

**Measuring the Test-Retest Reliability of Communicative Congruence and Communicative
Dysphoria**

by

Connor Mahon

Bachelor of Arts, Cleveland State University, 2022

Submitted to the Graduate Faculty of the
School of Health and Rehabilitation Sciences in partial fulfillment
of the requirements for the degree of
Master of Science in Communication Science and Disorders

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH
SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This thesis was presented

by

Connor Mahon

It was defended on

March 8, 2024

and approved by

James Coyle, PhD, CCC-SLP, BCS-S, ASHA Fellow, Professor, Department of Communication
Science and Disorders

Jason Bohland, PhD, Assistant Professor, Department of Communication Science and Disorders

Committee Chair: Leah B. Helou, PhD, CCC-SLP, Assistant Professor, Department of
Communication Science and Disorders

Copyright © by Connor Mahon

2024

Measuring the Test-Retest Reliability of Communicative Congruence and Communicative Dysphoria

Connor Mahon, BA

University of Pittsburgh, 2024

Various communicative qualities can convey meaningful information about a speaker's identity and personality. When the qualities and associated connotations of someone's communication behaviors do not align with their own self-identities and experiences, they may experience distress or discomfort. Two constructs describe this phenomenon. Communicative congruence (CC) is the experience of alignment between one's sense of self and their voice, speech, and communication. Communicative dysphoria (CD) is significant emotional, motivational, and/or cognitive distress due to communicative incongruence. Both constructs are framed within the Cybernetic Big Five Theory of personality.

CC and CD were previously measured in a sample of 366 individuals. In order to understand the behavior of these constructs over time, this study repeated 2 measures of CC and 2 measures of CD in subset of the 366 participants at three time points. Test-retest reliability was assessed using an interclass correlation coefficient. Findings indicated that communicative congruence and communicative dysphoria will be more reliable when measured in the short-term, and that CC has greater test-retest reliability than CD across measurement approaches. Clinical application of CC and CD measurement include as an outcome measure and prognostic indicator. Future work is necessary to understand cross-group variance and clinical value.

Table of Contents

1.0 Introduction.....	1
1.1 Relevant Pre-Existing Measures	2
1.2 Theoretical Basis for Communicative Congruence and Dysphoria.....	6
1.3 State Versus Trait.....	8
1.4 Research Question	9
2.0 Methods.....	11
2.1 Participants and Pre-Existing Data	11
2.2 Survey Modifications.....	12
2.3 Survey Distribution	16
2.4 Statistical Approach	16
3.0 Results	17
3.1 Communicative Congruence (CC) Scale	17
3.2 Communicative Congruence (CC) Slider.....	19
3.3 Dysphoria Severity Slider	20
3.4 Disposition Slider	21
4.0 Discussion.....	23
4.1 Clinical Implications	26
4.2 Limitations and Future Directions.....	27
5.0 Conclusions.....	30
Bibliography	31

List of Tables

Table 1 Communicative Congruence (CC) Likert Scale.....	13
Table 2 Visual Analog Scales	14

List of Figures

Figure 1 CC Scale scores measured across timepoints – Baseline & T1.....	18
Figure 2 CC Scale scores measured across timepoints – T1 & T2.....	18
Figure 3 CC Scale scores measured across timepoints – T2 & T3.....	18
Figure 4 CC Slider ratings measured across timepoints – Baseline & T1.....	19
Figure 5 CC Slider ratings measured across timepoints – T1 & T2.....	19
Figure 6 CC Slider ratings measured across timepoints – T2 & T3.....	20
Figure 7 Dysphoria Severity Slider ratings measured across timepoints – Baseline & T1..	20
Figure 8 Dysphoria Severity Slider ratings measured across timepoints – T1 & T2.....	21
Figure 9 Dysphoria Severity Slider ratings measured across timepoints – T2 & T3.....	21
Figure 10 Disposition Slider ratings measured across timepoints – Baseline & T1.....	22
Figure 11 Disposition Slider ratings measured across timepoints –T1 & T2.....	22
Figure 12 Disposition Slider ratings measured across timepoints – T2 & T3.....	22

1.0 Introduction

Various communicative qualities can convey meaningful information about a speaker's gender (Pernet & Belin, 2012; Simpson, 2009), race (Walton & Orlikoff, 1994), age (Pettorino et al., 2014), body size (Pisanki et al., 2016), and sexual orientation (Levon, 2006). They also carry additional culturally constructed information including that relating to group membership, value judgements such as trustworthiness, and aspects of personality (see review by Podesva & Callier, 2015; Welch et al., 2021). When the qualities and associated connotations of someone's communication behaviors do not align with their own self-identities and experiences, they may experience distress or discomfort.

Many individuals seek professional help in aligning their communication with their identity and experiences (i.e., their overall "sense of self") to address a primary complaint relating to perceived incongruence. For some individuals, such a misalignment between communication and sense of self may arise secondary to voice therapy, phonosurgery, or other processes that change voice and communication including voice disorders, aging, laryngectomy, or use of alternative and augmentative communication (Crow et al., 2019). Clinical anecdotes suggest that such misalignments may be an important factor in adherence to and motivation in voice therapy, especially when the new or target behaviors increase the sense of misalignment.

A limited body of research has been published on the topic, with the most comprehensive exploration to date from Welch & Helou (2022). They outline two newly named constructs, *communicative congruence* and *communicative dysphoria*. Communicative congruence is "the state in which a person's voice, speech, and/or language are consistent with their prevailing sense

of self/self-schemas." (Welch & Helou, 2022, p. 3421). When that ideal state is not met—i.e., the person’s voice, speech, and/or language is not consistent with their sense of self—they experience some level of communicative incongruence. Relatedly, communicative dysphoria represents an experience of disliking the way one sounds when they talk. The communicative dysphoria construct arose from transgender individuals’ lived experiences and descriptions of *vocal dysphoria*, which is a feeling of discord associated with the sound of one’s own voice. Welch and Helou expanded that term to include all features of spoken communication, hence the use of the term “communicative dysphoria.” This terminological choice was driven by the fact that the average person does not know the difference between voice, speech, and language, and because any of these features might comprise the element(s) of communication that feels incongruent or disliked by an individual. One major goal of gender affirming voice and communication training is to increase the client’s feelings of congruence between their sense of self and their communicative behaviors (i.e., improving communicative congruence). When that occurs, the experience of communicative dysphoria is expected to lessen in severity.

1.1 Relevant Pre-Existing Measures

Gender affirming voice and communication training is the process of facilitated behavioral change to modify the sound of one’s voice and/or other communication behaviors. Training may be sought for a variety of reasons, both externally motivated (e.g., “I want others to perceive my voice as a certain gender”) and/or internally motivated (e.g., “The sound of my voice is a trigger for gender dysphoria”). In most cases, outcomes are measured through client satisfaction or

acoustic measures of pitch, resonance, and other salient features for gender identification. Despite the fact that communicative incongruence and dysphoria have driven the treatment-seeking behaviors described in the literature for decades, existing measures of communicative congruence or dysphoria have only recently become available.

One tool used to measure outcomes in this population is the Trans Woman Voice Questionnaire (TWVQ), a 30-item questionnaire measuring vocal function and social participation with an emphasis on the experiences of those seeking vocal feminization (Dacakis et al., 2017). This measure was additionally adapted to be appropriate for transmasculine individuals, changing only the gendered words (Kreukels et al., 2012). It asks questions related to communicative congruence (e.g., “I feel my voice does not reflect the ‘true me’”), however, gender is the sole aspect of identity explored, and it is not designed to be suitable outside of the binary transgender community.

Another tool recently published is the Communication Participation Item Bank - Gender-Diverse Version, an adapted version of the Communication Participation Item Bank that makes several wording changes, such as the removal of pathologizing language, to make it more appropriate and applicable to gender diverse people seeking gender affirming voice and communication services (Baylor et al., 2024). This tool is suitable for any gender presentation or communication goals and is unique among similar outcome measures in its lack of questions regarding specific communication domains, such as voice. Rather, it poses a variety of scenarios, such as “How difficult is it for you to have a conversation when you are upset?” or “How difficult is it for you to talk with people you do NOT know?” Communication participation is certainly a relevant and functional measure for individuals experiencing communicative dysphoria. However,

this measure does not itself capture communicative congruence and communicative dysphoria as defined here, but rather the hypothesized impact on life participation.

The Utah Gender Presentation Scale for Communication is a new 10-item patient reported outcome measure for gender incongruence specific to voice and communication (Myers et al., 2024). In contrast to the lack of specific communication parameters seen in the Communication Participation Item Bank - Gender-Diverse Version, each item in the Utah Gender Presentation Scale for Communication is a 10 point scale from masculine to feminine for a specific communication parameter. It includes pitch, intonation, resonance, loudness, speech smoothness, speech clarity, word choice, facial expression, gesture, and posture. For each item, clients select a score for their current presentation, and their goal presentation. An incongruence score is recorded as the sum of the difference between the two points for each item.

An additional tool in development for the assessment of communicative congruence is the Gender Voice Index (GVI), previously the Gender Spectrum Voice Inventory (Van Hook & Duffy, 2021). This tool collects patient goals in a scale and free text form and includes a range of Likert scale questions assessing feelings as they relate to speech and voice. Words used to describe vocal congruence vary widely, including comfort, safety, authenticity, unity, alignment, satisfaction, reflective, accurate, connected, and representative. The scale is advertised as appropriate to use for “any trans* / gender diverse individual who is seeking trans voice training, or any cisgender individual who experiences gender-voice incongruence (e.g. with presbyphonia).” (Van Hook, 2023).

Both the U-GPS and GVI have a primary focus on gender as it relates the experience of communicative congruence. While they are inclusive of a broader gender spectrum than the TWVQ, the emphasis on gender limits their applicability to the broader concepts of

communicative congruence and dysphoria, and the variety of reasons they may or may not be experienced. Gender is a salient aspect of identity conveyed by voice. However, it is not the only aspect of identity that is meaningful to a given individual, nor the only one that may cause feelings of communicative dysphoria. Additionally, experiences of communicative congruence and dysphoria are not exclusive to the transgender and gender diverse community (Welch & Helou, 2022; Helou et al., 2022). Therefore, the predominant focus on gender expression and identity in these measures, as well as the Trans Woman Voice Questionnaire, limits their usefulness in measuring communicative congruence and dysphoria at a whole-person level of psychological robustness.

Other clinical researchers are interested in related concepts of vocal/communicative congruence. Specifically, the Vocal Congruence Scale (Crow et al. 2019) is a patient reported measure intended for use in the general population. It is rooted in the idea of “interoceptive awareness,” an individual’s awareness of interoceptive processes such as those indicated by gastrointestinal and cardiac sensations. This scale has been used as a method of conceptualizing and quantifying the body-brain relationship through the relationship of physiological and psychological states, thereby representing a “bodily sense of self” (Crow et al., 2019, p. 324.e16). High interoceptive awareness is shown to be correlated with higher scores on measures of anxiety, personality, and emotional lability across several studies. In their experiment utilizing a heartbeat detection task, participants exhibiting high interoceptive awareness were significantly more likely to have higher scores on the VCS, which authors interpreted as being indicative of a higher degree of vocal congruence. Authors hypothesized that those with high interoceptive awareness do “not rely on external cues to determine vocal identity” (Crow et al., 2019, p. 324.e17), and instead have a more stable bodily sense of self than those individuals with lower interoceptive awareness.

While the VCS does refer to gender in one question (“I feel as though my voice reflected my gender”), it differs from other measures presented here as it is not limited to a particular gender identity. Questions on the VCS are relatively more generalized in nature, for example, "it seemed like my voice belonged to me," and "it seemed like my voice reflected who I am." Still yet, this perspective is limited to voice and neglects to account for the full range of spoken communication behaviors, for example, accent or social ability.

1.2 Theoretical Basis for Communicative Congruence and Dysphoria

The communicative congruence and dysphoria constructs presented by Welch and Helou (2022) are high-order and multidimensional in that they are (a) inclusive of all gender identities; (b) inclusive of major aspects of self-identity; and (c) inclusive of all features of spoken communication. They frame the constructs within the Cybernetic Big Five Theory (CB5T) of personality. This is a modern personality theory that includes the five classically accepted “domains” of personality: Neuroticism, Agreeableness, Conscientiousness, Extraversion, and Openness/Intellect. Each domain is further divided into “aspects” that are subdivided into “facets” (e.g., the domain of Agreeableness is comprised of the aspects Compassion and Politeness, which are then subdivided into facets, such as Empathy and Concern from the former, and Compliance and Selflessness from the latter). The CB5T theory also includes two higher order *metatraits*, Stability and Plasticity, which combine certain domains. These theoretical structures of personality are considered alongside *characteristic adaptations*, which are defined as “relatively stable goals,

interpretations, and strategies, specified in relation to an individual's particular life circumstances" (DeYoung, 2015, p. 38). Self-identity can be conceptualized as coordination between these three components: goals, interpretations, and strategies (DeYoung, 2015, p. 38).

Under CB5T, an experience of incongruence between one's experience and characteristic adaptations is expected to result in psychological *entropy*, "a measure of disorder that describes the amount of uncertainty or unpredictability in any information system" (Shannon, 1948, in DeYoung, 2015). The higher order traits of Plasticity and Stability relate to how well one responds to entropy and its related uncertainty. Stability (comprised of three domains: Neuroticism, Agreeableness, and Conscientiousness) is the metatrait that protects existing characteristic adaptations, allowing one to remain stable when faced with entropy. Plasticity (comprised of two domains: Extraversion and Openness/Intellect) is the metatrait that allows one to adjust or generate new characteristic adaptations, including new goals, interpretations, and strategies (DeYoung, 2015; Welch & Helou, 2021).

Welch & Helou (2022) hypothesized that people with higher levels of Plasticity would report more communicative congruence. Their logic was that individuals higher in Plasticity are more likely to work to minimize any incongruencies with their communication behaviors due to the inherent flexibility of their characteristic adaptations. Utilizing CB5T's associated 100-item self-report personality measure, the Big Five Aspect Scales (BFAS, DeYoung et al., 2007), they found that participants who scored higher on the BFAS in the metatrait Plasticity did indeed report significantly higher levels of communicative congruence than those low in Plasticity. Thus, those more easily able to shift or modify their sense of identity are more likely to experience a high degree of consistency between their communication and sense of self.

To explore the negative associations with experiencing communicative incongruence, they also administered to participants the Center for Epidemiologic Studies-Depression Scale (CES-D) (Radloff, 1977). They found that lower communicative congruence predicted high scores on the CES-D, indicating higher reports of depression symptoms. This finding highlights the clinical importance of considering communicative congruence as a component of mental wellbeing.

1.3 State Versus Trait

Psychological constructs are commonly measured and understood in terms of their behavioral as a *state* that is highly variable depending on the situation and specific context of an interaction, or a *trait* that is more consistently observable of an individual (Horstmann & Ziegler, 2020). A widely recognized example of this is anxiety, which exists in everyone at both the state and trait level. As one example, anxiety about giving a presentation to a particular audience is considered state anxiety, while trait anxiety might manifest as a predisposition to experience anxiety in all situations where they are observed by others (Endler & Kocovski, 2001). In discussing the reliability of a state measure, it is necessary to consider that variance is a fundamental and expected property of the state constructs (Horstmann & Ziegler, 2020). Thus, test-retest reliability can serve as an indicator of the behavior of the constructs.

It remains unclear whether communicative dysphoria and communicative congruence exist more like a state or if they are more consistently observable as a trait of an individual. Some preliminary data exist on this front. Specifically, Welch & Helou (2021a) assessed the intra-responder variability of communicative congruence and communicative dysphoria by returning

to 71 of their previous participants. They replayed previously recorded speech samples to participants, who provided new scores for their self-perceived communicative congruence and communicative dysphoria. In that study, communicative congruence showed moderate test-retest reliability, indicating it may be a relatively stable construct within individuals (i.e., closer to a trait than state). In contrast, communicative dysphoria had low test-retest reliability, indicating that dysphoria might fluctuate more over time and therefore might exist more like a state than a trait. However, replication of this finding has not yet been pursued. Further, the time between those observations varied widely from 217 to 592 days, with a mean of 368 days. In such a wide time frame, numerous experiences might have impacted participants' perceptions of their communicative congruence. Though several formal guidelines exist for construction of trait measures, fewer exist for development of state measures. At minimum, in order to make any inferences about whether communicative congruence and dysphoria function as states, traits, or both, future studies would need to include multiple observations that are more proximal to each other than in the study by Welch and Helou (2021a).

1.4 Research Question

The primary research question of the present study is: What is the test-retest reliability of communicative dysphoria and communicative congruence when measured at long and short intervals? Building on the preliminary findings of Welch & Helou (2021a), we followed up with

a second set of experimental participants to redistribute previous self-report items on communicative congruence and dysphoria. We hypothesized that communicative congruence would demonstrate moderate test-retest reliability, and communicative dysphoria would demonstrate low test-retest reliability in this new sample. This hypothesis is based on the theoretical framework underpinning both constructs, which frames communicative congruence at the metatrait level of Plasticity and communicative dysphoria at the lower-order level of characteristic adaptations (Welch & Helou, 2021b). Theoretically, an individual's level of Plasticity is relatively more stable and trait-like whereas their characteristic adaptations are relatively more dynamic and state-like. Further, short intervals are hypothesized to result in greater test-retest reliability, as they minimize the effects of natural change and the influence of major life events when measuring personality constructs (Watson, 2004), thus allowing for a more accurate interpretation of the constructs' behaviors as a state or a trait.

2.0 Methods

2.1 Participants and Pre-Existing Data

The present study leveraged a convenience sample of pre-existing research participants. Between August 30, 2021 and November 8, 2022 a survey was distributed by the Helou Laboratory to the population at large and to specific online communities. The survey specifically recruited participants who dislike how they sound when they talk. It invited participants to share their experience and make ratings about communicative congruence and communicative dysphoria, as well as about symptoms of depression utilizing the Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977), and personality utilizing the Big Five Aspects Scale (BFAS) (DeYoung et al., 2007). Recruitment materials explicitly targeted adults who disliked the sound of their own voice or speech, using the following text:

Hate the sound of your voice? Hate the way you talk? Wish you communicated differently? We want to hear from people who don't like or even hate the sound of their voice, speech, or other communication features.

A total of 515 individuals began the survey and 380 completed it. Following manual review for free text responses inappropriate for the questions asked, 16 responses were removed and 366 participants remained in the dataset. Within this group, a total of 198 provided email addresses to be considered for the opportunity to receive a gift card for their participation.

2.2 Survey Modifications

The original survey was consolidated to narrow focus on two measures each of communicative congruence and communicative dysphoria. The modified survey also included demographic questions. It was hosted in REDCap (Harris et al., 2009; Harris et al., 2019), a web application for the hosting of online surveys. Table 1 presents 8 items related to the construct of communicative congruence, presented with a 7-point Likert scale. These were initially developed through expert opinion with feedback provided by a medical doctor and public health researcher, one personality research psychologist, one clinical psychologist, two voice-specialized speech-language pathologists, and one speech science researcher. Further revisions were made by the two voice-specialized speech-language pathologists and based on their extensive clinical experience in the treatment of voice disorders and gender affirming voice and communication services. Table 2 presents one question pertaining to communicative congruence and two addressing communicative dysphoria, all with 100-point visual analog scales.

Table 1 Communicative Congruence (CC) Likert Scale

1	My voice matches my gender.
2	I think my voice should belong to someone else.
3	I feel like my voice accurately represents "me."
4	The way I talk reflects my personality.
5	People tend to make correct assumptions about me based on how I communicate.
6	I am fine with hearing myself on recordings, because that voice generally matches the way I think I sound.
7	I have low self-esteem related to how I communicate.
8	I feel I have to compensate in other ways to express who I am (e.g., physical presentation), because my communication is not sufficient.

Table 2 Visual Analog Scales

<p>Communicative Congruence (CC) Slider</p>	<p>How congruent do you feel your voice and speech are with your sense of self?</p> <p>That is, individuals with HIGH congruence feel that their voice and speech accurately reflect who they are.</p> <p>Individuals with LOW congruence feel that their voice and speech do not match their sense of self.</p>
	<p><i>Extremely Incongruent (0) ----- Extremely Congruent (100)</i></p>
<p>Dysphoria Severity Slider</p>	<p>Communicative dysphoria is a feeling someone might have if they don't like their voice, speech, or other features of communication. The technical term refers to the significant distress, unease, dissatisfaction, and emotional, motivational, cognitive, or behavioral dysregulation that some people may experience due to their voice, speech, and/or other features of communication.</p> <p>How would you describe the severity of your communicative dysphoria?</p>
	<p><i>None (0) ---- Extreme (100)</i></p>
<p>Disposition Slider</p>	<p>How do you feel about your voice and speech?</p> <p><i>Hate it so much that it negatively affects my life (0) ----</i></p> <p><i>Love it so much that it positively impacts my life (100)</i></p>

Although not the focus of this investigation, participants were also asked to complete the CES-D, the Severity Measure for Generalized Anxiety Disorder—Adult (SMGAD-A, Craske et al., 2013, a concise measure of anxiety.), the Mini-International Personality Item Pool (Mini-IPIP, a concise measure of personality, Goldberg, 1999; Donnellan et al., 2006), and the Big Five Aspect Scales (BFAS). Participants were also given the opportunity to share additional information about their experiences of communicative congruence and dysphoria in a free text response field.

Questions were transferred identically from the survey previously completed by these individuals, with some exceptions. First, in a series of communicative congruence questions with standard 7-point Likert response options (where 1 represents “Strongly disagree” and 7 represents “Strongly agree”), two of the original questions included additional response options. Specifically, the question “People tend to make correct assumptions about me based on how I communicate” included the response option “I have no idea what assumptions people make about me based on how I communicate.” In addition, the question “I am fine with hearing myself on recordings, because that voice generally matches the way I think I sound” included the response option “Not applicable.” These options were originally included to be as thorough as possible, but the lack of parallel structure limited analysis and interpretation. As such, those extra response options were removed so that only the 7-point Likert scale remained.

The second exception is that a content warning was removed following the definition of communicative dysphoria. The original survey contained up to 40 questions related to experiences with communicative dysphoria, which had the potential to trigger an adverse emotional response. For the present study, the modified survey included only a single 1-100 scale to describe severity of communicative dysphoria, and therefore had limited potential to trigger an adverse emotional response.

2.3 Survey Distribution

Participants who shared their email address after completion of the baseline survey were invited via email to complete the newly modified survey (T1). Participants who responded were invited to participate in the same survey again, three weeks after their date of completion (T2). A third invitation to participate was sent after an additional three weeks following completion (T3). If a participant had not yet completed the survey for any of these time points, one reminder email was sent after 48 hours. Once a participant failed to complete any of these surveys, they were no longer included in the list of those individuals receiving follow-up invitations to later time points.

2.4 Statistical Approach

Following the data collection period, baseline data was matched to responses by matching email addresses of participants. Then, the analyses performed on the original test were repeated. Specifically, all scores were normalized to z scores to allow for easier comparison across measures. Test-retest reliability was examined using an interclass correlation coefficient (ICC) for single fixed raters (ICC3), which measures the relationship between two tests, to assess the reliability of responses over time. The ICC test generates a value between 0 and 1, where poor reliability is represented by values below 0.5, moderate reliability by values between 0.5 and 0.75, good reliability by values between 0.75 and 0.9, and excellent reliability by values above 0.9 (Koo & Li, 2016).

3.0 Results

Of those who completed the Baseline Survey, 198 provided email addresses and were contacted for the present survey. Of those contacted, 44 completed the T1 survey. Of these, 30 completed the survey at T2, and 24 completed the survey at T3. Of the T1 respondents, 20 (45.45%) identified as cisgender. The mean age was 28.75 (SD=6.95). The majority (84.09%) of participants reported their race as white, with 13.64% indicating Asian, and 4.55% Black. One participant (2.25%) identified as Hispanic or Latinx/e. Time between the baseline measure and T1 varied widely (Mean = 509.62 days, SD = 168.97). Time between T1 and T2 was an average of 22.89 days (SD = 1.81), and time between T2 and T3 was an average of 23.55 days (SD = 3.21).

3.1 Communicative Congruence (CC) Scale

From Baseline to T1, the Communicative Congruence Scale showed moderate test-retest reliability (ICC3 = .60). From T1-T2, and T2-T3, the scale showed good (ICC3 = .82) and excellent (ICC3 = .92) reliability, respectively.

Figure 1 CC Scale scores measured across timepoints – Baseline & T1

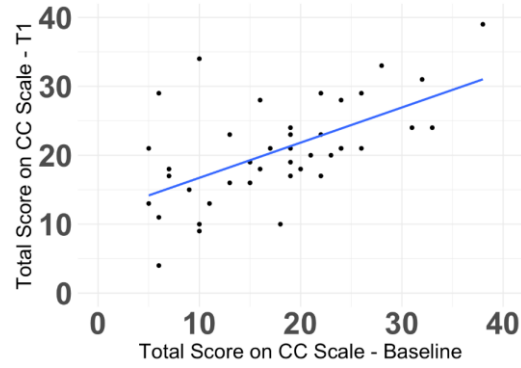


Figure 2 CC Scale scores measured across timepoints – T1 & T2

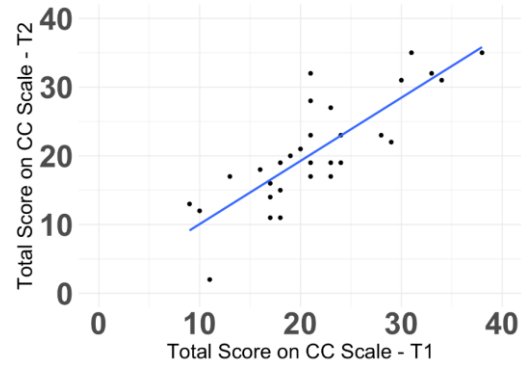
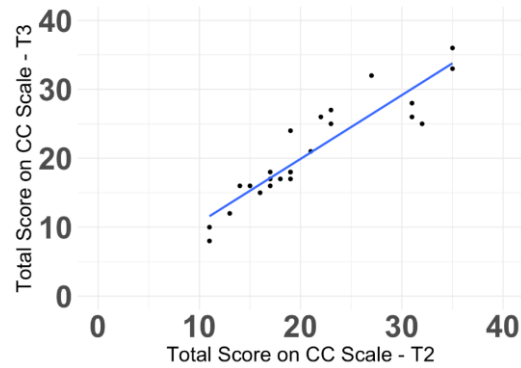


Figure 3 CC Scale scores measured across timepoints – T2 & T3



3.2 Communicative Congruence (CC) Slider

From Baseline to T1, the Communicative Congruence Slider showed moderate test-retest reliability (ICC3 = .50). From T1-T2, the scale showed moderate reliability (ICC3 = .68), as well as from T2-T3 (ICC3 = .74).

Figure 4 CC Slider ratings measured across timepoints – Baseline & T1

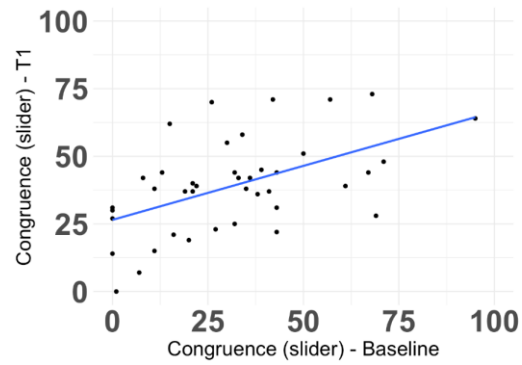


Figure 5 CC Slider ratings measured across timepoints – T1 & T2

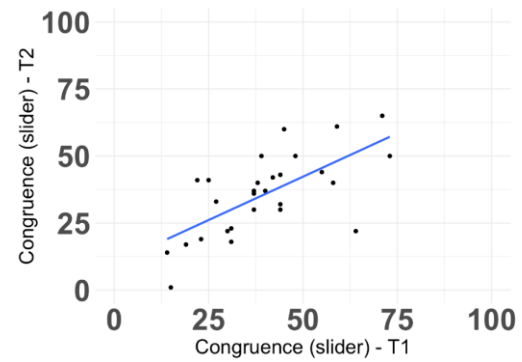
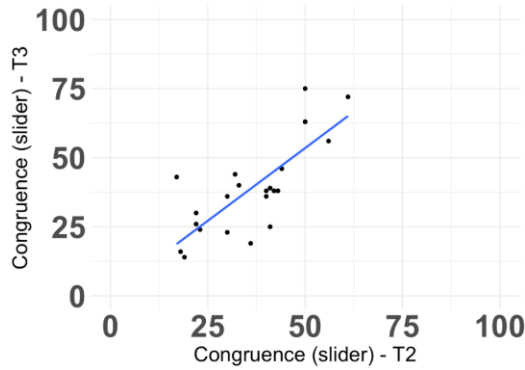


Figure 6 CC Slider ratings measured across timepoints – T2 & T3



3.3 Dysphoria Severity Slider

From Baseline to T1, the Dysphoria Severity Slider showed poor test-retest reliability (ICC3 = .37). From T1-T2, the scale showed moderate reliability (ICC3 = .52), however, it had poor reliability from T2-T3 (ICC3 = .28).

Figure 7 Dysphoria Severity Slider ratings measured across timepoints – Baseline & T1

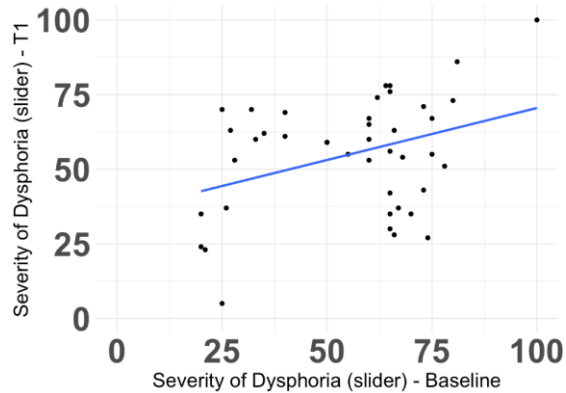


Figure 8 Dysphoria Severity Slider ratings measured across timepoints – T1 & T2

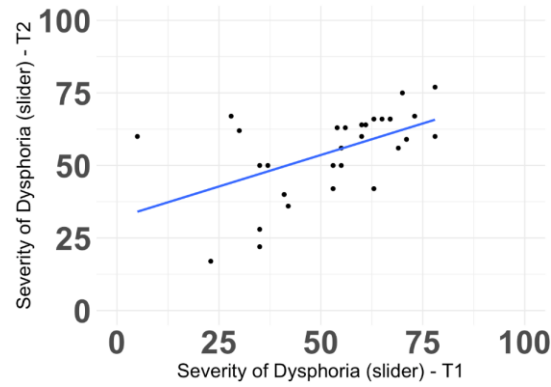
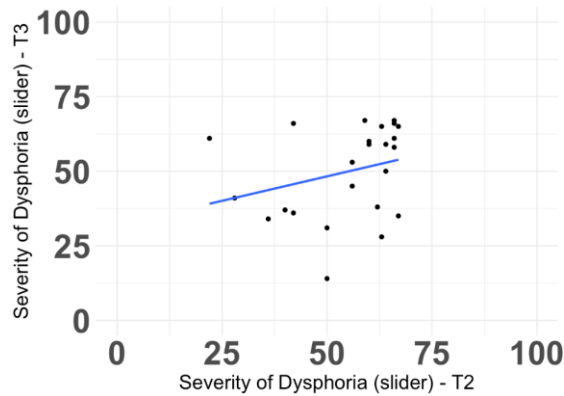


Figure 9 Dysphoria Severity Slider ratings measured across timepoints – T2 & T3



3.4 Disposition Slider

From Baseline to T1, the Disposition Slider showed moderate test-retest reliability (ICC3 = .54). From T1-T2, the scale showed moderate reliability (ICC3 = .71), and good reliability from T2-T3 (ICC3 = .80).

Figure 10 Disposition Slider ratings measured across timepoints – Baseline & T1

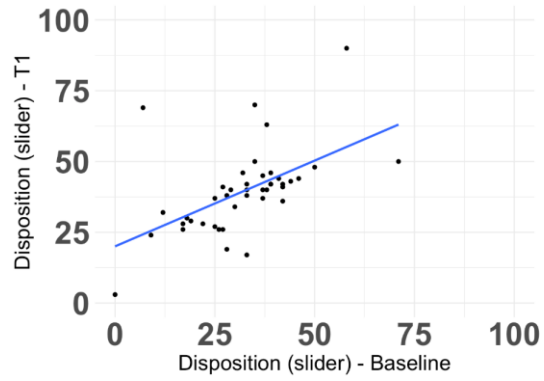


Figure 11 Disposition Slider ratings measured across timepoints –T1 & T2

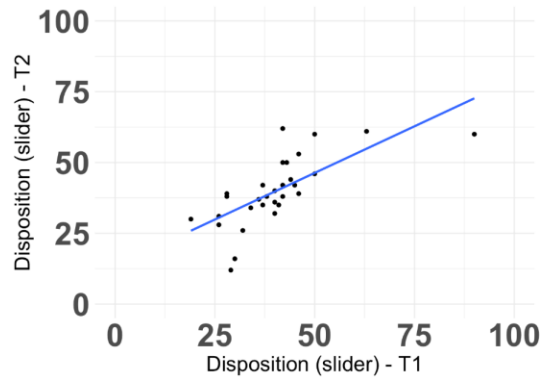
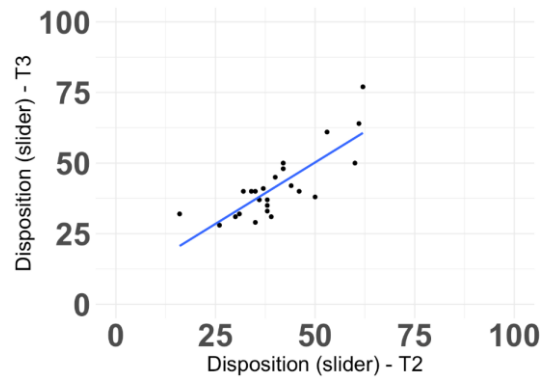


Figure 12 Disposition Slider ratings measured across timepoints – T2 & T3



4.0 Discussion

In this study, we returned to a convenience sample of research participants and asked them to re-rate their experiences of communicative congruence and dysphoria. Self-ratings were made using two different informal measures for each construct. This process was then repeated twice more at 3-week intervals in order to assess more short term test-retest reliability. By measuring these outcomes at both proximal and distal time points, we were able to explore the nature of these constructs as they may pertain to state- and trait levels.

We hypothesized that communicative congruence would demonstrate moderate test-retest reliability because of its theoretical positioning at a high-order level of personality (the meta-trait Plasticity). We also expected that reliability would improve at more proximal time points than at long-term points. In line with our hypothesis, CC showed moderate reliability from Baseline to T1 when measured using the CC Scale and CC Slider. Repeat surveys at three-week intervals (T1-T2 and T2-T3) showed moderate reliability with the CC Slider and good-to-excellent reliability with the eight-item CC Scale. Across all timepoints, the CC Scale was more reliable than the CC Slider.

Although the CC Scale and CC Slider are by design closely related, these measures appear to be assessing slightly different phenomena. Specifically, the CC Slider is singular in its question and concept. By defining congruence in terms of self-identification, it prompts specifically internal consideration about whether the participant's voice and speech accurately reflect who they are. On the other hand, the CC Scale has more breadth and more items. As such, it likely prompts consideration of both internal sense of self as well as identity in relation to others, encompassing feedback and reinforcement of societal roles (e.g., gender, age). For example, questions such as

“People tend to make correct assumptions about me based on how I communicate” and “I feel I have to compensate in other ways to express who I am (e.g., physical presentation), because my communication is not sufficient” both prompt the participant to consider their interaction with a communication partner, and how their communication influences that other person in addition to their own conceptualization of their identity. In contrast, the CC Slider prompt is more general and might not elicit as specific or nuanced a self-reflection as the CC Scale. The CC Scale represents an early attempt to validate a more robust measure of these concepts. Ideally a more developed scale will be inclusive of both internal and external influences and capture how communicative congruence changes across settings, partners, and other such factors.

We also hypothesized that communicative dysphoria would demonstrate low test-retest reliability compared to communicative congruence, again based on the preliminary findings of Welch & Helou (2021a) as well as their theoretical framework (2021b). Specifically, because of dysphoria’s theoretical mapping onto the characteristic adaptations in the CB5T, we expected that it would function more like a state and vary more over time than congruence. However, we again expected measures to be more similar when measured at more proximal time points than at more long-term time points. These hypotheses were generally supported. From Baseline to T1, CD showed poor-to-moderate reliability using the Dysphoria Slider and Disposition Slider. The Dysphoria Slider had moderate reliability from T1 to T2 and poor reliability from T2 to T3. Notably, the Disposition Slider had moderate reliability from T1 to T2 and good reliability from T2 to T3. Across all timepoints, the Disposition Slider was more reliable than the Dysphoria Slider.

Once again, these two measures appear to be assessing slightly different phenomena. The Disposition Slider exists on a scale from –“Hate it so much that it negatively affects my life” to

“Love it so much that it positively impacts my life.” Said another way, this is a spectrum from dysphoria to euphoria, where the middle point should represent relative neutrality. Just as the opposite of *scarcity* is *enough* rather than *plenty*, the opposite of *dysphoria* is not *euphoria* but simply *no dysphoria*. More in line with measuring the dysphoria and its opposite state is the Dysphoria Slider, which is a scale from “None” to “Extreme,” i.e., zero dysphoria to extreme dysphoria. Though this might be more structurally appropriate, removing the possibility of euphoria from the measurement scale will necessarily limit the scale by rendering it incapable of capturing the full breadth of experience in this domain. For instance, a transgender man whose voice deepens due to the initiation of masculinizing hormone replacement therapy might very well move from a state of high communicative dysphoria to one of euphoria. Another possible difference between these two scales returns to the concepts of internal and external calibration. The Disposition Slider specifically refers to impact on one’s own life, whereas the Dysphoria Slider may prompt calibration more in relation to others experience (e.g., “I hate my voice, but not as much as others I know”), or one’s own past experience (e.g., “I still totally hate how I sound and it negatively impacts my life, but hate it less now that my pitch has decreased since starting testosterone.”) Both ways of conceptualizing and measuring communicative dysphoria are valuable, especially when used in tandem.

Overall, the findings indicate that shorter-term measures of both CC and CD will be more reliable than longer-term measures. This is in line with the literature on personality science, in which it is well documented that with increasing time intervals, retest correlations systematically decline, likely owing to personality change over time (Watson, 2004). Additionally, the influence of major life events is more likely over a longer period. For instance, the prior example, initiation of masculinizing HRT for a transgender man, may represent a significant change to vocal quality,

societal role, and self-perceptions, creating a large difference between the distal datapoints that cannot be interpreted as representative of the constructs' behaviors. The proximal time-points minimize likelihood of major life changes, and thus the resulting data is a better indicator of the constructs' behaviors as a state or a trait. Our findings show that CD varies more over time than CC, indicating CD may behave more like a state, and CC, a trait. However, both may function in both a trait- and state-like manner, similarly to other psychological phenomena such as anxiety (Spielberger et al., 2012), in that there exists a relatively stable predisposition to experience CC or CD (evidenced by the findings of moderate test-retest reliability between baseline and T1 for both CC measures and one of the CD measures), along with state experiences related to specific situations, settings, or communication partners.

4.1 Clinical Implications

As previously described, numerous clinical populations might benefit from valid and reliable measurement of CC and CD, such as the trans and gender diverse community, those with voice disorders, and those using alternative or augmentative communication devices. With an improved understanding of how CC and CD function over time, they can more readily be employed in clinical practice.

One such scenario is clinical goal writing. While traditional goal writing for speech and language therapy often centers an acoustic product or functional communication goal, CC and CD can supplement these for improved patient satisfaction. For example, in addition to proficiency using a speech generating device, a goal targeting communicative congruence may seek to find an

appropriate electronic voice output that the patient feels aligned with in terms of age, gender, accent, personality, or other salient factors.

Additionally, baseline measurement of CC and CD may aid in goal prioritization as well as predict prognosis. For some individuals with high CD, acknowledgement and consideration of their experience and emotions surrounding their CD may be a prerequisite for further voice work, and may require adapted treatment approaches. For example, a trans woman whose high communicative dysphoria precludes independent practice may benefit from guidance on auditory masking.

The implementation of CC and CD in clinical practice is highly relevant to mental wellbeing, with Helou & Welch (2022) finding that lower communicative congruence predicted higher scores on the CES-D, indicating higher reports of depression symptoms. With a known link between functional voice disorders and mental health difficulties, including depression symptoms (Aldridge-Waddon, et al. 2023), targeting and monitoring CC and CD serves to treat the whole person as well as their voice complaints.

4.2 Limitations and Future Directions

Although this investigation had a diverse sample in terms of gender, the sample was not sufficiently large to compare results across groups (e.g., cisgender and transgender groups). Future work should compare CC and CD scores and reliability across groups including those pertaining to gender identity, racial/ethnic categories, individuals with and without voice disorders, and other relevant features. Due to the wide variety of potential sources of communicative incongruence,

results may vary significantly. For example, this sample may be legitimately representative of gender-related communicative dysphoria, but may not have the same findings as a sample with greater racial/ethnic diversity. If the sample were more racially and ethnically diverse, experiences of communicative dysphoria may differ as a function of effects relating to accent, dialect, or code-switching as opposed to gender identity.

There are several other considerations regarding the sample selection for future work. For example, it is possible that participants interested in or highly attuned to the concepts being measured were more likely to participate or continue participation for all three surveys. It is unclear how this possible self-selection bias may have affected the findings. Further, repeat administration may have primed participants to consider the constructs (and their experience of them) more deeply in the time between measurements. Another consideration is that the order of item presentation was not varied, so any possible order effect was not mitigated. Future research should vary presentation order and employ other methods to reduce possible effects of repeat administration on the responses.

Further research is also needed to assess the prognostic value of CC and CD for clinical outcomes like those described previously, including voice therapy outcomes, length of treatment needed, and patient satisfaction. Utilizing these measures during intervention that aims to improve CC (e.g., gender affirming voice and communication training) may offer valuable insight into the relationship between existing therapy approaches and these constructs. Comparison against concurrently presented valid and reliable measures (e.g., the TQVQ, or CESD) can be employed to better understand the behavior of CC and CD.

Overall, there is a need to continue developing the measurement approaches to create an inclusive, valid, and reliable clinical tool for measuring CC and CD based on this work and the findings of other ongoing research.

5.0 Conclusions

This project measured the test-retest reliability of the novel constructs communicative congruence and communicative dysphoria, measured in four ways. Findings indicated that communicative congruence and communicative dysphoria will be more reliable when measured in the short-term. In line with our hypotheses, CC has greater test-retest reliability than CD across measurement approaches. Both constructs may function in both a trait- and state-like manner, meaning they can be both highly variable depending on the situation and specific context of an interaction, or more consistently observable as a trait of an individual. Further research is needed to better understand these constructs across diverse populations, as well as their utility across the scope of clinical practice in speech-language pathology.

Bibliography

- Aldridge-Waddon, L., Hiles, C., Spence, V., & Hotton, M. (2023). Clinical Psychology and Voice Disorders: A Meta-Analytic Review of Studies Assessing Psychological Characteristics Across Individuals With and Without Voice Disorders. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2023.09.012>
- Baylor, C., Bamer, A., Brown, C., Jin, J. L., Teixeira, J., & Nuara, M. (n.d.). The Communicative Participation Item Bank–Gender-Diverse Version: Item Bank Calibration and Short Form. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/2023_AJSLP-23-00260
- Craske, M., Wittchen, U., Stein, M., Andrews, G., & Lebeu, R. (2013). Severity Measure for Generalized Anxiety Disorder—Adult.
- Crow, K. M., van Mersbergen, M., & Payne, A. E. (2019). Vocal Congruence: The Voice and the Self Measured by Interoceptive Awareness. *Journal of Voice*, 35(2), 324.e15-324.e28. <https://doi.org/10.1016/j.jvoice.2019.08.027>
- Dacakis, G., Oates, J. M., & Douglas, J. M. (2017). Further Evidence of the Construct Validity of the Transsexual Voice Questionnaire (TVQMtF) Using Principal Components Analysis. *Journal of Voice*, 31(2), 142–148. <https://doi.org/10.1016/j.jvoice.2016.07.001>
- DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, 56, 33–58. <https://doi.org/10.1016/j.jrp.2014.07.004>
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93, 880–896. <https://doi.org/10.1037/0022-3514.93.5.880>
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment*, 18, 192–203. <https://doi.org/10.1037/1040-3590.18.2.192>
- Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Journal of Anxiety Disorders*, 15(3), 231–245. [https://doi.org/10.1016/S0887-6185\(01\)00060-3](https://doi.org/10.1016/S0887-6185(01)00060-3)
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, 7(1), 7–28.
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O’Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The

- REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208. <https://doi.org/10.1016/j.jbi.2019.103208>
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Helou, L. B., Goldberg, A., & Welch, B. M. (2022, June 1). Characterizing vocal and communicative dysphoria in the general population. *The Voice Foundation’s 51st Annual Symposium: Care of the Professional Voice*, Philadelphia PA.
- Horstmann, K. T., & Ziegler, M. (2020). Assessing Personality States: What to Consider when Constructing Personality State Measures. *European Journal of Personality*, 34(6), 1037–1059. <https://doi.org/10.1002/per.2266>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kreukels, B. P. C., Haraldsen, I. R., Cuypere, G. D., Richter-Appelt, H., Gijs, L., & Cohen-Kettenis, P. T. (2012). A European network for the investigation of gender incongruence: The ENIGI initiative. *European Psychiatry*, 27(6), 445–450. <https://doi.org/10.1016/j.eurpsy.2010.04.009>
- Levon, E. (2006). Hearing “gay”: Prosody, interpretation, and the affective judgments of men’s speech. *American Speech*, 81(1), 56–78. <https://doi.org/10.1215/00031283-2006-003>
- Myers, B. R., Pierce, J. L., & Mathy, P. (2024). The Utah Gender Presentation Scale for Communication: Development and Validation of a Measure of Voice-Related Gender Incongruence. *Journal of Speech, Language, and Hearing Research: JSLHR*, 67(2), 415–428. https://doi.org/10.1044/2023_JSLHR-23-00549
- Pernet, C., & Belin, P. (2012). The Role of Pitch and Timbre in Voice Gender Categorization. *Frontiers in Psychology*, 3. <https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00023>
- Pettorino, M., Pellegrino, E., & Maffia, M. (2014). “Young” and “Old” Voice: The prosodic auto-transplantation technique for speaker’s age recognition. *Speech Prosody 2014*, 135–139. <https://doi.org/10.21437/SpeechProsody.2014-15>
- Pisanski, K., Oleszkiewicz, A., & Sorokowska, A. (2016). Can blind persons accurately assess body size from the voice? *Biology Letters*, 12(4), 20160063. <https://doi.org/10.1098/rsbl.2016.0063>
- Podesva, R. J., & Callier, P. (2015). Voice Quality and Identity. *Annual Review of Applied Linguistics*, 35, 173–194. <https://doi.org/10.1017/S0267190514000270>

- Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3), 385–401. <https://doi.org/10.1177/014662167700100306>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3(2), 621–640. <https://doi.org/10.1111/j.1749-818X.2009.00125.x>
- Spielberger, C. D. (2012). State-Trait Anxiety Inventory for Adults [dataset]. <https://doi.org/10.1037/t06496-000>
- Van Hook, E. (2023). The Gender Voice Index. <https://gender-voice-index.company.site/>
- Van Hook, E., & Duffy, S. (2021). Gender Spectrum Voice Inventory: A Pilot Study. *Perspectives of the ASHA Special Interest Groups*, 6(6), 1900–1908. https://doi.org/10.1044/2021_PERSP-21-00076
- Walton, J. H., & Orlikoff, R. F. (1994). Speaker Race Identification From Acoustic Cues in the Vocal Signal. *Journal of Speech, Language, and Hearing Research*, 37(4), 738–745. <https://doi.org/10.1044/jshr.3704.738>
- Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, 38(4), 319–350. <https://doi.org/10.1016/j.jrp.2004.03.001>
- Welch, B., & Helou, L. B. (2021a, October). Preliminary Examination of the Test-Retest Reliability of Communicative Congruence and Communicative Dysphoria. The Fall Voice Conference 2021, Miami, Florida.
- Welch, B., & Helou, L. B. (2021b). Communicative Congruence and Communicative Dysphoria: A Theory of Communication, Personality, and Identity. <https://doi.org/10.20944/preprints202108.0194.v1>.
- Welch, B., & Helou, L. B. (2022). Measuring Communicative Congruence and Communicative Dysphoria in a Sample of Individuals Without Voice Disorders. *Journal of Speech, Language, and Hearing Research*, 65(9), 3420–3437. https://doi.org/10.1044/2022_JSLHR-21-00459
- Welch, B., van, M. M. R., & Helou, L. B. (2021). Listeners' Perceptions of Speaker Personality Traits Based on Speech. *Journal of Speech, Language, and Hearing Research*, 64(12), 4762–4771. https://doi.org/10.1044/2021_JSLHR-20-00582