DEVELOPMENT OF A PEER-PEER INTELLIGIBILITY MEASURE

By

ROSWELL M. SUTHERLAND

(Under the Direction of Hannah Krimm)

ABSTRACT

In this study, we share the development of the Peer-Peer Assessment of Speech Intelligibility (PPASI), a child-friendly short form measure of single-word intelligibility based on the Preschool Speech Intelligibility Measure (PSIM; Morris, Wilcox, & Schooling, 1995) and created to equip future researchers to address the lack of empirical evidence indicating how well children understand their peers with Speech-Sound Disorders (SSDS). In the current study, we investigated the concurrent validity of the PPASI relative to the PSIM. Thirteen adult participants completed the PPASI and PSIM for two speech samples, one taken from a child with a diagnosed SSD and the other identified as having developmentally appropriate speech. Findings revealed a strong correlation between the two measures and evidence of the equivalence of scores obtained using the two measures, suggesting that the PPASI demonstrates concurrent validity with the PSIM. Continued development of the PPASI is warranted based upon the findings of the current study.

INDEX WORDS: Pediatric speech sound disorders, Communication sciences and disorders, Speech intelligibility, Intelligibility assessment

DEVELOPMENT OF A PEER-PEER INTELLIGIBILITY MEASURE

By

ROSWELL M. SUTHERLAND

B.S.Ed., University of Georgia, 2021

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2023

DEVELOPMENT OF A PEER-PEER INTELLIGIBILITY MEASURE

by

ROSWELL M. SUTHERLAND

Major Professor: Hannah Krimm

Committee: Anne Bothe Marcotte
Jennifer Brown

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

Introduction

Speech sound disorders (SSDs) are a collection of impairments that affect an individual's speech production. SSDs are characterized by errors in the production of target speech sounds (i.e., phonemes), and/or phonemic blends, and/or changes to the syllabic structure of a target word, when the speaker's age, language background, and other characteristics make it reasonable to expect a particular production (Namasivayam et al., 2013). Errors occur due to motoric impairments (referred to as articulation-based SSDs), phonological impairments (referred to as phonological SSDs), or both (Bauman-Waengler, 2016). Children with SSDs make up the third largest group of students receiving speech-language intervention services in school settings, with around 92% of school-based speech-language pathologists (SLPs) reporting that they provide SSD-related intervention (American Speech-Language-Hearing Association, 2022). The research conducted for this thesis represented preliminary steps related to determining whether the peers of children with SSDs find them intelligible, in part because of the importance of mutual intelligibility to the development and maintenance of social relationships.

Speech Intelligibility

Speech intelligibility is a perceptual, multifactorial construct. Weismer et al. (2008) proposes that the factors that contribute to speech intelligibility are (a) speech-sound production, (b) the identities of the speaker and listener, (c) what is spoken, and (d) where it is spoken. Because speech intelligibility is multifactorial, one's definition of speech intelligibility depends on which factor they are studying in research or targeting in clinical practice. The study described in this thesis represents the first step of a research line that seeks to investigate the degree to which the peers of children with SSDs find them intelligible. Thus, the theoretical framework of the current study revolves around Weismer et al.'s (2008) second factor: the identities of the speaker and listener. Consistent with this focus, Nicolosi et al. (1996) defined speech intelligibility as a measurement of the degree to which a listener can accurately perceive the speaker's message. We utilize this definition to serve as the foundation of this research project because it emphasizes the identity of the listener rather than other factors such as speech-sound production. The choice to focus on identity of the listener does not imply that this factor is more important or plays a larger role in intelligibility than the other factors; it is merely the factor being presented here.

Speech intelligibility, like speech-sound acquisition, follows a developmental trajectory. Typically developing kindergarten to first grade children (between 5 and 7 years old) demonstrate single-word intelligibility between 75 and 90% as measured by adult listeners (e.g., caregivers, teachers, and SLPs; Hustad et al., 2021). Similar data are available across languages and show that typically developing children produce most phonemes of their language correctly and are largely intelligible to familiar adults by age 4-5 years (McLeod & Crowe, 2020). Speech intelligibility can be used as a proxy to indicate a speaker's communication effectiveness. Monsen et al. (1981) found that intelligibility of at least 60% is necessary for a listener to mostly understand a speaker's message and that less intelligible speech often results in communication breakdowns. It is, therefore, reasonable to assert that when measures of intelligibility indicate that a speaker is less than 60% intelligible to a listener, the speaker may have difficulty in participating in meaningful and effective interpersonal communication with that listener.

Evaluation of Children with Suspected SSDs

When evaluating children with suspected SSDs and completing subsequent intervention planning, SLPs must use a holistic framework. Evaluations intended to describe a child's speech abilities and/or to determine the presence of a SSD combine parent report, professional judgment, formal testing, and informal testing. Evaluation commonly includes assessments of articulation and expressive phonology to assess the direct impact of motoric or phonological impairments on speech production, more specifically the types and frequency of speech sound errors, as well as measures of speech intelligibility to evaluate how these errors impact the speaker's ability to be understood by others. SSD evaluations, therefore, cannot be reduced to only one type of measurement and should instead include speech intelligibility and speech-sound accuracy measures, as both contribute to impaired communication effectiveness and associated impacts to social communication and socio-emotional wellbeing (Namasivayam et al., 2013).

Speech-Sound Accuracy

Assessing speech-sound accuracy commonly involves using measures of articulation and expressive phonology to assess the types and frequency of speech-sound errors; children's performance on these measures often is compared to the performance of their peers (Dodd, 2014; Skahan et al., 2007). Examples of speech-sound accuracy measures include the Goldman-Fristoe Test of Articulation, 3rd edition (GFTA-3;

Goldman et al., 2015) and the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2002). Speech-sound accuracy measures task the speaker with verbally producing a list of predetermined words which contain a variety of phonemes, phoneme blends, and syllable and word structures. Speaker productions are used to generate articulation and phonological inventories. Types and frequency of errors and/or demonstrated lack of acquisition of expected speech sounds can be an indication that the speaker presents with an SSD.

Speech-sound accuracy measures, however, should not be used as the sole source of information for diagnostic decisions or in determining eligibility for special education services in the school setting. Children who do well on such assessments may continue to have difficulties in communicating with their peers, caregivers, teachers, and unfamiliar communication partners (Namasivayam et al., 2013). Psychometric properties (e.g., reliability and validity) of most speech-sound accuracy assessments are somewhat weak (Fabiano-Smith, 2019). Inconsistencies often are attributed to differences across assessment developers regarding (a) their theories of the etiology of SSDs and (b) the frequency and types of errors that they believe should be considered outside of normal limits (Clausen & Fox-Boyer, 2022; Storkel, 2019). The effects of these inconsistencies are exacerbated for children who move between clinicians or school districts, where individual SLPs and school districts use different criteria for evaluating and determining eligibility for children with SSDs (Ireland, 2020).

Speech Intelligibility

Several measures exist to assess speech intelligibility. A clinician's choice of measure will depend on the factor of intelligibility (e.g., speech-sound production, the identities of the speaker and listener, what is spoken, and where it is spoken) that they are most interested in for evaluation purposes. Measures can broadly be categorized as either subjective or objective.

Subjective Intelligibility Measures. Subjective measures tend to be perceptual measures that require the listener to report on the intelligibility of the speaker using a pre-determined rating scale. These measures are subjective in that they require the listener to make arbitrary judgments. Listener effort scales are subjective perceptual measures; these task the listener with reporting the degree of effort they underwent in attempting to understand the speaker. Intelligibility rating scales also are subjective perceptual measures. On intelligibility rating scales, the listener uses a scale to report how intelligible the speaker is. The Intelligibility in Context Scale (ICS; McLeod et al., 2012a) is an evidenced-based intelligibility rating scale with inter-assessment validity. The child's caregiver(s) responds to seven questions that ask about the child's intelligibility with a variety of communication partners. The ICS is widely used by SLPs for its ease of answering, minimal time necessary to complete, availability in over 60 languages, and consistency with other evidence-based SSD assessment measures (McLeod, 2020).

Children with typically developing speech acquisition, while they may still produce speech with errors, should be 75-100% intelligible to their caregivers by around the age of 3;0 (Coplan & Gleason, 1988). Thus, if caregiver responses to the ICS indicate that a child older than three is not generally intelligible to their parents, then this is a strong indicator of a potential SSD. Children over the age of three who are not generally intelligible by their parents as reported on the ICS often do not score within normal limits on articulation assessments such as the Diagnostic Evaluation of Articulation and Phonology (Dodd et al., 2002) and are often consequently diagnosed with SSDs (McLeod et al., 2012b). Additionally, analysis of ICS responses across communication partners suggests that children are most intelligible to their parents followed by immediate family members, then familiar adults (e.g., teachers), and are least intelligible to unfamiliar adults (McLeod, 2015).

Objective Intelligibility Measures. More objective approaches to measuring speech intelligibility often involve tasking listeners with completing word or phrase recognition tasks such as transcribing what they hear or choosing what they heard from a field of options. These measures characterize the speaker's intelligibility as a percentage (e.g., of intelligible words or intelligible utterances).

Transcription. One approach to calculating percent intelligible units is by collecting a speech sample. The SLP who collected the sample using standardized protocols (e.g., SALT; Miller et al., 2012) can transcribe the sample and identify unintelligible units. Alternatively, an unfamiliar listener can also serve as the judge by transcribing and determining which units they perceived as intelligible or unintelligible. Speech sample analysis relies on spontaneous utterances; therefore, samples may vary in length and morphological complexity, both of which may affect the calculation of the child's intelligibility. It therefore becomes more difficult to objectively compare a child's speech intelligibility over time and across evaluators. Additionally, this approach requires substantially more time to complete than other methods because it involves creating a transcription of spontaneous dialogue and then calculating intelligibility.

Choice From a Field. Another approach to calculating percent intelligible utterances is to use standardized words and phrases that the child is asked to produce either independently or through delayed modeling. The Preschool Speech Intelligibility Measure (PSIM; Wilcox et al., 1991) is a multiple-choice intelligibility measure adapted from Yorkston and Beukelman's (1981) Assessment of Intelligibility of Dysarthric Speech (AIDS; Morris et al., 1995). The PSIM consists of 50 sets of 12 words; each 12-word set contains a target word the child is asked to produce as well as 11 foils. To use the PSIM, the child is recorded saying the 50 target words. This recording is later played to a listener who selects what they believe had been the intended word that the child had been instructed to say.

Morris et al. (1995) found the PSIM to be highly correlated with scores from the GFTA ($r = .73$), a finding they interpreted as evidence of the PSIM's concurrent validity. This result suggests that children who receive developmentally low intelligibility ratings on the PSIM, when compared to speech acquisition norms, are highly likely to have misarticulations and/or produce phonological error patterns in their speech production consistent with the diagnostic criteria for a speech-sound disorder. The PSIM's interrater reliability ranges from .92 to .94 and intra-rater reliability is .97-.98 (Morris et al., 1995).

Impact of SSDs on Communication and Social Relationships

Access to Meaningful Communication

Well beyond the details of speech-sound production, SSDs can directly and adversely influence a child's ability to fulfill their communication needs and to develop and maintain interpersonal relationships (Laughton & Hasenstab, 1986; McCormack et al., 2009). For some children with SSDs, this inability to communicate with others can contribute to underdeveloped language skills, as children who are unable to communicate effectively are less able to learn and practice language skills in conversational contexts (Laughton & Hasenstab, 1986). Additionally, without access to effective communication with their communication partners, children with SSDs frequently experience communication breakdowns (Yont et al., 2002). In extreme cases, these communication breakdowns can prevent children with SSDs from adequately participating in meaningful interpersonal relationships through verbal exchange of needs and ideas. McCormack et al. (2010) found that for some children, these communication breakdowns can be so overwhelming that they lead to frustration and undesired behaviors (e.g., tantrums) and that these issues only worsen the communication breakdown, making it more difficult for the listener to understand the child. The same study also found that some children develop maladaptive behaviors (e.g., choosing to no longer speak) when they become frustrated or overwhelmed. These circumstances, however, have been studied primarily in adult-child relationships; there is a scarcity of published research investigating the extent to which these circumstances manifest in peer-peer relationships and therefore should be a continued pursuit in the area of SSDs research.

Social and Emotional Wellbeing

Even when the articulation errors might be described as less severe, children with SSDs often are also more vulnerable to social isolation (McCormack et al., 2009), undeveloped social-communication skills (Farquharson & Boldini, 2018; Hitchcock et al., 2015), and reduced quality of life (Markham et al., 2009). Hadley and Rice (1991) found that children with language disorders and/or speech impairments were twice as likely to be ignored by others which can lead to the outcomes previously described. The Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition (DSM-V; American Psychiatric Association, 2013) notes that an important consideration for the treatment of developmental SSDs is the psychosocial impact of peer bullying and social isolation.

Negative Peer Perceptions. One specific relevant issue emerges from the fact that SSDs in pediatric populations are associated with negative inter-peer perceptions; that is, children with SSDs are perceived more negatively by their peers than children with developmentally appropriate speech are perceived (Crowe Hall, 1991). Specifically, Crowe Hall (1991) found that upper elementary and middle school students expressed more negative attitudes towards peers with articulation errors and more positive attitudes towards peers without articulation errors. These negative perceptions are not limited to older students and have, in fact, been documented among children as young as preschool age (Gertner et al., 1994). Work in other speech and language disorders also demonstrates young children’s negative views of peers with speech and language differences (Blaskova & Gibson, 2022; Gertner et al., 1994), suggesting that communication effectiveness contributes to how young children interact with and think of their peers. For example, Blaskova and Gibson (2022) found that school-aged children with communication difficulties, specifically language disorders, were more likely to have social statuses which the researchers labeled as “rejected” and “neglected.”

Overall, these findings suggest that even young children can distinguish between typical and disordered speech and make negative social decisions based on that information. Current evaluation methods in SSDs, however, focus on adult perceptions and have not explored children's perceptions of speech-sound disorders.

Study Purpose

As described in the previous sections, available methods for assessing the intelligibility of children with SSDs have focused on adults' perceptions and understanding of children's speech, despite some evidence that children make negative social judgments based on their peers' speech and language abilities. Evaluators in the area of SSDs recognize the social importance of peer reactions, but they appear to have assumed that children who are intelligible to adults will be perceived in the same way by their peers. We know of no research that has systematically investigated this belief. It is possible that children's judgments about intelligibility of their peers with SSDs might differ from adults' judgments of intelligibility. Thus, the purposes of this initial project were (a) to develop the Peer-Peer Assessment of Speech Intelligibility (PPASI) as a modified version of the PSIM, and (b) to investigate whether a determination of the degree of concurrent validity of the PPASI relative to the PSIM when completed by adult respondents could be made.

CHAPTER 2

Method

The following protocol was reviewed and approved by Sterling IRB on behalf of the Institutional Review Board (IRB) of the University of Georgia.

Design

This study constitutes the initial stage of development of the Peer-Peer Assessment of Speech Intelligibility (PPASI). We used a one-group within-subjects (repeated measures) experimental design (Orlikoff et al., 2022) to complete preliminary evaluations of the PPASI scores as compared with the Preschool Speech Intelligibility Measure (PSIM; Morris et al., 1995) scores, using a convenience sample of young adult participants as judges for this preliminary work.

Participants

Thirteen currently enrolled students in the Mary Frances Early College of Education who met eligibility criteria participated in the study. Inclusion criteria included being enrolled as a student in the Mary Frances Early College of Education at the University of Georgia and speaking English as a first language. Exclusion criteria included any reported history of developmental disorders, any medical diagnosis that would affect the participant's ability to perceive audio recordings, and any diagnosis that would affect the participant's ability to touch a picture on a tablet among a field of four images (i.e., individuals with hearing, visual, motor, or intellectual impairment were excluded). Eighteen potential participants submitted the informed consent document digitally using Qualtrics and were subsequently provided with a web-based demographic survey (see Appendix A) that addressed each of these areas to confirm eligibility. Thirteen potential participants completed the demographic survey. All individuals who completed the demographic survey met all requirements and were invited to participate; all 13 agreed and completed all study tasks. Demographic data about the 13 participants are provided in Table 1. Note that Participant 1 indicated being a bilingual English-Spanish Speaker.

Table 1

Demographic Data for the Thirteen Study Participants

Participant	Age	Sex	First Language	Current Student Level of Education	Area of Study	Coursework in SSDs
1	21;1	F	English	Undergraduate	CMSD	Yes
2	21;5	F	English	Undergraduate	CMSD	Yes
3	22;10	F	English	Master’s	CMSD	Yes
4	22;11	F	English	Master’s	CMSD	Yes
5	23;11	F	English	Master’s	CMSD	Yes
6	22;5	F	English	Master’s	CMSD	Yes
7	23;4	F	English	Master’s	CMSD	Yes
8	22;10	F	English	Master’s	CMSD	Yes
9	23;3	F	English	Master’s	CMSD	Yes
10	30;6	F	English	Doctoral	EDSE	Unknown
11	21;10	F	English	Undergraduate	CMSD	Yes
12	21;8	F	English	Undergraduate	CMSD	Yes
13	26;11	F	English	Doctoral	CMSD	Yes

Note. CMSD = Communication Sciences and Disorders; EDSE = Special Education

Materials

Preschool Speech Intelligibility Measure

The Preschool Speech Intelligibility Measure (PSIM; Wilcox et al., 1991) is a multiple-choice intelligibility measure adapted from Yorkston and Beukelman's (1981) Assessment of Intelligibility of Dysarthric Speech (AIDS; Morris et al., 1995). As described in Chapter One, the PSIM consists of 50 test sets. Each set contains the target word, which the child is asked to imitate, and 11 foils. To administer the PSIM, the child is audio recorded saying the 50 target words. This recording is then played to a listener, who selects which word in the set they believe had been the intended word the child was tasked with saying. PSIM scores correlate highly with GFTA scores (r = .73) and show high interrater (r = .92 - .94) and intra-rater reliability (r = .97 - .98; Morris et al., 1995).

Peer-Peer Assessment of Speech Intelligibility (PPASI)

The PPASI was designed for this study and as a potential peer-peer measure of speech intelligibility.

Overall Design and Purpose. The PPASI was designed as an adapted version of the PSIM so that the two measures could be used together in the evaluation of children with a suspected speech-sound disorder.

Development. The following procedures were implemented in the development of the PPASI.

Initial Pool of Possible Test Items. To maintain the desired consistency between the PSIM and the PPASI, words from the 50 test sets from the PSIM constituted the initial pool of possible target words for the PPASI. These 600 words were subjected to a series of inclusion criteria before being selected for use on the PPASI. The series of inclusion criteria identified which words from the PSIM (a) are expected to be within the receptive lexicon of kindergarten and first-grade children and (b) could be easily represented with an image.

Word Appropriateness. Zeno et al.'s (1995) *The Educator's Word Frequency Guide* served to determine which words from the PSIM are consistent with the lexicon of the target age range. The guide provides a detailed index of word frequencies in children's literature. Each word in the guide is accompanied by a prevalence per million words. All 600 words on the PSIM were compared against the word frequency guide. Words that were identified as high frequency (defined as greater than three presentations per million words) for children in kindergarten and first grade met inclusion criteria; all others were excluded. Completion of this step resulted in 371 remaining potential target words for the PPASI.

Image Representability. To account for the limited reading abilities of children who would complete the measure, words on the PPASI were to be presented as images. Thus, words from the PSIM that met the Word Appropriateness criterion were subsequently characterized by the author as being either easily representable with an image, or not easily representable by an image, from the perspective of school-aged children with typical language. At this stage, 291 words were characterized as easily representable and progressed to the next stage of inclusion criteria; all others were excluded.

Test Set Inclusion Criteria. All previous stages of inclusion criteria critically analyzed the PSIM's 600 individual words, which are grouped in sets of 12 to form 50 test sets. To ensure measure efficacy and comparability to the PSIM while also creating a simpler task, the PPASI reduces the number of foils but does not otherwise alter test sets from the PSIM. Thus, all PSIM test sets were treated as integral units so that a word from one PSIM test set could not be grouped with words from another PSIM test set when presented on the PPASI. After the Word Appropriateness and Image Representability criteria had been applied, 42 PSIM sets had four or more words remaining, and 8 PSIM sets had fewer than four words remaining. The sets of test items in which fewer than four words met both word-level inclusion criteria were excluded. At this stage of development, therefore, a pool of 42 potential test sets for the PPASI, each with at least 4 words, had been identified.

PPASI Target and Foil Word Selection and Final Test Set Selection. A review of each of the 42 potential test sets remaining from the PSIM was completed to identify phonetic features and possible speech-sound errors that could alter the intelligibility of the word production. Target words and foils for the PPASI were then manually selected by the research team to represent common speech-sound errors. Each final test set included one target word and three foils, with each foil differing in one speech sound from the target word and from the other foils. For test sets that had more than four words remaining, those that represented more common articulation or phonological errors were selected to be either the target word or one of the three foils for that test set. Twenty of the 42 remaining test sets were manually chosen by the research team.

Administration. The PPASI is administered as a digital, receptive task on a tablet using Microsoft PowerPoint software. It consists of two practice items followed by 20 test items. On each item, the experimenter presents participants with four

images that represent the target word and three phonologically similar foils presented in a

2x2 grid. The experimenter points at and verbally labels each image to ensure that the participants know each of the possible responses. Then, participants hear a single word from an audio recording and touch the image they believe had been the intended word the child was asked to say. Participants hear the child production of the target word only once and are encouraged to guess if they are unsure. The examiner records the participants' responses on a paper test form.

Audio Recording of Target Words

Audio recordings of the PPASI and PSIM target words were obtained from a child with a diagnosed SSD (white, female, age 6 years) and a child with typically developing speech (white, female, age 7 years). The recordings were made in-person by a speech-language pathologist (for the child with a SSD) or by a parent familiar with the child (for the child with typically developing speech). During the recordings, the adult said each target word and the child repeated each word. Because the target words for the PPASI were same target words used for the PSIM, there was no need to record samples for these measures separately. Recordings were edited so that each word could be presented individually to the study participants without the clinician or parent model.

Procedure

Each participant completed all study tasks in a single, one-on-one session with the researcher that lasted about 45 minutes. All sessions were held in the same research lab at the University of Georgia. Participants completed all study tasks at a table with only the necessary materials for the specific task in front of them. The researcher sat perpendicular to the participant at the table and manipulated the audio recordings, provided the participant with the necessary materials (e.g., the PSIM data response sheet only when completing the PSIM), and took any relevant notes during administration.

Task Instructions and Information Provided to Participants

All participants were provided with general verbal instructions prior to beginning the first task. They were provided with an abbreviated form of these instructions prior to each subsequent task; see Appendix B. During the verbal instructions, participants were informed that they would be listening to child speech samples and would complete measures of intelligibility. They were not informed of the number of recordings, that one of the children had been diagnosed as having an SSD, that the child they were about to listen to had or had not been diagnosed as having an SSD, or that they would hear each child twice. Instructions also informed participants that the target word would only be played one time and that they were encouraged to guess if they were unsure of the child's production. Participants were also informed that they could change their answer as long as the next target word had not yet been played.

Technical Specifications

Audio Recordings. Both speech samples were obtained using the audio-recording function on Apple iPhones. Recordings were converted to .MP4 files and edited to remove the parent/SLP model so that only the child productions remained.

Stimuli Playback and Audio Settings. During study tasks, stimuli were presented using the native QuickTime Media Player software on an Apple Macbook Air device. Participants used over-the-ear noise reducing headphones to hear each stimulus word. Audio settings were configured to present stimuli at maximal loudness. Participants were instructed to inform the experimenter if the loudness of the recordings caused any discomfort; however, no participants communicated this during the study, so volume was not reduced for any participant. Conversely, no participants noted that they were unable to hear the presented stimuli.

PPASI Presentation. The PPASI is administered as a digital, receptive task on a tablet using Microsoft PowerPoint software and was presented on an Apple iPad. Each image was obtained using an image search using Google Images, filtered for usage rights. The research team selected appropriate pictures which were determined to accurately represent the words chosen on the PPASI. The research team presented these images to the members of the research lab during a routine lab meeting; all images were agreed to be appropriate by all lab members. All images on the PPASI are in full color and are artistic renditions rather than real images. The Apple iPad was set to maximum brightness prior to administration. A sample of the PPASI visual stimulus for one test set is included as Appendix C.

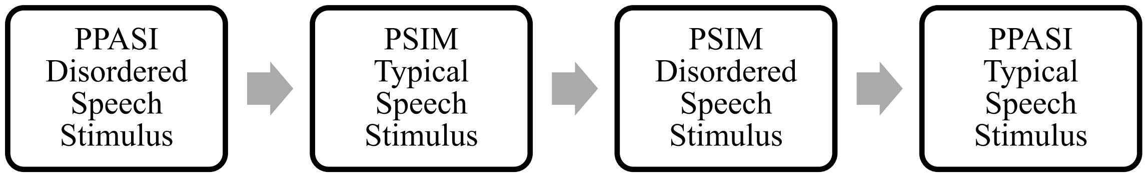
PSIM Response Form. Participants filled out the PSIM response form independently during PSIM tasks. The response form was a single front and back page consisting of a table with the PSIM test sets. The full measure can be found within Morris et al. (1995).

Task Completion

Participation in the study included listening to the typical speech and the disordered speech recordings taken from the two children. Participants scored each set of stimuli using the PSIM and the PPASI. The sequence of measure completion and which child recording was presented was counterbalanced to control for order effects. Figure 1 illustrates one possible order of study tasks.

Figure 1

Example Order of Study Tasks



Administration of the PSIM followed the PSIM assessment protocol. The experimenter first read standard instructions to the participant, provided the participant with the response form, and verbally confirmed the participant's understanding of the task. The experimenter then played the audio recordings one stimulus target word at a time of either the disordered or the typical child's productions. The participant listened to each recorded stimulus once and marked their own selection on the test form. Administration of the PSIM took approximately 10-15 minutes each time. The experimenter played the next stimulus target word after the participant had confirmed their selection.

Administration of the PPASI followed a similar routine. On the first attempt at completing the PPASI, the participants attempted two practice items that resemble the test items found on the measure. The audio stimuli used as practice items were presented live by the examiner. Successful completion of the practice items ensured that the participant understood the task. All participants successfully completed both practice items. To make their selection, the participant pointed at the one of the four images making up that test set's visual stimulus that they believed had been the intended word the child was tasked with saying. The experimenter then confirmed the participant's selection verbally before marking their response on a paper response form. The PPASI response form is provided as Appendix D. Note, this action was not originally included in the task instructions; inclusion of this action in task instructions is discussed in Chapter 3.

Data Analysis

Pre-analysis

All participants completed all study tasks and demonstrated task competency. There was no indication that any participant responses or data collected were invalid; therefore, all data collected were used for analysis and interpretation. All participant responses and calculated percent-correct scores were entered manually by two members of the research team and were compared for errors. Errors documented were all typographical errors and were corrected during a meeting between the two team members.

Scoring of Individual Measures

Participant responses for each test set were coded by the author as either 0 (incorrect) or 1 (correct), where correct indicates that the participant selected the target word (PSIM) or image (PPASI) the child had been instructed to produce (i.e., the adult form of the word). Correct selections were totaled and calculated as a percent-correct score that was interpreted as an estimate of the participant’s ability to correctly identify the word the child had been asked to say and also as an estimate of the child’s speech intelligibility.

Correlations Between Measures

To evaluate relative linear relationships between scores obtained using the two measures, listeners' percent correct scores from the two measures were compared using Pearson *r* correlations. Correlations were completed for all data combined and also separately for data from each child.

Equivalency Testing

The Two One-Sided t-Tests (TOST; Schuirmann, 1987) procedure was implemented, using *t* tests for dependent samples, to evaluate absolute similarities and differences between scores obtained using the two measures. Because many speech intelligibility rating scales use ranges of 10 percentage points to descriptively note the speaker's intelligibility (Monsen, 1981), score bounds for the TOST were first established at $\pm$ 5 percentage points (so scores would fall within a range of 10 percentage points; $\alpha = 0.1$). As further described in Chapter 3, bounds of $\pm$ 6, 7, and 8 points $\alpha = 0.1$ were used for additional analyses, given the structure of obtained data PPASI percent-correct scores were then interpreted as estimates of the children's intelligibility

Severity Ratings

To do so, percent-correct scores were converted to severity ratings using Monson's (1981) intelligibility scale. Analysis aimed at investigating whether the two measures would have classified the children's severity differently. These analyses were done descriptively by hand based on Monson's (1981) definitions: profound, less than 60% intelligible; severe, 60-69% intelligible; moderate, 70-79% intelligible; mild, 80-89% intelligible; and typical, 90-100% intelligible. These labels were applied to scores to determine if participant responses on both measures would place the same child in the same level of the scale.

Item Analysis

Simple item analysis was completed by visually searching for general patterns in participant responses. The scope of this thesis was primarily focused on comparison of binary accuracy scores; therefore, statistical item analysis was not completed. Statistical item analysis will be completed and presented in follow up publications.

CHAPTER 3

Results

The purposes of this study were to (a) develop the PPASI as a modified version of the PSIM, and then (b) to investigate the degree of concurrent validity of the PPASI relative to the PSIM. In this chapter, we present both statistical and observational findings, providing brief conclusions which are expanded further in Chapter 4.

Purpose 1: PPASI Development

Development of this first version of the PPASI was successful. Inclusion criteria based on the age appropriateness and image representability of words from the PSIM resulted in the identification of 42 possible sets of at least four phonetically related words. The final selection of 20 word sets, with each set including four phonetically related words, was successfully used to create 20 sets of four pictures. Recorded productions were gathered from two children, one with a known SSD and one showing typical speech development. Thirteen adult listeners then completed the entire judgment task successfully for both children.

Analysis of the instrument itself, including the instructions developed and the task as administered and as originally envisaged, also identified several issues for future continued development. First, we identified during the data collection process that the word “crawl” appeared twice on the measure—an oversight when creating the test sets. Additional review of the instrument or obtained data will be used to determine from which of the two test sets the word should be removed and to identify an appropriate replacement.

Second, we realized during the conduct of this study that the instructions for listeners, as currently written, may not be appropriate for the school-aged children with whom we plan to implement the PPASI in the future. Therefore, a new, more child-appropriate instructional script must be developed before moving forward in future studies.

Finally, the PPASI visual stimulus was originally intended to serve also as the participants' response form. In that way, it was intended to be navigated independently by the participant (i.e., the software would automatically register the screen location of the participant's tactile input as being the participant's response and would automatically present the next visual stimulus following their selection). For the current study, this automation was not implemented. Rather, the participant would still use tactile direct selection, but their selection was not automatically recorded. During testing sessions, the experimenter verbally confirmed the participant's selection before marking it on a separate paper response form completed by the experimenter. Participants were not informed that the experimenter would do this during the instructional script. This action did not appear to influence any of the participant responses, as no participant changed their selection following the verbal prompt. If this method of data collection is to be continued, mention of this verbal prompt will need to be added to the current script to avoid influencing future participants' responses. However, if response automation is implemented, then this concern will be naturally eliminated.

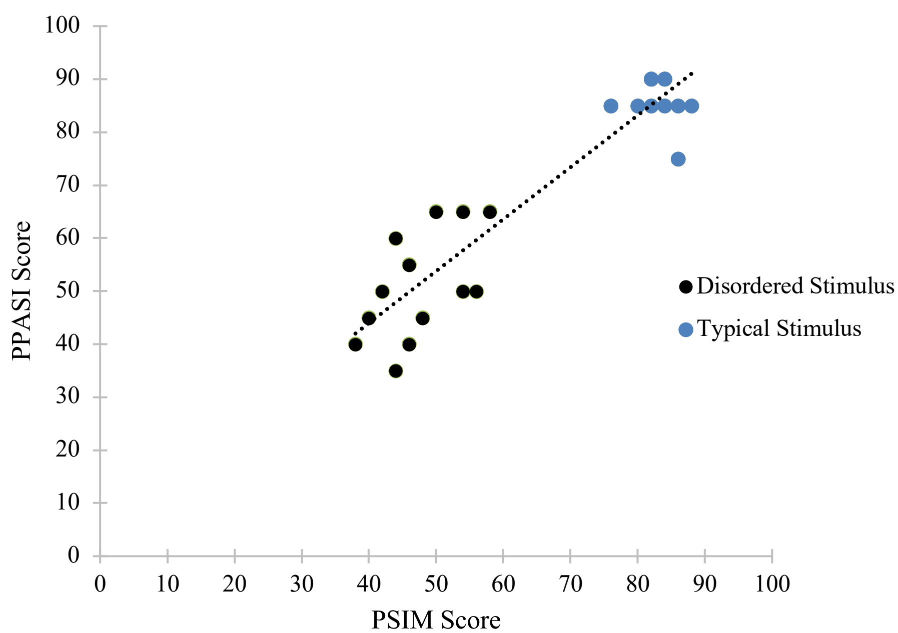
Purpose 2: Concurrent Validity Explorations

Individual Judgment Data

Figure 2 provides a visual representation of data collected from all participants for the recordings from both children using both measures, the PPASI and the PSIM. PPASI scores showed a general similarity to PSIM scores, and scores from both measures appear to distinguish between the child with the known SSD and the typically developing child.

Figure 2

Participants' Percent-Correct Scores on the Peer-Peer Assessment of Speech Intelligibility (PPASI) and the Preschool Intelligibility Measure (PSIM), for two children (Disordered Speech or Typical Speech)



Correlation Analysis

Upon visual analysis of the data scatter plot shown in Figure 2, we determined that correlation analysis using Pearson's *r* was appropriate given the apparently linear relationship between the two measures for the complete data. For all data combined, a correlation coefficient of $r = .934$ ($p < .001$; $a = .05$) was obtained, suggesting a strong correlation between the PPASI and PSIM scores.

We then opted to do an exploratory follow up to evaluate the strength of the correlation between the two measures for the two children separately. A reduction in strength of the correlations when separate is theoretically expected due to the limited quantity of data points associated with each child (13). When isolated, the correlation between the two measures was $r = .573$ ($p < .05$) for the child with disordered speech, suggesting a moderate, statistically significant correlation. A correlation coefficient of $r = -.204$ (p > .05) was observed between the two measures when completed for the child with typical speech. We acknowledge that visual analysis of Figure 2 would normally contraindicate calculation of a linear correlation coefficient for the typical child's speech, as it does not appear to be a linear relationship. However, we deemed it appropriate to do so that it can be compared to future data and because it provides evidence for a ceiling effect that is explored in depth within the Discussion chapter.

Analysis of Group Mean Scores

We next combined all judges' percent-correct scores and calculated group mean percent correct ratings from the two measures; data are presented in Table 2.

Table 2

Participants' Group Mean Percent-Correct Scores (With Standard Deviation and Range) From Two Measures, the Peer-Peer Assessment of Speech Intelligibility (PPASI) and the Preschool Intelligibility Measure (PSIM), for two children (Disordered Speech or Typical Speech)

Child	Measure	Mean	SD	Range
Disordered	PPASI	51.15	10.24	35-65
	PSIM	47.69	6.32	38-58
Typical	PPASI	86.54	4.28	75-90
	PSIM	83.08	3.01	76-86

These averaged data show that the mean percent correct scores between measures are relatively similar for each child, but with greater variability in PPASI percent-correct scores than in PSIM percent-correct scores for both children. Additionally, comparison of ranges and standard deviations reveals that for both measures, there is greater variability in percent-correct scores for the child with disordered speech than for the child with typical speech.

Equivalency Testing

Equivalency testing was completed using the Two One-Sided t-Tests Procedure (TOST; Schuirmann, 1987). The TOST procedure was implemented as an exploratory analysis to examine the bounds within which scores on the PPASI could be considered statistically equivalent to scores on the PSIM. As shown at the top of Table 3, the PPASI and PSIM scores were not statistically equivalent for the child with disordered speech ($p > .05$) or for the child with typical speech ($p > .05$) using score bounds of $\pm$ 5 percentage points.

Table 3

Equivalence Testing of Percent-Correct Scores Between the Peer-Peer Assessment of Speech Intelligibility (PPASI) and the Preschool Speech Intelligibility Measure (PSIM) Using the Two One-Sided T-tests (TOST) Procedure

Score Equivalence Bound	Child	Cohen's dz Equivalency Bounds	Cohen's dz Effect Size	p
(-5, 5)	Disordered	(-0.593, 0.593)	0.411	.261
	Typical	(-0.876, 0.876)	0.606	.175
(-6, 6)	Disordered	(-0.712, 0.712)	0.411	.149
	Typical	(-1.051, 1.051)	0.606	.067
(-7,7)	Disordered	(-0.831, 0.831)	0.411	.078
	Typical	(-1.226, 1.226)	0.606	.023*
(-8,8)	Disordered	(-.949, .949)	0.411	.038*
	Typical	(-1.401, 1.401)	0.606	.007*

Note. **p* < .05.

For this exploratory analysis, we then widened the equivalency bounds by one percentage point in subsequent calculations, to determine the smallest possible range in which scores showed statistically significant equivalency. At $\pm$ 7-point percent-correct score bounds, responses for the typical speech recordings were statistically equivalent (using the unadjusted alpha level of .05) and at $\pm$ 8-point percent-correct score bounds, responses for the disordered speech recordings were statistically equivalent ($p < .05$). These values demonstrate the bounds within which the PPASI and PSIM demonstrate equivalency.

Comparison of Mathematically Possible Scores to Obtained Scores

We added another exploratory analysis of similarity between PSIM and PPASI scores at this stage of our work, to complement the TOST results and to assist in our interpretation of the score ranges identified as equivalent by the TOST results. Due to the difference in the quantity of test sets on the two measures (20 on the PPASI and 50 on the PSIM), identical scores on the two measures are impossible to obtain for a single child. Each of the 50 items on the PSIM contributes 2 percentage points to the final percent-correct score; all PSIM scores are even whole numbers. On the PPASI, however, each of the 20 items contributes 5 percentage points to the final percent-correct score, and all scores are multiples of 5. For example, although a final percent-correct score of 42% (21/50) is possible on the PSIM, this score could not be replicated on the PPASI. The closest percent-correct scores that could be obtained on the PPASI would be 40% (8/20) or 45% (9/20).

Table 4 addresses this issue by comparing the mathematically calculated predicted scores for all participants on the PPASI, based on their PSIM score, with their obtained PPASI score. As shown in Table 4, the obtained PPASI scores were as close as mathematically possible to the obtained PSIM score for 3 of the 13 participants, for the child with disordered speech, and for 4 different participants, for the child with typically developing speech. An additional 7 (disordered) and 8 (typical) obtained PPASI scores were within one PPASI item of the mathematically closest possible predicted PPASI score. Only four of the 26 comparisons identified PPASI scores that differed by more than one PPASI item from the closest mathematically possible PPASI score (this result occurred for 3 participants, for the child with disordered speech, and for 1 participant, for the child with typically developing speech; see Table 4).

Table 4

Mathematically Possible (Predicted) Percent-Correct Scores for All Participants on the Peer-Peer Assessment of Speech Intelligibility (PPASI) Based on Percent-Correct Score on the Preschool Speech Intelligibility Measure (PSIM), with Obtained PPASI Scores.

Participant	Disordered			Typical		
	PSIM Score	Predicted PPASI Scores	Obtained PPASI Score*	PSIM Score	Predicted PPASI Scores	Obtained PPASI Score*
1	54	50, 55	65	80	80	85
2	42	40, 45	50	82	80, 85	90
3	38	35, 40	**40**	76	75, 80	85
4	58	55, 60	65	88	85, 90	**85**
5	40	40	45	84	80, 85	90
6	54	50, 55	**50**	84	80, 85	90
7	56	55, 60	50	82	80, 85	**85**
8	46	45, 50	40	84	80, 85	**85**
9	50	50	65	84	80, 85	90
10	46	45, 50	55	82	80, 85	90
11	48	45, 50	**45**	82	80, 85	90
12	44	40, 45	35	86	85, 90	75
13	44	40, 45	60	86	85, 90	**85**

Note. *Obtained PPASI scores that were as close to the obtained PSIM score as the mathematical structure of the two instruments allowed are in **bold type**.

Severity Rating

Monsen's (1981) percentage-based intelligibility scale was implemented to determine if the two measures consistently resulted in the same severity rating for each child; findings are displayed in Table 5. Monsen's (1981) severity ratings are as follows: profound, less than 60% intelligible; severe, 60-69% intelligible; moderate, 70-79% intelligible; mild, 80-89% intelligible; and typical, 90-100% intelligible.

Table 5

Consistency of Severity Ratings Based on Percent-Correct Scores from the Peer-Peer Assessment of Speech Intelligibility (PPASI) and the Preschool Speech Intelligibility Measure (PSIM) for Both Children (Disordered and Typical)

Participant	Disordered				Typical			
	PSIM		PPASI		PSIM		PPASI	
	Score	Severity	Score	Severity	Score	Severity	Score	Severity
1	**54**	**Profound**	**65**	**Severe**	80	Mild	85	Mild
2	42	Profound	50	Profound	**82**	**Mild**	**90**	**Typical**
3	38	Profound	40	Profound	**76**	**Moderate**	**85**	**Mild**
4	**58**	**Profound**	**65**	**Severe**	88	Mild	85	Mild
5	40	Profound	45	Profound	**84**	**Mild**	**90**	**Typical**
6	54	Profound	50	Profound	**84**	**Mild**	**90**	**Typical**
7	56	Profound	50	Profound	82	Mild	85	Mild
8	46	Profound	40	Profound	84	Mild	85	Mild
9	**50**	**Profound**	**65**	**Severe**	**84**	**Mild**	**90**	**Typical**
10	46	Profound	55	Profound	**82**	**Mild**	**90**	**Typical**
11	48	Profound	45	Profound	**82**	**Mild**	**90**	**Typical**
12	44	Profound	35	Profound	**86**	**Mild**	**75**	**Moderate**
13	**44**	**Profound**	**60**	**Severe**	86	Mild	85	Mild

Note. **Bolded** items show discrepancies in severity ratings between the PPASI and PSIM

Ratings between the two measures were consistent in 53.85% of instances (14 of 26 measure pairings). Inconsistencies are presented in bold in Table 5. Note that all severity discrepancies differ by only one level on the severity scale. Additionally, in most discrepancies (7 of 12) the difference was by a factor of one percent (e.g., score on the PPASI of 90% and on the PSIM of 84% where the mild severity range is 80%-89%). The PPASI was biased for both the disordered and typical stimuli, as compared with the PSIM, to result in a score associated with a less severe (i.e., more intelligible) rating on the Monsen (1981) severity scale.

Additional Analyses of Obtained Data

Several analyses of obtained data were completed to complement the study's planned purposes of developing the PPASI instrument and comparing its scores to scores obtained from the PSIM.

PPASI Item Analysis

Simple item analysis was completed by visually searching for general patterns in participant responses. Statistical item analysis will be completed and presented in follow up publications as the aim of this thesis was to analyze binary responses (i.e., correct or incorrect). Some general observations were identified among participant responses. These observations are described in further detail below.

For the child with typical speech, 6 of the 20 test sets had at least one participant respond incorrectly. Of these six instances, three showed a majority of participants responding incorrectly. In all instances, the incorrect participants all selected the same foil. For the recordings from the child with a known SSD, at least one participant responded incorrectly on 18 of the 20 PPASI test sets; of the 18 instances, 13 showed a majority of participants selecting incorrectly. However, unlike for the child with typical speech, where all participants selected the same incorrect foil for all questions with at least one participant responding incorrectly, only 4 of the 18 instances for the child with disordered speech showed all participants selecting the same foil. This finding is somewhat expected, in part because the typical stimuli were from a child who had speech-sound errors that were typical, age-appropriate, and generally infrequent. Their errors, therefore, might be more easily recognizable by participants, which led to consistency in incorrect responses, especially when only one error was present during any one production. Conversely, the disordered stimuli were from a child with frequent speech sound errors, often with multiple errors co-occurring within the same word. When multiple speech errors were present for this child, their production might not have been perceived as matching any of the options on the PPASI. Depending on which error the participant thought was the most salient or most important, they might have selected the PPASI option that was the closest to what they heard. Variability, therefore, could be expected for this child.

This difference in the consistency between participant responses for the typical and disordered speech recordings can also be seen when looking specifically at a given speech sound error. PPASI test sets were designed to demonstrate a variety of phonological processes and articulation errors. One phonological process present on the PPASI is cluster reduction which appears 12 times amongst the PPASI target words selected for this study. In each test set where the target word contained a consonant cluster, one of the foils in that test set was a minimal pair of the target word with this cluster reduced. For example, for the test set containing the target word "slick," the minimal pair with consonant reduction "sick" also appeared as a foil. Of the 12 target words with consonant clusters, the stimuli from the child with typical speech development resulted in greater consistency in correct responses amongst participants with all participants selecting the correct word in 10 of the 12 instances. This could be interpreted as the child likely not presenting with cluster reduction. Amongst the typical stimuli, only two test sets had at least one participant who selected the minimal pair with cluster reduction. In both instances, all participants who selected incorrectly did select the same foil—the foil with cluster reduction. For the disordered stimuli, all participants selected the correct target word in 2 of the 12 instances of cluster reduction; cluster reduction was present in the incorrect foil chosen by all participants in 9 of the 10 remaining instances. This seems to suggest that this child did in fact present with cluster reduction in their speech. Where cluster reduction was present in the incorrect foil chosen, the incorrect participants selected the same foil in 2 of the 9 instances. In the remaining seven instances, participants differed in their selections in that some selected the foil with cluster reduction only, whereas others chose the foil with cluster reduction and another phonological or articulation error. This pattern seems to suggest that participants were sensitive to identifying when cluster reduction had occurred (9 of 10 instances), but that actual participant selections also depended on their perception of other speech errors and how they made a final choice when multiple errors were present in the child's production.

In addition to these observations, we completed a review of all test sets. During this review, we identified that the target word "hide" was not correctly chosen by any participant when completing the PPASI for the child with disordered speech and by only one participant when completing the PPASI for the child with typical speech. For the disordered speech recording, 8 of the 13 participants chose the "high" foil; for the typical speech recording, 12 of the 13 chose "high." It was concluded that this consistent error was likely caused by the two children producing an unreleased final /d/ sound in the word "hide." Note that for many English speakers, this final consonant sound is often unreleased and thus could easily be heard as "high" when the listener is exposed to this word in isolation without conversational context. These findings support replacing this test set.

Analysis by Judge

In reviewing the demographic surveys of the participants, two factors that clearly separate the participants into subgroups were age and current level of education. Of the undergraduate students, all were aged 21 years, and all were students in the Communication Sciences and Disorders undergraduate program - a common undergraduate program for individuals who aim to become speech-language pathologists. All master's students were 22 or 23 years old, students in Communication Sciences and Disorders, and considered to be pre-professional speech-language pathologists. Of the two doctoral students, one was a 26-year-old student in Communication Sciences and Disorders; the other was a 30-year-old student in Special Education. To assess whether these differences may have influenced the participants' responses, we analyzed percent-correct scores and severity ranges by the participants' level of education.

Table 6 shows no clear patterns of association between participant percent-correct scores and their levels of education. This stability of scores across participant education levels is also apparent in Table 7, which provides participant group mean scores by level of education.

Table 6

Participant Percent-Correct Scores and Associated Severity Rating on the Peer-Peer Assessment of Speech Intelligibility (PPASI) and the Preschool Speech Intelligibility Measure (PSIM) by Current Level of Education

Current Level of Education	Disordered				Typical			
	PSIM		PPASI		PSIM		PPASI	
	Score	Severity	Score	Severity	Score	Severity	Score	Severity
Doctoral Student	46	Profound	55	Profound	**82**	**Mild**	**90**	**Typical**
	44	**Profound**	**60**	**Severe**	86	Mild	85	Mild
Master's Student	38	Profound	40	Profound	**76**	**Moderate**	**85**	**Mild**
	58	**Profound**	**65**	**Severe**	88	Mild	85	Mild
	40	Profound	45	Profound	**84**	**Mild**	**90**	**Typical**
	54	Profound	50	Profound	**84**	**Mild**	**90**	**Typical**
	56	Profound	50	Profound	82	Mild	85	Mild
	46	Profound	40	Profound	84	Mild	85	Mild
	50	**Profound**	**65**	**Severe**	**84**	**Mild**	**90**	**Typical**
Under-graduate Student	**54**	**Profound**	**65**	**Severe**	80	Mild	85	Mild
	42	Profound	50	Profound	**82**	**Mild**	**90**	**Typical**
	48	Profound	45	Profound	**82**	**Mild**	**90**	**Typical**
	44	Profound	35	Profound	**86**	**Mild**	**75**	**Moderate**

Table 7

Group Mean Percent-Correct Scores on the Peer-Peer Assessment of Speech Intelligibility (PPASI) and the Preschool Speech Intelligibility Measure (PSIM) and Associated Severity Ratings by Current Level of Education

Current Level of Education	Disordered				Typical			
	PSIM		PPASI		PSIM		PPASI	
	Mean Score	Severity	Mean Score	Severity	Mean Score	Severity	Mean Score	Severity
Doctoral Students	45	Profound	58	Profound	84	Mild	88	Mild
Master's Students	49	Profound	51	Profound	83	Mild	87	Mild
Under-graduate Students	47	Profound	49	Profound	83	Mild	85	Mild

As shown in Tables 6 and 7, percent-correct scores and associated severity ratings did not differ consistently by the participants' current level of education. Interestingly, the undergraduate participants (*n* = 4) showed the greatest consistency across the two measures for the typical speech recording, and their ratings also showed the same small difference between measures for the disordered speech recording as the master's students (n=7). The undergraduates' mean ratings were also the lowest of the three means for all four comparisons (Disordered and Typical child on the PSIM and on the PPASI), but the differences between the experience groups were small and the doctoral students did not consistently provide the highest scores. The PhD (n=2) students did show the largest differences between the two measures for the disordered speech recording, but they had the same difference between means for the typical speech recording as the master's students. We acknowledge that the small sample size of each of these groups prevents formulating conclusions based on level of education, and we revisit this question in the Discussion.

No other exploration into participant demographic data revealed any remarkable findings or patterns among participant responses. Direct experience and interactions with children with SSDs might reveal additional information pertaining to participant responses; this information was not obtained from the demographic survey completed by this study's participants but may be an important source of information in subsequent studies associated with this measure.

CHAPTER 4

Discussion

In this study, we designed the Peer-Peer Assessment of Speech Intelligibility (PPASI) as a modified version of the Preschool Speech Intelligibility Measure (PSIM: Morris et al., 1995). The purpose of the PPASI is to assist in the evaluation and subsequent treatment of children with speech sound disorders. While speech sound accuracy assessments provide objective data regarding a child's speech sound inventory, they are limited in evaluating the functional impact a speech-sound disorder may have on the child's everyday communication with others. The communication effectiveness between two school-aged peers when one student has a SSD appears to be a gap in our current literature. A novel assessment measure is necessary to investigate how well children understand their peers with SSDs. We propose the PPASI as a potential measure for investigation in this area. Though several independent studies are necessary to fully evaluate the PPASI's psychometric properties and subsequently characterize the PPASI's level of utility in clinical practice, the current study acts as a first step towards doing so.

For the current study, our major purposes were to create the PPASI and to then explore the PPASI's degree of concurrent validity with the PSIM. We assert that concurrent validity between the two measures could be established if (a) the PPASI results in similar findings as the PSIM or if (b) the participant responses on the PPASI show a similar rating pattern or correlation with participant responses on the PSIM. Results from 13 adult participants provided initial support for both of these types of validity evidence. Scores on the PPASI were highly correlated with scores on the PSIM. Equivalency testing using the TOST procedure (Schuirmann, 1987) revealed that participant responses were statistically equivalent within 16-point bands. Additionally, slightly more than half of pairwise comparisons were within the same fixed 10-point severity ratings bands, and 22 of 26 pairwise comparisons of percent-correct scores showed that the obtained PPASI score was within one PPASI item of the mathematically closest-possible PPASI score as predicted from a PSIM score. These results and some possible implications are addressed in this Discussion.

Study Findings

Results from the study provide preliminary support for concurrent validity between the PPASI and the PSIM as evidenced by a strong correlation between participant percent-correct scores on the two measures as well as findings from other secondary analyses.

Correlation Analysis

The correlation analysis revealed a strong correlation between the two measures when data from both speech recordings were combined. This correlation provides preliminary evidence of concurrent validity of the PPASI relative to the PSIM across the range of speech intelligibility. The strength of this relationship reduced substantially when the data between the two measures was separated by the stimuli (i.e., disordered and typical), in part because of a ceiling effect observed for the child with typical speech. Participant demographics, the nature of the disordered speech sample, and plateau effects are potential reasons for this reduction in correlation strength. We elaborate upon these potential explanations in the following sections.

Equivalency

Equivalency testing using the TOST procedure (Schuirmann, 1987) revealed that percent-correct scores on the PPASI and PSIM were significantly equivalent within ± 7-point score bounds for the typical speech and within ± 8-point percent-correct score bounds for the disordered speech. This result needs to be interpreted in the context of the scoring systems for the two measures, in part because each of the 50 items on the PSIM contributes 2 percentage points to the final percent-correct score, whereas each of the 20 items on the PPASI contributes 5 percentage points to the final percent-correct score.

Consider, for example, a percent-correct score of 85% on the PPASI, resulting from a raw score of 17/20. If that percent-correct score is interpreted as meaning that 85% of items were scored correctly, then we might expect the same participant to receive a PSIM percent-correct score equivalent to 85% of items having been answered correctly on the PSIM. However, 85% of 50 items is not possible for the PSIM; that participant would have to receive a score of either 42/50 (84%) or 43/50 (86%). Similarly, in the other direction, a PSIM percent-correct score of 84% could be interpreted as an expectation for the PPASI that 84% of items should be answered correctly, but 84% of 20 items is not possible; the participant would receive a score of either 16/20 items (80%) or 17/20 items (85%). Thus, the hypothetical child whose articulation abilities could be described as at approximately 85% (i.e., who has a true score of 85%) could receive scores between 80% and 86% on these measures solely as a feature of the number of items on the PSIM and the PPASI.

In this context, we find it to be a strength of the PPASI that 7 of the 26 pairs of scores were identical, and another 15 differed by only one PPASI test item, when comparisons were made between an obtained PPASI score and the PPASI score that could be predicted based on the participant's PSIM score for that child and taking into account the mathematically possible scores on the two instruments; see Table 4.

An initial assertion of general equivalence between the two instruments is also supported by the analyses we completed using Monsen's (1981) 10-point fixed severity bands. In most cases, it is reasonable to assume that the current equivalency of the PPASI to the PSIM (within a 14- to 16-point range, rather than the 10-point range established by Monsen) would not change diagnostic decisions, as shown in Table 5. This conclusion is evidenced by findings which demonstrate that the majority of the participants' percent-correct scores on the two measures placed the given child within the same severity rating band and that discrepancies were minimal, with a change in severity rating by only one percentage point in most instances. Overall, the data obtained in this initial study support a conclusion of a reasonable general equivalency between the two measures.

Comparing Understanding of Disordered and Typical Speech

An observation made in completing analysis of findings suggests that there is more consistency in rating typical speech as opposed to disordered speech. Descriptive statistics revealed a much larger range of percent-correct scores by participants when completing both measures for the child with disordered speech. This observation is consistent with the literature; for example, Hustad et al. (2015) also observed this phenomenon in their study of variability of intelligibility scores in children with dysarthria secondary to cerebral palsy.

Although the cause of this variability is not easily attributable to a singular variable, one potential cause for increased variability is listeners' experience listening to disordered speech. When listening to disordered speech, the listener must make assumptions regarding target speech sounds. Often, functional SSDs, especially those due to a phonological impairment, follow consistent, predictable patterns of substitution, omission, and/or distortion (Potter et al., 2019; American Speech-Language-Hearing Association, n.d.). Prior exposure to disordered speech likely increases a listener's knowledge of these predictable patterns and may increase their ability to understand disordered speech. For example, a judge who is familiar with SSDs might know that when the phonological process of gliding is present in a child's speech, the word "red" might be produced as "wed." That judge may then, upon recognizing that the student has an SSD, assume that the word the child was *attempting* to produce was in fact "red" and was not what was actually produced by the child, in this case "wed."

Analysis by judge on the basis of current level of education did not reveal a clear pattern consistent with this claim. We had originally expected that level of education could be used as a proxy to determine the individual's experience with SSDs, either having one class that covered SSDs at the undergraduate level, having a class that covered SSDs at the graduate level, or having clinical experience at the graduate or post-graduate level. As shown in Table 7, the two doctoral-student participants, with the current highest levels of education, did provide the highest percent-correct scores for the typically developing child (on both measures) and for the child with an SSD on the PPASI, but this pattern did not hold for the PSIM. Because one of the two doctoral-student participants had expertise in special education, however, not in speech-language pathology, it is less clear that experience with children with SSDs was the controlling factor.

Completing a more in-depth evaluation of theoretical knowledge and real experience with SSDs of participants in future studies associated with this project will be a priority. Experience listening to disordered speech and familiarity with the speaker, as discussed, does affect adults' ratings of children's intelligibility (Flipsen Jr, 1995; Connolly 1986; Van Doornik et al., 2018). We expect that these factors also affect peers' understanding of the speech of children with suspected SSDs. Thus, enlisting multiple listeners, including peers, to assess intelligibility of children with suspected SSDs may provide additional information that allows for a more complete understanding of communication ability across contexts and communication partners.

Ceiling Effect

One area of possible concern pertaining to the PPASI relative to the PSIM was an observed plateau of percent-correct scores for the typical speech sample. Analysis of participant responses seen in Figure 2 and Table 3 reveals a ceiling effect of percent-correct scores at 85% to 90% on the PPASI for the typical speech recordings; percent-correct scores from the PSIM show a larger range of 76% to 88%. Although several factors may contribute to this ceiling effect, the analysis to adequately confirm the origin was outside the scope of the current study. However, this ceiling effect likely contributed to reduced correlation coefficient for the typical speech recording. Potential reasons for a plateau effect may include low test-item reliability of one or more test items relative to the final score. Alternatively, one could argue that the plateau is expected because speech intelligibility may follow a logistic relationship rather than a linear one. Data from the study revealed that scores in the disordered condition spanned a range of 20 for the PSIM and 30 for the PPASI. For the typical speech, the range of scores reduced to 10 for the

PSIM and 15 for the PPASI. This observation was also made in analyzing equivalency testing which demonstrated that scores for the typical speech were statistically equivalent within a smaller range than scores for the disordered speech. These findings may suggest that there is for more consistency in rating typical speech compared to disordered speech as previously discussed and as seen in Hustad et al. (2015). For the purposes of developing the PPASI, the observed plateau effect may therefore be a natural phenomenon rather than an error in the PPASI measure.

Implications and Possible Uses of the PPASI

Use of the PPASI as a Short Form

Findings from this study support the potential utility of the PPASI as a short form of the PSIM for adults (e.g., teachers, parents, etc.) to complete when judging children's speech intelligibility. The design of this study specifically evaluated the correlation of scores between the PPASI and PSIM when completed by adult raters. The strong correlation coefficient, the equivalency of the severity ratings generated from PSIM scores and from PPASI scores , and the finding that most PPASI scores were within one item of the closest possible score that could have been obtained given the two instruments' mathematical characteristics all suggest that the PPASI could be used to measure children's overall intelligibility.

We propose, therefore, that the PPASI can be as an informal measure to document subjective interpretations of a child's intelligibility that may have several advantages as compared with the PSIM. Per Morris et al. (1995), the PSIM requires up to 15 minutes to administer and at least 5 to 10 minutes to score. In clinical settings, where other assessments must be completed, administration time is often a contributing factor for which assessments are administered during an evaluation. Thus, while the PSIM may be an effective tool for measuring speech intelligibility, it may be overlooked to make space for other assessments due to time constraints. The PPASI can be administered in substantially less time than the PSIM because it has fewer items (i.e., 20 compared to 50) and fewer response choices for each item (i.e., 4 compared to 12). Additionally, although in this study we used manual scoring of the PPASI, its digital nature makes it a candidate for automatic scoring which reduces administration time and mitigates a potential source of error.

Treatment Programming

In its present version and based upon the findings of the current study, the PPASI may have value in determining whether a child's speech is considered disordered or typical from the perspective of adult raters. However, its primary utility likely lies in its ability to quantifiably describe the effect of SSDs on the intelligibility of a child's speech independent from the constraints of disordered versus typical classification. As previously discussed, speech intelligibility is multifactorial and is determined by both the speaker as well as the listener. The PPASI was designed to allow a wide range of individuals to rate the speech intelligibility of a child. Responses from various listeners could provide critical information in determining when (i.e., with what listeners) inaccurate speech sound production impacts a child's ability to effectively communicate. Thus, we recommend that the PPASI could be used as a supplemental measure of intelligibility for researchers and clinicians who wish to understand the relation between impaired speech and intelligibility across listeners using a standardized tool.

Eligibility Determination

As mentioned throughout this thesis, the PPASI was developed to evaluate how well children without SSDs understand the speech of their peers with SSDs. To receive speech-language intervention services in the public educational setting, a child must present with a speech-language impairment, and that impairment must result in an adverse educational impact. An educational impact includes the child's ability to establish and maintain social relationships with their peers. Currently available measures of speech intelligibility, as discussed in Chapter 1, are exclusively done from the perspective of adult listeners, for example, a speech-language pathologist or the child's classroom teacher. If, based on these measures and formal speech sound accuracy assessments, a child is found to have a speech deficit that does not impact their educational experience, that child might be determined to be ineligible for school-based intervention. However, in evaluating intelligibility simply from an adult perspective, the impact of the SSD on communication with peers does not reach beyond anecdotal teacher reports and the student's self-reports. Until now, a measure did not exist to evaluate if such an impact did exist; however, the PPASI, through this study, has been designed as a measure to directly evaluate this area. Implementing the PPASI in this way might assist in documenting adverse educational impact for children with SSDs eligible for the speech intervention they need to improve their speech intelligibility and subsequently their peer relationships. As discussed in Chapter 1, evidence has consistently shown that children are able to recognize impaired speech production and go on to make social determinations that can lead to social isolation, bullying, and reduced quality of life for the child with an SSD. Effectively measuring the functional impact of reduced speech intelligibility, therefore, should be viewed as a necessary component of speech-language evaluation. In doing so, the effects of impaired intelligibility may be managed or eliminated through intervention that a child might not have previously been deemed eligible to receive. The PPASI has potential clinical utility in assisting practicing school-based speech-language pathologists in doing so.

Future Directions

Further Analysis of Available Data and Additional Data Collection

Future studies could further evaluate test-item reliability on the PPASI to evaluate if all test sets appropriately represent the variability that we see amongst the speech production of children. This would include randomly selecting and testing other target words from each test, because this study utilized the same target words for both children. Additionally, relations between demographic attributes (e.g., age, familiarity with disordered speech, knowledge of speech sound disorders, etc.) and scores on the PPASI should continue to be explored.

As part of the research design for the current study, participants were tasked with completing the PPASI and PSIM for the speech recordings taken from a child with typical speech and a child with disordered speech. Due to constraints in study resources, the stimuli presented were limited to the two samples, one from a child with an SSD and the other from a child with typical speech. This design allowed comparisons to be made directly between all participants, because they were presented with the same stimuli. This decision, however, has limited our ability to adequately evaluate whether the PPASI maintains an overall strong correlation across the range of speech intelligibility. Reduction in correlation was seen when the two children were analyzed independently.

Strong conclusions cannot be made about the validity of the PPASI without additional data based on children's speech across the range of intelligibility.

This study employed a relatively small sample (*N*=13). Findings from this study are promising despite this small sample size, but a larger sample size in subsequent studies would likely provide a more precise estimate of the correlation of scores between the PPASI and the PSIM.

Modifications to the PPASI

As analysis of the current version of the PPASI continues, findings will be instrumental in making necessary modification to the PPASI to develop future versions of the measure to improve its utility in clinical practice and for research. In its present version, some test sets on the PPASI, as discussed in Chapter 3, require replacing. Further, the number of test-items on the PPASI is subject to potential modifications based upon future data analysis, to ensure that a variety of phonological processes and articulation errors are encompassed by the measure so that the PPASI represents the variability of reduced intelligibility amongst children. As the PPASI continues to be developed, subsequent studies will be necessary to continue to evaluate its utility in clinical practice and research and to guide additional modifications before it can be implemented for its original intent: as a tool to evaluate peer-peer intelligibility. These future studies will focus on adult participants until the PPASI can be endorsed as a valid measure based upon strong psychometric properties across all areas.

Future Research

The original intent behind the development of the PPASI was to evaluate how well children understand the speech of their peers with speech sound disorders; this remains its current purpose. However, several steps will be necessary in doing so. Initial implementation of the PPASI with school-aged children will focus primarily on establishing inter- and intra-rater reliability among school-aged listeners. Implementation will concentrate on developing administration protocols, providing of task instructions, and ensuring consistent responses within and between participants. Then, when this process is complete, the measure can be implemented for its original purpose in evaluating how well children understand disordered speech produced by their same-aged peers.

Beyond Assessing Disordered Speech. The PPASI also has potential beyond assessing disordered speech. For example, it could provide insight into speech-sound processing in children and could inform researchers and clinicians on developmental norms for speech-sound discrimination, identification, and meaning association of the speech of adult speakers. Alternatively, the PPASI may also has potential in progress monitoring of intervention for SSDs intervention.

CHAPTER 5

Conclusion

The purpose of this study was to develop and complete initial testing of the Peer-Peer Assessment of Speech Intelligibility (PPASI), with the future goal of implementing the tool to evaluate how well children understand the speech of their peers with speech-sound disorders. The guiding purpose of this study was to determine whether the PPASI demonstrated concurrent validity with the PSIM. Through correlation analysis, equivalency testing, and other secondary analyses the PPASI was shown to have preliminary concurrent validity with the PPASI. The strength of its validity cannot be determined from this study alone, but we contend that the PPASI may be a beneficial addition to currently available speech-intelligibility assessments and may, therefore, have both clinical applications as well as support future research investigation in SSDs. Additional investigation and modifications may be necessary before the PPASI can be endorsed as a valid and reliable measure of peer-peer speech intelligibility. Nonetheless, promising findings from the current study do support the conclusion that this additional development and investigation are warranted and should be continued.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

American Speech-Language-Hearing Association (n.d.) *Speech sound disorders: Articulation and phonology.* Practice Portal. Retrieved February 17, 2022, from www.asha.org/Practice-Portal/Clinical-Topics/Articulation-and-Phonology/.

American Speech-Language-Hearing Association. (2022). *2022 schools survey report: SLP caseload and workload characteristics.* www.asha.org/Research/memberdata/Schools-Survey/

Bauman-Waengler, J. A. (2016). *Articulation and phonology in speech sound disorders: A clinical focus* (5th ed.). [ebook edition]. Pearson Education. https://www.pearson.com/en-us/subject-catalog/p/articulation-and-phonology-in-speech-sound-disorders-a-clinical-focus/P200000000757/9780136631583

Blaskova, L. J., & Gibson, J. L. (2023). Children with language disorder as friends: Interviews with classroom peers to gather their perspectives. *Child Language Teaching and Therapy*, *39*(1), 39–57. https://doi.org/10.1177/02656590221139231

Brunnegård, K., Lohmander, A., & van Doorn, J. (2009). Untrained listeners' ratings of speech disorders in a group with cleft palate: A comparison with speech and language pathologists' ratings. *International Journal of Language and Communication* Disorders, *44*, 656–674. https://doi.org/10.1080/13682820802295203

Cabbage, K. (2019). Speech sound disorders in children. *Seminars in Speech and Language*, *40*(2), 79-80. https://doi.org/10.1055/s-0039-1677758

Cabbage, K. L., Farquharson, K., & Hogan, T. P. (2015). Speech perception and working memory in children with residual speech errors: A case study analysis. *Seminars in Speech and Language*, *36*(4), 234-246. https://doi.org/10.1055/s-0035-1562907

Clausen, M. C., & Fox-Boyer, A. V. (2022). Diagnostic validity, accuracy, and inter-rater reliability of a phonological assessment for Danish-speaking children. *Journal of Communication Disorders*, *95*. https://doi.org/10.1016/j.jcomdis.2021.106168

Coplan, J., & Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics, 82*(3), 447–452.

Crowe Hall, B. J. (1991). Attitudes of fourth and sixth graders toward peers with mild articulation disorders. *Language, Speech, and Hearing Sciences in Schools, 22*(1), 334-340. https://doi.org/10.1044/0161-1461.2201.334

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports*, *1*, 189-196. https://doi.org/10.1007/s40474-014-0017-3

Dodd, B. J., Hua Z., Crosbie ,S, Holm, A., Ozanne, A. (2002) *Diagnostic Evaluation of Articulation and Phonology*. Psychological Corporation.

Ertmer, D. J. (2010). Relationships between speech intelligibility and word articulation scores in children with hearing loss. *Journal of Speech, Language, and Hearing Research*, *53*(5). 1075-1086. https://doi.org/10.1044/1092-4388(2010/09-0250)

Fabiano-Smith, L. (2019). Standardized tests and the diagnosis of speech sound disorders. *Perspectives of the ASHA Special Interests Groups, 4,* 58-66. https://doi.org/10.1044/2018_PERS-SIG1-2018-0018

Farquharson, K., & Boldini, L. (2018). Variability in interpreting "educational performance" for children with speech sound disorders. *Language, Speech, and Hearing Services in Schools*, *49*(4), 938-949. https://doi.org/10.1044/2018_LSHSS-17-0159

Flipsen Jr, P. (1995). Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders*, *28*(1), 3-19. https://doi.org/10.1016/0021-9924(94)00015-r

Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research, 37*(4), 913-923. https://doi.org/10.1044/jshr.3704.913

Goldman, R., Fristoe, M., (2015). *GFTA -3: Goldman Fristoe 3 Test of Articulation*. Pearson Education, Inc., & PsychCorp

Hadley, P. A., & Rice, M. L. (1991). Conversational responsiveness of speech- and language-impaired preschoolers. *Journal of Speech & Hearing Research, 34*(6), 1308–1317. https://doi.org/10.1044/jshr.3406.1308

Hitchcock, E. R., Harel, D., & Byun, T. M. (2015). Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. *Seminars in Speech and Language*, *36*(4). 283-294. https://doi.org/10.1055/s-0035-1562911

Hustad, K. C., Mahr, T. J., Natzke, P., & Rathouz, P. J. (2021). Speech development between 30 and 119 months in typical children: Intelligibility growth curves for single-word and multiword productions. *Journal of Speech, Language, and Hearing Research*, *64*(10), 3707-3719. https://doi.org/10.1044/2021_JSLHR-21-00142

Ireland, M., McLeod, S., Farquharson, K., & Crowe, K. (2020). Evaluating children in U.S. public schools with speech sound disorders: Considering federal and state laws, guidance, and research. *Topics in Language Disorders*, *40*(4), 326. https://doi.org/10.1097/TLD.0000000000000226

Jin, F., Schjølberg, S., Eadie, P., Nes, R. B., & Røysamb, E. (2020). Preschool speech intelligibility and 8-year literacy: A moderated mediation analysis. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3380-3391. https://doi.org/10.1044/2020_JSLHR-19-00394

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology*, *3*(2), 81-95. https://doi.org/10.1044/1058-0360.0302.81

Krueger, B. I. (2019). Eligibility and speech sound disorders: Assessment of social impact. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 85-90. https://doi.org/10.1044/2018_PERS-SIG1-2018-0016

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders*, *35*, 165-188. https://doi.org/10.1080/136828200247133

Laughton, & Hasenstab, M. S. (1986). *The language learning process : Implications for management of disorders.* Aspen Publishers, Inc.

Lynch, J. I., Brookshire, B. L., & Fox, D. R. (1980). *A Parent - child cleft palate curriculum: Developing speech and language.* CC Publications

Markham, C., van Laar, D., Gibbard, D., & Dean, T. (2009). Children with speech, language and communication needs: Their perceptions of their quality of life. *International Journal of Language & Communication Disorders*, *44*(5), 748–768. https://doi.org/10.1080/13682820802359892

McAllister, L., McCormack, J., McLeod, S., & Harrison, L. J. (2011). Expectations and experiences of accessing and participating in services for childhood speech impairment. *International Journal of Speech-Language Pathology*, *13*(3), 251-267. https://doi.org/10.3109/17549507.2011.535565

McCormack, J., Mcleod, S., Mcallister, L., & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology*, *11*(2), 155-170. https://doi.org/10.1080/17549500802676859

McLeod, S. (2020). Intelligibility in Context Scale: Cross-linguistic use, validity, and reliability. *Speech, Language and Hearing*, *23*(1), 9-16. https://doi.org/10.1080/2050571X.2020.1718837

McLeod, S., Harrison, L. J., & McCormack, J. (2012a). *Intelligibility in Context Scale*. Charles Sturt University.

McLeod, S., Harrison, L. J., & McCormack, J.(2012b). Intelligibility in Context Scale: Validity and reliability of a subjective measure. *Journal of Speech, Language, and Hearing Research*, *55*(2), 648–656. https://doi.org/10.1044/1092-4388(2011/10-0130)

Miller, J. F., Andriacchi, K., Nockerts, A., Westerveld, M., & Gillon, G. (2012). *Introduction to LSA using SALT. Assessing language production using SALT software: A clinician's guide to language sample analysis.* SALT Software.

Monsen, R. B. (1981). A usable test for the speech intelligibility of deaf talkers. *American Annals of the Deaf, 156*(7), 845-852.

Monsen, R. B., Moog, J., & Geers, A. (1988). *Picture Speech Intelligibility Evaluation*. Central Institute for the Deaf.

Morris, S. R., Wilcox, K. A., & Schooling, T. L. (1995). The preschool speech intelligibility measure. *American Journal of Speech-Language Pathology*, *4*(4), 22-28. https://doi.org/10.1044/1058-0360.0404.22

Namasivayam, A. K., Pukonen, M., Goshulak, D., Vickie, Y. Y., Kadis, D. S., Kroll, R., & Luc, F. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders*, *46*(3), 264-280. https://doi.org/10.1016/j.jcomdis.2013.02.003

Nicolosi, L., Harryman, E., & Kresheck, J. (1996). *Terminology of communication disorders* (4th ed.). [ebook edition]. Williams & Wilkins.

Orlikoff, R. F., Schiavetti, N. E., & Metz, D. E. (2014). *Evaluating research in communication disorders.* Pearson Higher Ed.

Peterson, H. A., & Marquardt, T. P. (1994). *Articulation testing. Appraisal and diagnosis of speech and language disorders* (3rd ed). [ebook edition]. Prentice-Hall.

Potter, N. L., Nievergelt, Y., & VanDam, M. (2019). Tongue strength in children with and without speech sound disorders. *American Journal of Speech-Language Pathology*, *28*(2), 612-622. https://doi.org/10.1044/2018_AJSLP-18-0023

Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool speech error patterns predict articulation and phonological awareness outcomes in children with histories of speech sound disorders. *American Journal of Speech-Language Pathology*, *22*(2), 173–184. https://doi.org/10.1044/1058-0360(2012/12-0022)

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657–680. https://doi.org/10.1007/BF01068419

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16*(3), 246-259. https://doi.org/10.1044/1058-0360(2007/029)

Storkel, H. L. (2019). Using developmental norms for speech sounds as a means of determining treatment eligibility in schools. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 67-75. https://doi.org/10.1044/2018_PERS-SIG1-2018-0014

Van Doornik, A., Gerrits, E., McLeod, S., & Terband, H. (2018). Impact of communication partner familiarity and speech accuracy on parents' ratings of their child for the Intelligibility in Context Scale: Dutch. *International Journal of Speech-Language Pathology*, *20*(3), 350-360. https://doi.org/10.1080/17549507.2018.1472808

Weismer, G., Ball, M. J., Perkins, M., & Müller, N. (2008). Speech Intelligibility. In, *The handbook of clinical linguistics* (pp. 568-582). Blackwell Publishing. https://doi.org/10.1002/9781444301007

Wilcox, K. A., Schooling, T. L., & Morris, S. R. (1991). *The Preschool Speech Intelligibility Measure (P-SIM).* [paper presentation] Annual Convention of the American Speech-Language-Hearing Association 1991, Atlanta, GA.

Yont, K. M., Hewitt, L. E., & Miccio, A. W. (2002). "What did you say?": Understanding conversational breakdowns in children with speech and language impairments. *Clinical Linguistics & Phonetics*, *16*(4), 265–285. https://doi.org/10.1080/02699200210126523

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. Touchstone Applied Science Associates Inc.

Appendix A: Demographic Survey Completed by Study Participants

Demographic Survey

Section 1: Participant Identity Confirmation

Please enter your unique participant ID code that was provided to you via email.

____

Section 2: Survey Instructions

Please read the following questions carefully and answer to the best of your ability. These questions are optional. You may skip any or all of the questions.

We are asking these questions so that we can describe the study participants when we publish or present the study results.

Section 3: Questions

Q1 Date of Birth:

____

Q2 Sex

○ Male

○ Female

Q3 Race (select all that apply)

○ American Indian or Alaska Native

○ Asian

○ Black or African American

- ◯ Native Hawaiian or other Pacific Islander
- ◯ White
- ◯ Other
- ◯ I prefer not to respond

Q4 Ethnicity

- ◯ Hispanic / Latino
- ◯ NOT Hispanic / Latino

Q5 Highest degree earned

- ◯ Less than high school diploma or GED certificate
- ◯ High school diploma or GED certificate
- ◯ Associate degree or 2-year college degree
- ◯ Bachelor's degree
- ◯ Graduate degree

Q6 Highest degree earned

- ◯ Less than high school diploma or GED certificate
- ◯ High school diploma or GED certificate
- ◯ Associate degree or 2-year college degree
- ◯ Bachelor's degree
- ◯ Graduate degree

Q7 Do you consider your primary language to be English?

- ◯ Yes
- ◯ No

Q8 What do you consider to be your primary language?

____

Q9 Did you ever repeat a grade?

- ◯ Yes
- ◯ No

Q10 Have you ever been diagnosed with any of the following (check all that apply):

- ◯ Autism spectrum disorder
- ◯ Intellectual disability or Cognitive impairment
- ◯ Traumatic brain injury
- ◯ Hearing loss
- ◯ Language impairment
- ◯ Speech impairment
- ◯ ADD or ADHD
- ◯ Other developmental difficulties
- ◯ I have not been diagnosed with one of the above

Appendix B: Verbal Instructions Read to All Participants Prior to Study Tasks

Full Verbal Instructions

I. General Instructions (Read before the first study task. These do not have to be repeated prior to each subsequent task.)

a. Today you will listen to several speech samples taken from children. The children in the recordings were instructed to repeat a word. For each recording, you will choose which word you believe the child was told to say using different measures of intelligibility. I will play each recording only one time, so listen carefully. You can change your answer as long as I have not played the next recording.

II. Before Beginning the Peer-Peer Assessment of Speech Intelligibility (PPASI)

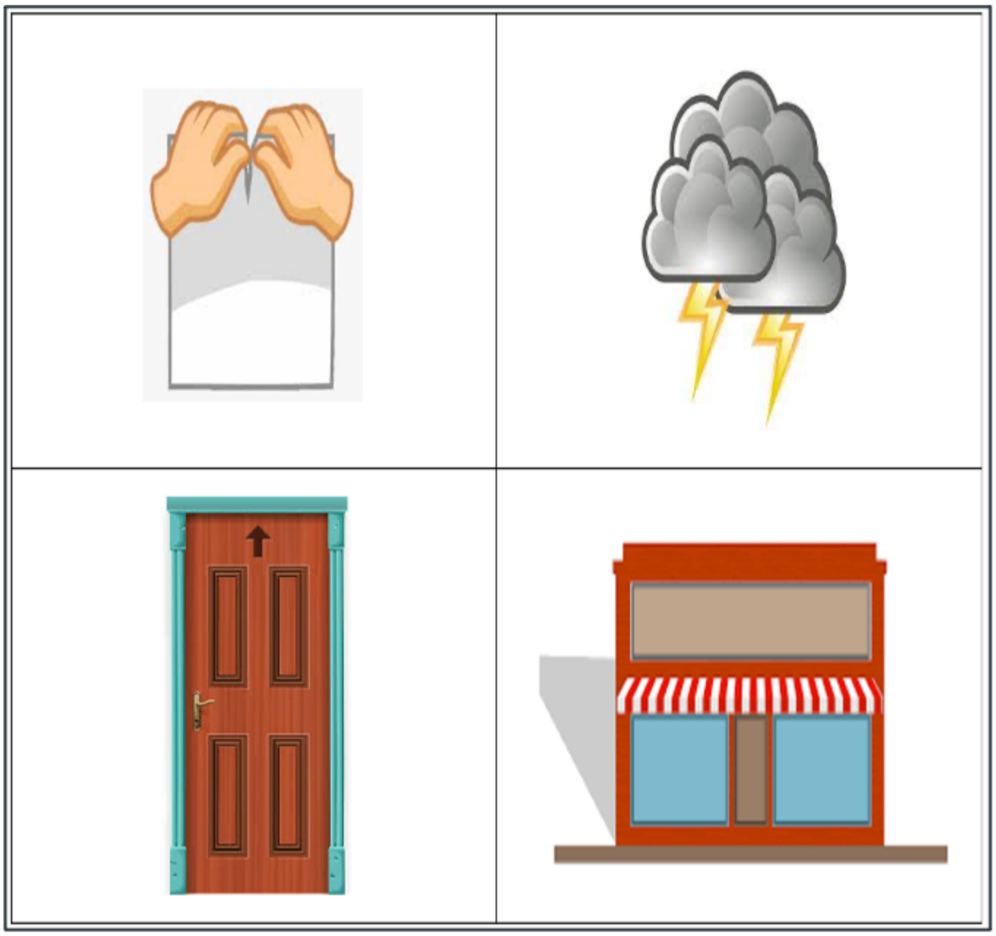
a. You will now listen to a recording of a child saying a list of words. Before you hear the recording, I will tell you what each of the pictures on the screen means like this: [Show the first practice item. Point and label the four items on the screen in the order of 1- Top left; 2- top right; 3- bottom left; 4- bottom right]. You will then hear the recording. Your job is to select the word you believe the child was told to say. If you are unsure, make your best guess. I will play the recording for each word only one time, so listen carefully.
b. If the participant has already completed the PPASI, skip to Part D. If this is the first attempt at completing the PPASI, continue to Part C.
c. Let's try one [examiner provides a live production of the target word for practice item 1]. Which of these four images represents the word that you think I said? [If correct, move on to practice item 2; if incorrect, correct the participant before moving onto practice item 2]. Let's try one more. [present the next practice item]. [If correct, begin the measure; if incorrect, terminate the session].
d. Great job! We will continue doing this a few more times, but now you will make your decisions based on the recording, not me. Are you ready?
e. [Begin measure].

III. Before Beginning the Preschool Speech Intelligibility Measure (PSIM)

a. You will now listen to a recording of a child saying a list of words. Each time, you will pick which word in the list you believe is the word the child was told to say. If you are unsure, make your best guess. I will only play the recording for each word one time, so listen carefully. Are you ready?

b. [Begin measure].

Appendix C: Sample Visual Stimulus for One Test Set of the Peer-Peer Assessment of Speech Intelligibility (PPASI)



1. Top Left: **Torn**
2. Top Right: **Storm**
3. Bottom Left: **Door**
4. Bottom Right: **Store**

Appendix D: Peer-Peer Assessment of Speech Intelligibility (PPASI) Response Form

Completed by the Experimenter for All Participants

PPASI Response Form

_______ Disordered Speech _______ Non-Disordered Speech

Practice Items

1	**Fruit**	Soup	Suit	Root	**1**	2	3	4
2	Bat	Cat	**Rat**	Cow	1	2	**3**	4

Task Items

1	**Cage**	Cape	Tape	Cake	**1**	2	3	4
2	Waste	Wake	Rake	**Raced**	1	2	3	**4**
3	Wall	**Crawl**	Call	Tall	1	**2**	3	4
4	Tear	Hear	**Steer**	Deer	1	2	**3**	4
5	Bark	**Dark**	Park	Art	1	**2**	3	4
6	**Leave**	Lean	Leash	Leap	**1**	2	3	4
7	**Sharp**	Shop	Shark	Shot	**1**	2	3	4
8	Bag	Bat	**Back**	Bank	1	2	**3**	4
9	Dark	heart	**Cart**	Dart	1	2	**3**	4
10	Picture	**Creature**	Pasture	Teacher	1	**2**	3	4
11	Bother	Mother	Butter	**Brother**	1	2	3	**4**
12	Cough	Cloth	**Crawl**	Claw	1	2	**3**	4
13	Torn	**Storm**	Door	Store	1	**2**	3	4
14	Red	Rest	Best	**Dress**	1	2	3	**4**
15	Store	Torch	**Scorch**	Score	1	2	**3**	4
16	High	Eye	**Hide**	Side	1	2	**3**	4
17	Sick	**Slick**	Lit	Lick	1	**2**	3	4
18	Rake	**Train**	Trade	Rain	1	**2**	3	4
19	**Scream**	Stream	Screen	Green	**1**	2	3	4
20	**Spade**	paid	paint	pain	**1**	2	3	4