy, J.R. (2006). Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. *Journal of Speech, Language, and Hearing Research, 49,* 395 – 411.

Rosen, K.M., Kent, R.D., & Duffy, J.R. (2005). Task-based profile of vocal intensity decline in Parkinson's disease. *Folia Phoniatrica et Logopaedia, 57,* 28 – 37.

Santangelo, G., Siciliano, M. Pedone, R., Vitale, C., Falco, F. Bisogno, R., … Trojano, L. (2015). Normative data for the Montreal Cognitive Assessment in an Italian population sample. *Neurological Sciences, 36*(4), 585 – 591.

Sapir, S. (2014). Multiple factors are involved in the dysarthria associated with Parkinson's Disease: A review with implications for clinical practice and research. *Journal of Speech, Language, and Hearing Research, 57*, 1330-1343.

Sapir, S., Spielman, J.L., Ramig, L.O., Story, B.H., & Fox, C. (2007). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 50,* 899 – 912.

Schalling, E., Gustafsson, J., Ternström, S., Bulukin Wilén, F., Södersten, M. (2013). Effects of tactile biofeedback by a portable voice accumulator on voice sound level in speakers with Parkinson's disease. *Journal of Voice, 27*(6), 729 – 737.

Schmidt, C.P., Andrews, M.L., & McCutcheon, J.W. (1998). An acoustical and perceptual analysis of the vocal behavior of classroom teachers. *Journal of Voice, 12*(4), 434 – 443.

Schneider, S.A., & Obeso, J.A. (2014). Clinical and pathological features of Parkinson's Disease. *Current Topics in Behavioral Neuroscience, 10,* 126-128.

Schrag, A. Barone, P., Brown, R.G., Leentjens, A.F.G., McDonald, W.M., Starstein, S., … Goetz, C.G. (2007). Depression rating scales in Parkinson's disease: Critique and recommendations. *Movement Disorders, 22*(8), 1077 – 1092.

Shaik, M.A., Chan, Q.L., Xu, J., Xu, X., Hui, R.J.Y., Chong, S.S.T., … Dong, Y. (2016). Risk factors of cognitive impairment and brief cognitive tests to predict cognitive performance determined by a formal neuropsychological evaluation of primary health care patients. *Journal of the American Medical Directors Association, 17*(4), 343 – 347.

Sheikh, J.I., & Yesavage, J.A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist, 5*(1/2), 165 – 173.

Siciliano, M., Chiorri, C., Passaniti, C., Sant'Elia, V., Trojano, L., & Santangelo, G. (2019). Comparison of alternate and original forms of the Montreal Cognitive Assessment (MoCA): An Italian normative study. *Neurological Sciences, 40*(4), 691 – 702.

Simberg, S., Rae, J., Kallvik, E., Salo, B., & Martikainen, K. (2012). Effects of speech therapy on voice and speech in Parkinson's after a 15-day rehabilitation course: A pilot study. *International Journal of Therapy and Rehabilitation, 19*(5), 273 – 285.

Skodda, S., Flasskamp, A., & Schlegel, U. (2011). Instability of syllable repetition as a marker of disease progression in Parkinson's disease: A longitudinal study. *Movement Disorders, 26*(1), 59 – 64.

Smidt, N., van der Windt, D.A., Assendelft, W.J., Mourits, A.J., Devillé, W.L., de Winter, A.F., & Bouter, L.M. (2002). Interobserver reproducibility of the assessment of severity of complaints, grip, strength, and pressure pain threshold in patients with lateral epicondylitis. *Archives of Physical Medicine and Rehabilitation, 83,* 1145 – 1150.

Stathopoulos, E.T., Huber, J.E., Richardson, K., Kamphaus, J., DeCicco, D., Darling, M., … Sussman, J.E. (2014). Increased vocal intensity due to the Lombard effect in speakers with Parkinson's disease: Simultaneous laryngeal and respiratory strategies. *Journal of Communication Disorders, 48,* 1 – 17.

Steffen, T. & Seney, M. (2008). Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form healthy survey, and the Unified Parkinson's Disease Rating Sale in people with Parkinsonism. *Physical Therapy, 88*(6), 733 – 746.

Stipanic, K.L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research, 59*(2), 230 – 238.

Sukanen, O., Sihvo, M., Rorarius, E., Lehtihalmes, M., Autio, V., & Kleemola, L. (2007). Voice activity and participation profile (VAPP) in assessing the effects of voice disorders on patients' quality of life: Validity and reliability of the Finnish version of VAPP. *Logopedics Phoniatrics Vocology, 32*(1), 3-8.

Sussman, J.E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research, 55,* 1208 – 1219.

Tanner, M. Rammage, L., & Liu, L. (2016). Does singing and vocal strengthening improve vocal ability in people with Parkinson's disease? *Arts and Health, 8*(3), 199 – 212.

Thomann, A.E., Goettel, N., Monsch, R.J., Berres, M., Jahn, T., Steiner, L.A., & Monsch, A.U. (2018). The Montreal Cognitive Assessment: Normative data from a German-

speaking cohort and comparison with international normative samples. *Journal of Alzheimer's Disease, 64*(2), 643 – 655.

Tjaden, K.K. & Liss, J.M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics, 9*(2), 139 – 154.

Tjaden, K., Richards, E., Kuo, C., Wilding, G., & Sussman, J. (2013). Acoustic and perceptual consequences of clear and loud speech. *Folia Poniatria et Logopaedica, 65,* 214 – 220.

Tjaden, K., Sussman, J.E., & Wilding, G.E. (1994). 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, 57*(3), 779 – 792.

Tjaden, K., & Wilding, G. (2011). Effects of speaking task on intelligibility in Parkinson's disease. *Clinical Linguistics & Phonetics, 25*(2), 155 – 168.

Vaz, S., Falkmer, T., Passmore, A.E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One, 9;8*(9).

Walsh, B., & Smith, A. (2012). Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease. *Movement Disorders, 27*(7), 843 – 850.

Walshe, M., & Miller, N. (2010). Living with acquired dysarthria: the speaker's perspective. *Disability and Rehabilitation, Early Online,* 1 – 9.

Weintraub, D., Oehlberg, K.A., Katz, I.R., & Stern, M.B. (2006). Test characteristics of the 15-item Geriatric Depression Scale and Hamilton Depression Rating Scale in Parkinson's disease. *American Journal of Geriatric Psychiatry, 14*(2), 169 – 175.

Weintraub, D., Saboe, K., & Stern, M.B. (2007a). Effect of age on geriatric depression scale performance in Parkinson's disease. *Movement Disorders, 22*(9), 1331 – 1335.

Weintraub, D., Xie, S., Karlawish, J., & Siderowf, A. (2007b). Differences in depression symptoms in patients with Alzheimer's and Parkinson's diseases: Evidence from the 15-item Geriatric Depression Scale (GDS-15). *International Journal of Geriatric Psychiatry, 22*(10), 1025 – 1030.

Wirdefeldt, K., Adami, H., Cole, P., Trichopoulos, D., Mandel, J. (2011). Epidemiology and etiology of Parkinson's disease: A review of the evidence. *European Journal of Epidemiology, 26,* 1-58.

World Health Organization. (1997). *International classification of impairment, disability and handicap – Beta – 1: A manual of dimensions of disablement and participation.* Geneva, Switzerland: Author.

World Health Organization. (2001). *ICF: International Classification of Functioning, Disability and Health.* Geneva: Author.

Yorkston, K.M., Beukelman, D.R., & Tice, R. (1996). *Sentence Intelligibility Test.* Lincoln, NE: Institute for Rehabilitation Science and Engineering at Madonna Rehabilitation Hospital.

Yorkston, K. M., Beukelman, D. R., & Tice, R. (2011). *Sentence Intelligibility Test for Windows*. Lincoln, Communication Disorders Software. Lincoln, NE: Tice Technology Services.

Yorkston, K.M., Kuehn, C.M., Johnson, K.L., Ehde, D.M., Jensen, M.P., & Amtmann, D. (2008). Measuring participation in people living with multiple sclerosis: A comparison of self-reported frequency, importance and self-efficacy. *Disability and Rehabilitation, 30*(2), 88 – 97.

Zheng, L., Teng, E.L., Varma, R., Mack, W.J., Mungas, D., Lu., P.H., & Chui, H.C. (2012). Chinese-language Montreal Cognitive Assessment for Cantonese or Mandarin speakers: Age, education, and gender effects. *International Journal of Alzheimer's Disease, 2012,* 1 – 10.

Appendices

Appendix A: Geriatric Depression Scale - 15-Item Short Form

Instructions: Choose the best answer for how you have felt over the past week.

No.	Question	Answer	Score
1.	Are you basically satisfied with your life?	YES/NO	
2.	Have you dropped many of your activities and interests?	YES/NO	
3.	Do you feel that your life is empty?	YES/NO	
4.	Do you often get bored?	YES/NO	
5.	Are you in good spirits most of the time?	YES/NO	
6.	Are you afraid that something bad is going to happen to you?	YES/NO	
7.	Do you feel happy most of the time?	YES/NO	
8.	Do you often feel helpless?	YES/NO	
9.	Do you prefer to stay at home, rather than going out and doing new things?	YES/NO	
10.	Do you feel you have more problems with memory than most people?	YES/NO	
11.	Do you think it is wonderful to be alive?	YES/NO	
12.	Do you feel pretty worthless the way you are now?	YES/NO	
13.	Do you feel full of energy?	YES/NO	
14.	Do you feel that your situation is hopeless?	YES/NO	
15.	Do you think that most people are better off than you are?	YES/NO	
		TOTAL	

Sheikh, J.I., & Yesavage, J.A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist, 5(1/2), 165 – 173.

Appendix B: Medication Effectiveness Scale

With this scale, we ask you to rate the effectiveness of your medication. Please read the statement below. Then rate the effectiveness of your medication at managing your symptoms of Parkinson's disease. If you feel your medication is very effective, mark the 7. If you feel your medication is not at all effective, mark the 1. Feel free to use any number on the scale

How effective is your medication at managing your symptoms of Parkinson's disease?

Not at all effective	1	2	3	4	5	6	7	Very effective

Appendix C: Typical Speech Loudness

In this survey we ask you to rate your typical speech loudness. Please respond to the following questions by putting a cross ("X") on the line which best represents your response. A cross towards the left side means your speech is very quiet while a cross towards the right side means your speech is normal loudness.

What is your typical speech loudness?

Very quiet _____ Normal loudness

Appendix D: Communication Effectiveness Survey

In this survey we ask you to rate how effective your speech is in different communication situations. Please read each statement. Then rate how effectively you communicate in that situation. If you feel your speech is very effective, mark the 4. If your speech does not allow you to communicate at all in a situation, mark the 1. Feel free to use any number on the scale

1. Having a conversation with a family member or friends at home.

Not at all effective			Very effective
1	2	3	4

2. Participating in conversation with strangers in a quiet place.

Not at all effective			Very effective
1	2	3	4

3. Conversing with a familiar person over the telephone.

Not at all effective			Very effective
1	2	3	4

4. Conversing with a stranger over the telephone.

Not at all effective			Very effective
1	2	3	4

5. Being part of a conversation in a noisy environment (social gathering).

Not at all effective			Very effective
1	2	3	4

6. Speaking to a friend when you are emotionally upset or you are angry.

Not at all effective			Very effective
1	2	3	4

7. Having a conversation while traveling in a car.

Not at all effective			Very effective
1	2	3	4

8. Having a conversation with someone at a distance (across a room).

Not at all effective			Very effective
1	2	3	4

Donovan, N.J., Kendall, D.J., 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 – 47.

Appendix E: Voice Activity and Participation Profile

Please answer the following questions by putting a cross ("X") on the line which best represents your answer. A cross towards the left side means you are never affected while a cross towards the right side means you are always affected.

Self-perceived severity of voice problem

1. How severe is your voice problem now?

   Normal _____ Severe

Effect on job

2. Is your job affected by voice problem?

   Never _____ Always

3. In the last 6 months, have you thought of changing your job because of your voice problem?

   Never _____ Always

4. Has your voice problem created any pressure on your job?

   Never _____ Always

5. In the last 6 months, has your voice problem affected your decisions for your future career?

   Never _____ Always

Effect on daily communication

6. Do people ask you to repeat what you have just said because of your voice problem?

   Never _____ Always

7. In the last 6 months, have you ever avoided talking to people because of your voice problem?

   Never _____ Always

8. Do people have difficulty understanding you on the phone because of your voice problem?

   Never _____ Always

9. In the last 6 months, have you reduced the use of the telephone because of your voice problem?

   Never _____ Always

10. Does your voice problem affect your communication in quiet environments?

    Never _____ Always

11. In the last 6 months, have you ever avoided having conversations in quiet environments because of your voice problem?

    Never _____ Always

12. Does your voice problem affect your communication in noisy environments?

    Never _____ Always

13. In the last 6 months, have you ever avoided having conversations in noisy environments because of your voice problem?

    Never _____ Always

14. Does your voice problem affect your message when speaking to a group of people?

    Never _____ Always

15. In the last 6 months, have you ever avoided having conversations in a group because of your voice problem?

    Never _____ Always

16. Does your voice problem affect getting your message across?

    Never _____ Always

17. In the last 6 months, have you ever avoided speaking because of your voice problem?

    Never _____ Always

Effect on social communication

18. Does your voice problem affect you in social activities?

    Never _____ Always

19. In the last 6 months, have you ever avoided social activities because of your voice problem?

Never _____ Always

20. Are your family, friends, or co-workers annoyed by your voice problem?

Never _____ Always

21. In the last 6 months, have you ever avoided communicating with your family, friends, or co-workers because of your voice problem?

Never _____ Always

Effect on your emotion

22. Do you feel upset about your voice problem?

Never _____ Always

23. Are you embarrassed by your voice problem?

Never _____ Always

24. Do you have low self-esteem because of your voice problem?

Never _____ Always

25. Are you worried about your voice problem?

Never _____ Always

26. Do you feel dissatisfied because of your voice problem?

Never _____ Always

27. Does your voice problem affect your personality?

Never _____ Always

28. Does your voice problem affect your self-image?

Never _____ Always

Ma, E.P.-M., & Yiu, E.M.-L. (2001). Voice activity and participation profile: Assessing the impact of voice disorders on daily activities. Journal of Speech, Language, and Hearing Research, 44, 511 – 524.

Appendix F: Communication Participation Item Bank

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 this 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.

Question	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?**	○	○	○	○
2. Does your condition interfere with … **… communicating when you need to say something quickly?**	○	○	○	○
3. Does your condition interfere with … **… talking with people you do NOT know?**	○	○	○	○
4. Does your condition interfere with … **… communicating when you are out in your community (e.g., errands, appointments)?**	○	○	○	○
5. Does your condition interfere with … **… asking questions in a conversation?**	○	○	○	○
6. Does your condition interfere with … **… communicating in a small group of people?**	○	○	○	○
7. Does your condition interfere with … **… having a long conversation with someone you know about a book, movie, show, or sports event?**	○	○	○	○
8. Does your condition interfere with … **… giving someone DETAILED information?**	○	○	○	○
9. Does your condition interfere with … **… getting your turn in a fast-moving conversation?**	○	○	○	○
10. Does your condition interfere with … **… trying to persuade a friend of family member to see a different point of view?**	○	○	○	○

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, 56, 1190-1208.

Appendix G: Level of Speech Usage Scale

How Do You Use Your Speech?
While communication is important to everyone, different people use their speech in different ways. Think of how you have typically used your speech over the *past year.* Choose the category below that best describes you.

_____ ***Undemanding:***
Quiet for long periods of time almost every day:
Almost never
- talk for long periods
- raise your voice above a conversational level,
- participate in group discussions, give a speech or other presentation

_____ ***Intermittent:***
Quiet for long periods of time on **many days**
Most talking is **typical conversational speech**
Occasionally:
- talk for longer periods
- raise voice above conversational level
- participate in group discussions, give a speech or other presentation

_____ ***Routine:***
Frequent periods of talking on **most days**
Most talking is **typical conversational speech**
Occasionally:
- talk for longer periods
- raise voice above conversational level
- participate in group discussions, give a speech or other presentation

_____ ***Extensive:***
Speech usage **consistently goes beyond everyday conversational speech.**
Regularly:
- talk for long periods
- talk in a loud voice
- participate in group discussions, give presentations or performances

Although the demands of your speech are often high, you are able to continue with most work or social activities even if your speech is not perfect.

_____ ***Extraordinary:***
Very high speech demands
Regularly:
- talk for long periods of time
- talk with loud or expressive speech or
- give presentations or performances.

The success of your work or personal goals depends almost entirely on the quality of your speech and voice.

Baylor, C., Yorkston, K.M., Eadie, T., Miller, R., & Amtmann, D. (2008). Levels of speech usage: A self-report scale for describing how people use speech. Journal of Medical Speech-Language Pathology, 16(4), 191 – 198.

Appendix H: Letter of Information for Participants with Parkinson's Disease

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

Letter of Information for Participants with Parkinson's disease

1. **Invitation to Participate**
   You are invited to participate in this research study investigating the variability of speech loudness, speech intelligibility, and communicative participation in individuals with Parkinson's disease (PD) and hypophonia (reduced loudness). You have been invited to participate because you have been diagnosed with PD and hypophonia.

2. **Purpose of the Letter**
   The purpose of this letter is to provide you with information required to make an informed decision regarding participation in this research study.

3. **Purpose of this Study**
   The purpose of this study is to examine the variability of speech and psychosocial measures over time in individuals with a speech impairment resulting from

Parkinson's disease. Psychosocial measures refer to the self-assessment of communication in different social situations. For example, we could ask you to rate your effectiveness as a communicator when speaking to a stranger on the telephone. This study also aims to explore relationships among speech measures, psychosocial measures, and non-speech factors

4. **Inclusion Criteria**
   To be eligible to participate in this study, you must:
   1. have been diagnosed with idiopathic PD for a minimum of 3 years
   2. be between 55 and 85 years of age
   3. speak English as your first language.
5. **Exclusion Criteria**
   You are ineligible to participate if you:
   1. cannot read and/or write
   2. do not pass a 40 dB hearing screening test
   3. have a history of speech, language, hearing, or neurological impairments, with the exception of those that related to PD
   4. are currently receiving speech language therapy, or will be receiving speech language therapy within the next month
   5. have undergone deep brain stimulation surgery, or will be undergoing deep brain stimulation surgery within the next month
   6. receive a score below 21 on the Montreal Cognitive Assessment (a questionnaire used to assess cognitive ability).
6. **Study Procedures**
   This study will be conducted over 3 visits. The first and second visits will occur within a 5-day period. The third visit will occur approximately 4 weeks following the second visit. The first and second visits will each last approximately **90 minutes**. The third visit will last approximately **60 minutes**.

   During **visit one**, you will be asked to take a basic hearing screening. It is anticipated that completion of the hearing screening will take approximately 5 minutes. The principle experimenter or another member of the research team will ask you questions regarding your date of birth, general medical history, neurological history, and speech and hearing history. You will also be asked to complete a series of seven questionnaires related to your speech and communication. These questionnaires will look at how you use your speech on a daily basis, your typical speech loudness, your effectiveness as a communicator in different social situations, your participation in communication settings, and the impact of your speech on everyday life. You will also be asked to complete two additional questionnaires that will examine your cognitive abilities and screen for depression. It is anticipated that completion of all nine questionnaires will take approximately 45 minutes. In addition, you will be asked to perform various

speech tasks. These speech tasks will include syllable repetition, reading aloud multiple sentences, and taking part in conversation while being audio-recorded with a microphone. Only the researchers will have access to the speech recordings. The audio files will be encrypted and stored on a secure computer at Western University. It is anticipated that the completion of the speech tasks will take approximately 20 minutes.

During **visit 2**, you will be asked to complete the Unified Parkinson's Disease Rating Scale (UPDRS; a scale used to assess your PD symptoms). It is anticipated that completion of the UPDRS will take approximately 20 minutes. You will also be asked to complete the same series of seven questionnaires related to speech and communication from visit 1. It is anticipated that completion of the questionnaires will take approximately 30 minutes. Additionally, you will be asked to perform the same speech tasks from visit 1. It is anticipated that completion of the speech tasks will take approximately 20 minutes.

During **visit 3**, you will be asked to complete the same series of seven questionnaires from visits 1 and 2. It is anticipated that completion of the questionnaires will take approximately 30 minutes. In addition, you will be asked to perform the same speech tasks from visits 1 and 2. It is anticipated that the completion of the speech tasks will take approximately 20 minutes.

It is anticipated that the total participation time for this study will be approximately 4 hours, upon completion of all 3-study visits. This study will involve 50 participants with PD, 50 communication partners of participants with PD, 20 healthy control participants, and 10 naïve listeners.

If you agree to participate, you will be asked to come to the Principal Investigator's Lab in Elborn College (Room 2212) at Western University for repeated administration of questionnaires, and multiple speech recordings. While at Elborn College, you will be provided with free parking.

7. Possible Risks and Harms

There are no known or anticipated physical risks or discomfort associated with participation in this study. The experiment will be conducted in a safe and hygienic university laboratory with adequate lighting and ventilation. You will be seated in a comfortable chair throughout the procedures. To help counteract any fatigue you may experience through the duration of the experiment, you will be given rest breaks at approximately ten-minute intervals or more frequently if requested. You have the right to withdrawal from the study any time.

As part of this study, the Geriatric Depression Scale (a questionnaire used for screening depression) will be used. In the event that this questionnaire indicates the presence of depression, your physician, Dr. Mandar Jog (co-investigator in this study), will be informed of the findings.

As part of this this study, you will undergo a brief hearing screening. This hearing screening should not be considered a substitute for a formal hearing assessment conducted by a licensed audiologist. Should your hearing threshold results be above 40dB during our screening procedure, we will offer you a referral to the H.A. Leeper Speech and Hearing Audiology Clinic, located within Elborn College at Western University, or you may discuss an audiology referral with your family physician.

8. Possible Benefits

The procedures that will be used during this study are experimental in nature and will not provide any direct benefit to the participant's medical condition. However, it is anticipated that results from this study may provide important information about the stability of speech and communication in individuals with PD and hypophonia. The potential benefits to society include the improvement of the understanding of hypophonia in individuals with PD, and the development of more effective tools for identifying the impact and severity of hypophonia in individuals with PD.

9. Compensation

You will not be compensated for your participation in this study. However, free on-site parking will be provided. A daily visitor's parking pass will be provided to you upon your arrival to the Elborn College parking lot.

10. Voluntary Participation

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions, or withdraw from the study at any time with no effect on your future treatment or medical care.

11. Confidentiality

All data collected will remain confidential. Your name and any identifying information will be collected separately from the data. All data collected with no personal identifiers will be retained indefinitely. All data collected with personal identifiers will be retained for 15 years following publication. If you choose to withdraw from this study, your data will be immediately removed and destroyed from our database. Our research records will be locked in a cabinet in the principal investigator's secure lab in Elborn College, Western University. Audio recordings from participants will be de-identified. The de-identified audio files

will be encrypted and stored on the secure hard drive of a single desktop computer in the principle investigator's lab in Elborn College at Western University. Listener participants will make perceptual ratings from de-identified audio recordings. All other data collected will remain accessible only to the investigators of this study. Representatives of the University of Western Ontario Health Sciences Research Ethics Board may require access to your study-related documents to oversee the ethical conduct of this study. Representatives of Lawson Quality Assurance Education Program may require access to your study-related documents to ensure that proper laws and guidelines are being followed. You do not waive any legal rights by signing the consent form.

12. Publication

If the results of the study are published, your name will not be used and no information that discloses your identity will be released or published. If requested, you will be provided with a copy of any publication related to the results of this study when it becomes available.

13. Contacts for Further Information

If you require any further information regarding this research project or your participation in the study, you may contact Dr. Allyson Page at [redacted], or Dr. Scott Adams at [redacted]

If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Human Research Ethics [redacted]

If you agree to participate in this study, please sign the consent form on the next page.

Sincerely,

Allyson Page, PhD	Scott Adams, PhD	Cynthia Mancinelli
Associate Professor	Professor	MClSc/PhD candidate

This letter is yours to keep for future reference.

Appendix I: Consent Form for Participants with Parkinson's Disease

Consent Form for Participants with Parkinson's Disease

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

I have read the Letter of Information and have had the nature of the study explained to me, and I agree to participate. All questions have been answered to my satisfaction.

Signature of Research Participant	Printed Name	Date

Signature of Person Obtaining Informed Consent	Printed Name	Date

Appendix J: Letter of Information for Control Participants

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

Letter of Information for Participants without Parkinson's Disease

14. Invitation to Participate

You are invited to participate in this research study investigating the variability of speech loudness, speech intelligibility, and communicative participation in individuals with Parkinson's disease (PD) and hypophonia (speech loudness). You have been invited to participate because you have not been diagnosed with PD and/or hypophonia.

15. Purpose of the Letter

The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research study.

16. Purpose of this Study

The purpose of this study is to examine the variability of speech and psychosocial measures over time in individuals with a speech impairment resulting from

Parkinson's disease. Psychosocial measures refer to the self-assessment of communication in different social situations. For example, we could ask you to rate your effectiveness as a communicator when speaking to a stranger on the telephone. This study also aims to explore relationships among speech measures, psychosocial measures, and non-speech factors.

17. Inclusion Criteria

To be eligible to participate in this study, you must:

4. be between 55 and 85 years of age
5. have not been diagnosed with idiopathic PD
6. speak English as your first language.

18. Exclusion Criteria

You are ineligible to participate if you:

7. cannot read and/or write
8. do not pass a 40 dB hearing screening test
9. have a history of speech, language, hearing, or neurological impairments
10. receive a score below 26 on the Montreal Cognitive Assessment (a questionnaire used to assess cognitive ability).

19. Study Procedures

This study will be conducted over 3 visits. The first and second visits will occur within a 5-day period. The third visit will occur approximately 4 weeks following the second visit. The first visit will last approximately **90 minutes**. The second and third visits will each last approximately **60 minutes**.

During **visit one**, you will be asked to take a basic hearing test. It is anticipated that completion of the hearing screening will take approximately 5 minutes. The principle experimenter, or another member of the research team, will ask you questions regarding your date of birth, general medical history, neurological history, and speech and hearing history. You will also be asked to complete a series of seven questionnaires related to your speech and communication. These questionnaires will look at how you use your speech on a daily basis, your typical speech loudness, your effectiveness as a communicator in different social situations, your participation in different communication settings, and the impact of your speech on your everyday life. You will also be asked to complete two additional questionnaires that examine your cognitive abilities and screen for depression. It is anticipated that completion of all nine questionnaires will take approximately 45 minutes. In addition, you will be asked to perform various speech tasks. These speech tasks will include syllable repetition, reading aloud multiple sentences, and taking part in conversation while being audio-recorded with a microphone. Only the researchers will have access to these speech recordings. The audio files will be encrypted and stored on a secure computer at

Western University. It is anticipated that completion of the speech tasks will take approximately 20 minutes.

During **visit 2**, you will be asked to complete the same series of seven questionnaires related to your speech and communication from visit 1. It is anticipated that completion of the questionnaires will take approximately 30 minutes. Additionally, you will be asked to perform the same speech tasks from visit 1. It is anticipated that completion of the speech tasks will take approximately 20 minutes.

During **visit 3**, you will be asked to complete the same series of seven questionnaires related to your speech and communication from visits 1 and 2. It is anticipated that completion of the questionnaires will take approximately 30 minutes. Additionally, you will be asked to perform the same speech tasks from visits 1 and 2. It is anticipated that completion of the speech tasks will take approximately 20 minutes.

It is anticipated that the total participation time for this study will be 3.5 hours upon completion of all 3-study visits. This study will involve 50 participants with PD, 50 communication partners of participants with PD, 20 healthy control participants, and 10 naïve listeners.

If you agree to participate, you will be asked to come to the Principal Investigator's Lab in Elborn College (Room 2212) at the University of Western Ontario for repeated administration of questionnaires, and multiple speech recordings. While at Elborn College, you will be provided with free parking.

20. Possible Risks and Harms

There are no known or anticipated physical risks or discomfort associated with participation in this study. The experiment will be conducted in a safe and hygienic laboratory with adequate lighting and ventilation. You will be seated in a comfortable chair throughout the procedures. To help counteract any fatigue you may experience through the duration of the experiment, you will be given rest breaks at approximately ten-minute intervals or more frequently if requested. You have the right to withdrawal from the study any time if you feel discomfort.

As part of this study, the Geriatric Depression Scale (a questionnaire used for screening depression) will be used. In the event that this questionnaire indicates the presence of depression, you will be informed of the findings and recommended that you discuss the results with your family physician. You will also be provided with a list of clinics and helplines.

As part of this this study, you will undergo a brief hearing screening. This hearing screening should not be considered a substitute for a formal hearing assessment conducted by a licensed audiologist. Should your hearing threshold results be above 40dB during our screening procedure, we will offer you a referral to the H.A. Leeper Speech and Hearing Audiology Clinic, located within the School of Communication Sciences and Disorders at Western University, or you may discuss an audiology referral with your family physician.

21. Possible Benefits

The procedures that will be used during this study are experimental in nature and will not provide any direct benefit to the participant. However, it is anticipated that the results from this study may provide important information about the stability of speech and communication in individuals with PD and hypophonia. The potential benefits to society include the improvement of the understanding of hypophonia in individuals with PD, and the development of more effective tools for identifying the impact and severity of hypophonia in individuals with PD.

22. Compensation

You will not be compensated for your participation in this study. However, free on-site parking will be provided. A daily visitor's parking pass will be provided to you upon your arrival to the Elborn College parking lot.

23. Voluntary Participation

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions, or withdraw from the study at any time, with no effects on your future treatment or medical care.

24. Confidentiality

All data collected will remain confidential. Your name and any identifying information will be collected separately from the data. All data collected with no personal identifiers will be retained indefinitely. All data collected with personal identifiers will be retained for 15 years following publication. If you choose to withdraw from this study, your data will be immediately removed and destroyed from our database. Our research records will be locked in a cabinet in the principal investigator's secure lab in Elborn College, Western University. Audio recordings from participants will be de-identified. The de-identified audio files will be encrypted and stored on the secure hard drive of a single desktop computer in the principal investigator's lab in Elborn College at Western University. Listener participants will make perceptual ratings from de-identified audio recordings. All other data collected will remain accessible only to the investigators of this study. Representatives of the University of Western Ontario

Health Sciences Research Ethics Board may require access to your study-related documents to oversee the ethical conduct of this study. Representatives of Lawson Quality Assurance Education Program may require access to your study-related documents to ensure that proper laws and guidelines are being followed. You do not waive any legal rights by signing the consent form.

25. Publication

If the results of the study are published, your name will not be used, and no information that discloses your identity will be released or published. If requested, you will be provided with a copy of any publication related to the results of this study when it becomes available.

26. Contacts for Further Information

If you require any further information regarding this research project or your participation in the study, you may contact Dr. Allyson Page at [redacted], or Dr. Scott Adams at [redacted]

If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Human Research Ethics [redacted].

If you agree to participate in this study, please sign the consent form on the next page.

Sincerely,

Allyson Page, PhD	Scott Adams, PhD	Cynthia Mancinelli
Associate Professor	Professor	MClSc/PhD candidate

This letter is yours to keep for future reference.

Appendix K: Consent Form for Control Participants

Consent Form for Participants without Parkinson's Disease

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

I have read the Letter of Information and have had the nature of the study explained to me, and I agree to participate. All questions have been answered to my satisfaction.

Signature of Research Participant	Printed Name	Date

Signature of Person Obtaining Informed Consent	Printed Name	Date

Appendix L: Letter of Information for Primary Communication Partners of Participants with Parkinson's Disease

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

Letter of Information for Communication Partner Participants

27. Invitation to Participate

You are invited to participate in this research study investigating the variability of speech loudness, speech intelligibility, and communicative participation in individuals with Parkinson's disease (PD) and hypophonia (reduced loudness). You have been invited to participate because you are the primary communication partner of an individual who has been diagnosed with PD and hypophonia.

28. Purpose of the Letter

The purpose of this letter is to provide you with information required to make an informed decision regarding participation in this research study.

29. Purpose of this Study

The purpose of this study is to examine the variability of speech and psychosocial measures over time in individuals with a speech impairment resulting from Parkinson's disease. Psychosocial measures refer to the self-assessment of communication in different social situations. For example, we could ask you to rate your partner's effectiveness as a communicator when speaking to a stranger on the telephone. This study also aims to explore relationships among speech measures, psychosocial measures, and non-speech factors.

30. Inclusion Criteria

To be eligible to participate in this study, you must:

7. be between 18 and 85 years of age
8. be the primary communication partner of a participant with PD, or someone who regularly converses with a participant with PD
9. speak English as your first language.

31. Exclusion Criteria

You are ineligible to participate if you:

11. cannot read and/or write
12. do not pass a 40 dB hearing screening test
13. are not the primary communication partner of a participant with PD, or someone who regularly converse with a participant with PD.

32. Study Procedures

If you agree to participate, you will be asked to attend 3 visits with your communication partner with PD. This study will be conducted over 3 visits. The first and second visits will occur within a 5-day period. The third visit will occur approximately 4 weeks following the second visit. The first and second visits will each last approximately **90 minutes**. The third visit will last approximately **60 minutes**.

During visit 1, you will first be asked to take basic hearing screening test and to provide your age. It is anticipated that completion of the hearing screening will take approximately 5 minutes. During all three visits, you will be asked to complete a series of seven questionnaires related to the speech and communication of your partner who has PD. These questionnaires will look at your perceptions of how your communication partner with PD uses his/her speech on a daily basis, his/her typical speech loudness, his/her effectiveness as a communicator in different social situations, his/her participation in communication settings, and the impact of his/her voice on their everyday life. It is anticipated that the seven questionnaires will take approximately 30 minutes to complete during each visit. It is anticipated that your total active participation time for this study is 1.5 hours over the course of all 3-study visits. It is

anticipated that the total participation time for your communication partner with PD is approximately 4 hours over the course of the three study visits.

If you agree to participate you will be asked to come to the Principal Investigator's Lab in Elborn College (Room 2212) at Western University for a hearing screening, and repeated administration of questionnaires. While at Elborn College, you will be provided with free parking.

33. Possible Risks and Harms

There are no known or anticipated physical risks or discomfort associated with participation in this study. The experiment will be conducted in a safe and hygienic university laboratory with adequate lighting and ventilation. You will be seated in a comfortable chair throughout the procedures. To help counteract any fatigue you may experience through the duration of the experiment, you will be given rest breaks at approximately ten-minute intervals, or more frequently if requested. You have the right to withdrawal from the study any time.

As part of this this study, you will undergo a brief hearing screening. This hearing screening should not be considered a substitute for a formal hearing assessment conducted by a licensed audiologist. Should your hearing threshold results be above 40dB during our screening procedure, we will offer you a referral to the H.A. Leeper Speech and Hearing Audiology Clinic, located within Elborn College at Western University, or you may discuss an audiology referral with your family physician.

34. Possible Benefits

The procedures that will be used during this study are experimental in nature and will not provide any direct benefit to the participant. However, it is anticipated that results from this study may provide important information about the stability of speech and communication in individuals with PD and hypophonia. The potential benefits to society include the improvement of the understanding of hypophonia in individuals with PD, and the development of more effective tools for identifying the impact and severity of hypophonia in individuals with PD.

35. Compensation

You will not be compensated for your participation in this study. However, free on-site parking will be provided. A daily visitor's parking pass will be provided to you upon your arrival to the Elborn College parking lot.

36. Voluntary Participation

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions, or withdraw from the study at any time with no effects to you or your communication partner's future treatment or medical care.

37. Confidentiality

All data collected will remain confidential. Your name and any identifying information will be collected separately from the data. All data collected with no personal identifiers will be retained indefinitely. All data collected with personal identifiers will be retained for 15 years following publication. If you choose to withdraw from this study, your data will be immediately removed and destroyed from our database. Our research records will be locked in a cabinet in the principal investigator's secure lab in Elborn College, Western University. Representatives of the University of Western Ontario Health Sciences Research Ethics Board may require access to your study-related documents to oversee the ethical conduct of this study. Representatives of Lawson Quality Assurance Education Program may require access to your study-related documents to ensure that proper laws and guidelines are being followed. You do not waive any legal rights by signing the consent form.

38. Publication

If the results of the study are published, your name will not be used and no information that discloses your identity will be released or published. If requested, you will be provided with a copy of any publication related to the results of this study when it becomes available.

39. Contacts for Further Information

If you require any further information regarding this research project or your participation in the study, you may contact Dr. Allyson Page at [redacted], or Dr. Scott Adams at [redacted]

If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Human Research Ethics [redacted].

If you agree to participate in this study, please sign the consent form on the next page.

Sincerely,

Allyson Page, PhD	Scott Adams, PhD	Cynthia Mancinelli
Associate Professor	Professor	MClSc/PhD candidate

This letter is yours to keep for future reference.

Appendix M: Consent Form for Primary Communication Partners of Participants with Parkinson's Disease

Consent Form for Communication Partner Participants

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

I have read the Letter of Information and have had the nature of the study explained to me, and I agree to participate. All questions have been answered to my satisfaction.

Signature of Research Participant	Printed Name	Date

Signature of Person Obtaining Informed Consent	Printed Name	Date

Appendix N: Letter of Information for Listeners

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders; Health and Rehabilitation Sciences
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

Letter of Information for Listener Participants

1. **Invitation to Participate**
   You are invited to participate in this research study investigating the variability of speech loudness, speech intelligibility, and communicative participation in individuals with Parkinson's disease (PD) and hypophonia (speech loudness). You have been invited to participate because you have normal hearing ability and English is your first language.

2. **Purpose of the Letter**
   The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research study.

3. **Purpose of this Study**

The purpose of this study is to examine the variability of speech and psychosocial measures over time in individuals with a speech impairment resulting from Parkinson's

disease. Psychosocial measures refer to the self-assessment of communication in different social situations. For example, this may be a rating of one's effectiveness as a communicator when speaking to a stranger on the telephone. This study also aims to explore relationships among speech measures, psychosocial measures, and non-speech factors

4. **Inclusion Criteria**

To be eligible to participate as a listener in this study, you must be 18 years of age or older, have normal hearing ability, and speak English as your first language.

5. **Exclusion Criteria**

   If you have a history of hearing, language, or cognitive impairments, cannot read or write, or do not pass a 25 dB hearing screening, you are not eligible to participate in this study. Additionally, you will be excluded from the study if you have extensive research or clinical experience with individuals with PD.

6. **Study Procedures**

   If you agree to participate, you will first be asked to take a basic hearing screening and to provide your age as well as general information about your medical, speech and hearing, and neurological history. The study involves listening to pre-recorded speech samples of individuals with PD. You will be asked to transcribe the audio samples heard. You will also be asked to rate the audio samples with regard to speech intelligibility.

   This study will involve 50 participants with PD, 50 communication partners of participants with PD, 20 healthy control participants, and 10 naïve listeners. It is anticipated that the entire experiment will take approximately 4 hours to complete over 2 two-hour sessions. The tasks will be conducted in Dr. Allyson Page's lab, which is located in Elborn College, room 2504. There will be a total of 10 listeners participating in this study.

7. **Possible Risks and Harms**

   There are no known or anticipated risks or discomforts associated with participation in this study. The experiment will be conducted in a safe, hygienic, university laboratory with adequate lighting and ventilation. The experimental procedures will require minimal physical effort, you will be seated in a comfortable chair, and given rest breaks at approximately ten-minute intervals or more frequently if requested.

   As part of this this study, you will undergo a brief hearing screening. This hearing screening should not be considered a substitute for a formal hearing assessment conducted by a licensed audiologist. Should your hearing threshold results be

above 25dB during our screening procedure, we will offer you a referral to the H.A. Leeper Speech and Hearing Audiology Clinic, located within the Elborn College at Western University, or you may discuss an audiology referral with your family physician.

8. Possible Benefits

There is no direct benefit to participation in this study. The potential benefits to society include an improved understanding of the speech and psychosocial measures associated with PD.

9. Compensation

You will not be compensated for your participation in this study. However, free on-site parking will be provided. A daily visitor's parking pass will be provided to you upon your arrival to the Elborn College parking lot.

10. Voluntary Participation

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions, or withdraw from the study at any time, with no effect on your academic status, course evaluation, or grades in any way.

11. Confidentiality

All data collected will remain confidential. Your name and any identifying information will be collected separately from the data. All data collected with no personal identifiers will be retained indefinitely. All data collected with personal identifiers will be retained for 15 years following publication. If you choose to withdraw from this study, your data will be immediately removed and destroyed from our database. Our research records will be locked in a cabinet in the principal investigator's secure lab in Elborn College, Western University. Listener participants will make perceptual ratings from de-identified audio recordings. All other data collected will remain accessible only to the investigators of this study. Representatives of the University of Western Ontario Health Sciences Research Ethics Board may require access to your study-related documents to oversee the ethical conduct of this study. Representatives of Lawson Quality Assurance Education Program may require access to your study-related documents to ensure that proper laws and guidelines are being followed. You do not waive any legal rights by signing the consent form.

12. Contacts for Further Information

If you require any further information regarding this research project or your participation in the study, you may contact Dr. Allyson Page at [redacted], or Dr. Scott Adams at [redacted]

If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Human Research Ethics [redacted], email: [redacted]

13. Publication

If the results of the study are published, your name will not be used and no information that discloses your identity will be released or published. If you would like to receive a copy of any potential study results, please contact Dr. Allyson Page or Dr. Scott Adams.

If you agree to participate in this study, please sign the consent form on the next page.

Sincerely,

Allyson Page, PhD	Scott Adams, PhD	Cynthia Mancinelli
Associate Professor	Professor	MClSc/PhD candidate

This letter is yours to keep for future reference.

Appendix O: Consent Form for Listeners

Consent Form for Listener Participants

Project Title:
Examining the temporal variability of speech intensity, speech intelligibility, and communicative participation in individuals with hypophonia and Parkinson's disease

Principal Investigators:
Allyson Page, PhD
Associate Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Scott Adams, PhD
Professor
School of Communication Sciences and Disorders
Western University
[redacted]

Co-Investigators:
Cynthia Mancinelli
MClSc/PhD Candidate, Speech and Language Science
Health and Rehabilitation Sciences
Western University

Dr. Mandar Jog, MD, FRCPC
Director
Movement Disorders Program; Clinical Neurological Sciences
London Health Sciences Centre, University Campus, and Western University

I have read the Letter of Information and have had the nature of the study explained to me, and I agree to participate. All questions have been answered to my satisfaction.

Signature of Research Participant	Printed Name	Date

Signature of Person Obtaining Informed Consent	Printed Name	Date

Appendix P: Visual Analogue Scale for Ratings of Speech Intelligibility

Curriculum Vitae

Name: Cynthia Mancinelli

Post-secondary Education and Degrees:
Concordia University
Montreal, Quebec, Canada
2009-2013 B.Sc. (Hons)

Western University
London, Ontario, Canada
2013-2019 M.Cl.Sc.

Western University
London, Ontario, Canada
2013-2019 Ph.D.

Honours and Awards:
Province of Ontario Graduate Scholarship
1993-1994, 1994-1995

Related Work Experience:
Teaching Assistant
Western University
2015-2017

Research
NSERC Undergraduate Student Award

Publications:

Abeyesekera, A., Adams, S., Mancinelli, C., Knowles, T., Gilmore, G., Delrobaei, M., Jog, M. (2019). Effects of Deep Brain Stimulation of the Subthalamic Nucleus Settings on Voice Quality, Intensity, and Prosody in Parkinson's Disease: Preliminary Evidence for Speech Optimization. Canadian Journal of Neurological Sciences, 46(3), 287-294

Knowles, T., Adams, S., Abeyesekera, A., Mancinelli, C., Gilmore, G., Jog, M. (2018). Deep Brain Stimulation of the Subthalamic Nucleus Parameter Optimization for Vowel Acoustics and Speech Intelligibility in Parkinson's Disease. Journal of Speech, Language, and Hearing Research, 61(3), 510-524

Almey, A., Arena, L., Oliel, J. Hafez, N., Mancinelli, C., Henning, L., Tsanev, A., Brake, W.G. (2017). Interactions Between Estradiol and Haloperidol on Perseveration and Reversal Learning in Amphetamine-Sensitized Female Rats. Hormones and Behavior, 89(1), 113-120.

Poster Presentations:

Mancinelli, C., Dykstra-Page, A., Dworschak-Stokan, A., & Husein, M. (2016). Evaluating Self-Perceived Communication Competence in Adult Speakers with Velopharyngeal Insufficiency. American Speech and Hearing Association Conference, Pennsylvania, USA

Mancinelli, C., Adams, S., Abeyesekera, A., Knowles, T., Delrobaei, M., & Jog, M. (2016). Deep Brain Stimulation Parameter Optimization for Rate of Speech in Parkinson's Disease. American Speech and Hearing Association Conference, Pennsylvania, USA

Dykstra, A., Siegel, L., Mancinelli, C., Wilson, C., & Jog, M. (2016). Examining Speech Intelligibility and Self-Rated Communication-Related Quality of Life in Individuals with Oromandibular Dystonia Receiving Botulinum Toxin Therapy. Motor Speech Conference, California, USA

Abeyesekera, A., Adams, S., Knowles, T., Mancinelli, C., Delrobaei, M. & Jog, M. (2016). Deep Brain Stimulation Parameter Optimization for Voice Quality, Speech Intensity, and Prosody of Speech in Parkinson's Disease. Motor Speech Conference, California, USA

Knowles, T., Adams, S., Abeyesekera, A., Mancinelli, C., Delrobaei, M., & Jog, M. (2016). Deep Brain Stimulation Parameter Optimization for Speech Intelligibility and Acoustics in Parkinson's Disease. Motor Speech Conference, California, USA

Mancinelli, C., Dykstra, A., Dworschak-Stokan, A., & Husein, M. (2015). Evaluating Communication-Related Quality-of-Life in Adult Speakers with Velopharyngeal Insufficiency. American Speech and Hearing Association Conference, Denver, USA

Wilson, C., Dykstra, A., Siegel, L., Mancinelli, C., & Jog, M. (2015). An Exploration of the Perception of Communication Competence in Speakers with OMD Receiving Botulinum Toxin Injections. Faculty of Health Sciences Research Day, Western University

Wilson, C., Dykstra, A., Siegel, L., Mancinelli, C., & Jog, M. (2015). Exploring Self-Perceived Communication Competence in Speakers with Oromandibular Dystonia. Rehabilitation Research Colloquium, Queen's University

Mancinelli, C., Dykstra, A., Wilson, C., Siegel, L., Dworschak-Stokan, A., & Husein, M. (2015). Communication-Related Quality of Life and Speech Intelligibility in Adults with Velopharyngeal Insufficiency. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Siegel, L., Dykstra, A., Mancinelli, C., & Wilson, C. (2015). Examining Ratings of Communication-Related Quality of Life in Speakers with Oromandibular Dystonia Receiving Botulinum Toxin Therapy. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Wilson, C., Dykstra, A., Siegel, L., Mancinelli C., & Jog, M. (2015). Examining Ratings of Self-Perceived Communication Competence in Speakers with Oromandibular Dystonia Receiving Botulinum Toxin Therapy. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Siegel, L., Dykstra, A., Mancinelli, C., & Wilson, C. (2015). Examining Ratings of Communication-Related Quality of Life in Speakers with Oromandibular Dystonia Receiving Botulinum Toxin Therapy. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Mancinelli, C., Adams, S., Abeyesekera, A., Knowles, T., Delrobaei, M., & Jog, M. (2015). Deep Brain Stimulation Parameter Optimization for Rate of Speech in Parkinson's Disease. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Abeyesekera, A., Adams, S., Knowles, T., Mancinelli, C., Delrobaei, M., & Jog, M,. (2015). Effect of Different Deep Brain Stimulation Parameters on Voice Quality, Speech Intensity and Prosody in Parkinson's Disease. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Knowles, T., Adams, S., Abeyesekera, A., Mancinelli, C., Delrobaei, M., & Jog, M. (2015). Effect of Deep Brain Stimulation Parameters on Speech Intelligibility and Speech Acoustics in Parkinson's Disease. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Dykstra, A., Mancinelli, C., Domingo, Y., Dworschak-Stokan, A., & Husein, M. (2014). Examining Communicative Effectiveness in Adult Speakers with Velopharyngeal Insufficiency. American Speech and Hearing Association Conference, Florida, USA

Abeyesekera, A., Adams, S., Mancinelli, C., Rahimi, F., Delrobaei, M., & Jog, M. (2014). Deep Brain Stimulation Parameter Optimization for Speech in Parkinson's Disease, Southern Ontario Neuroscience Association, Western University

Mancinelli, C., Domingo, Y., Dykstra, A., Dworscak-Stokan, A., & Husein, M. (2014). An Exploration of the Relationships Between Speech Intelligibility, Hypernasality, and Self-Ratings of Communicative Effectiveness in Adults with Velopharyngeal Insufficiency. Aging, Rehabilitation, and Geriatric Care Symposium, Western University

Abeyesekera, A., Adams, S., Mancinelli, C., Rahimi, F., Delrobaei, M., & Jog, M. Deep Brain Stimulation Parameter Optimization for Speech in Parkinson's Disease. Aging, Rehabilitation, and Geriatric Care Symposium, Western University

Mancinelli, C., Domingo, Y., Dykstra, A., Dworscak-Stokan, A., & Husein, M. (2014). An Examination of Speech Intelligibility, Hypernasality, and Self-Ratings of Communicative Effectiveness in Adults with Velopharyngeal Insufficiency. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Abeyesekera, A., Adams, S., Mancinelli, C., Rahimi, F., Delrobaei, M., & Jog, M. (2014). Optimal Deep Brain Stimulation Settings for Speech in Parkinson's Disease. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Oral Presentations:

Mancinelli, C. (2017). Speech Therapy. Guest lecture for Resonance, Western University

Abeyesekera, A., Knowles, T., Mancinelli, C., Adams, S., Delrobaei, M., & Jog, M. Deep Brain Stimulation and Speech. 3rd Annual Movement Disorders Research Retreat, May 2016

Mancinelli, C. (2016). Pursuing a Doctorate. Guest lecture for Issues in Professional Practice, Western University

Mancinelli, C. (2016). Exploring the Temporal Variability of Speech Intensity, Speech Intelligibility, and Communicative Participation in Individuals with Hypophonia and Parkinson's Disease. Speech Language Sciences Seminar, Western University

Knowles, T., Adams, S., Abeyesekera, A., Mancinelli, C., Delrobaei, M., & Jog, M. (2016). Deep Brain Stimulation Parameter Optimization for Speech Intelligibility and Vowel Acoustics in Parkinson's Disease. International Clinical Phonetics and Linguistics Association Conference, Halifax, Nova Scotia, Canada

Mancinelli, C. (2016). Exploring Communicative Participation in Hypophonia and Parkinson's Disease. Health and Rehabilitation Sciences Graduate Research Conference, Western University

Mancinelli, C. (2015). Pursuing a Doctorate. Guest lecture for Issues in Professional Practice, Western University

Mancinelli, C. (2015). Evaluating Communication-Related Quality-of-Life in Adult Speakers with Velopharyngeal Insufficiency. Speech Language Sciences Seminar, Western University

Mancinelli, C. (2015). Evaluating communication-related quality of life in adults with velopharyngeal insufficiency. Rehabilitation Research Colloquium, Queen's University.

Mancinelli, C. (2015). Deep brain stimulation parameter optimization of rate of speech in Parkinson's disease. Rehabilitation Research Colloquium, Queen's University.

Mancinelli, C., Adams, S., Abeyesekera, A., Knowles, T., Delrobaei, M., & Jog, M. (2015). Deep brain stimulation parameter optimization for rate of speech in Parkinson's disease. 2nd Annual Movement Disorders Research Retreat, University Hospital.

Abeyesekera, A., & Mancinelli, C. (2014). Deep brain stimulation parameter optimization for speech in Parkinson's disease. 1st Annual Movement Disorders Research Retreat, University Hospital.