Survey of Pediatric SLP Methods for Measuring Speech Intelligibility

by

Abigail Laine Bennett

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 6, 2022

Keywords: intelligibility measurement, pediatric speech therapy

Approved by

Mary J. Sandage, Chair, Associate Professor of Speech, Language, and Hearing Sciences
Marisha Speights Atkins, Assistant Professor of Speech, Language, and Learning
Megan-Brette Hamilton, Assistant Professor of Speech, Language, and Hearing Sciences
Allison M. Plumb, Associate Professor of Speech, Language, and Hearing Sciences

Abstract

The purpose of this study was to explore the speech intelligibility measures currently being used by pediatric speech language-pathologists (SLPs), the most frequent measure, the variables considered in intelligibility measurement, and the tools SLPs believe are needed to improve intelligibility measurement. Data were collected through an anonymous 36-question, web-based survey. The survey was distributed via social media and the American Speech-Language-Hearing Association community groups and special interest groups. A total of 140 completed responses were received. Results indicated variation in intelligibility measurements currently in use and in the variables being considered. Many SLPs reported the need for a new or standardized approach to intelligibility measurement. There was disagreement regarding the existence of adequate tools for intelligibility measurement. It can be concluded that there is a need to develop a standard process for measuring speech intelligibility in children as there is currently not agreement among SLPs as to how to measure.

Acknowledgments

There are countless people who have made this thesis possible. Thank you first and foremost to my committee members for agreeing to lend their expertise and time to this study. I would like to thank my mentor, Dr. Marisha Speights Atkins, for the continued guidance, encouragement, positivity, and support over the past few years. Being your student has given me unique and invaluable experiences that extend beyond the classroom. As I transition from being a student to being a clinician, I know the lessons you have taught me have set me up for success. Dr. Mary J. Sandage, thank you for agreeing to take on the role of my thesis chair and for all you have taught me about being a thorough researcher and clinician. Dr. Megan-Brette Hamilton, your dedication to asking the right questions in the right ways ensured we received the most complete and useful responses possible. Dr. Allison Plumb, your knowledge of data analyses translated into results that highlight the need for further research in this area. I also wish to thank my cohort for the words of encouragement, the way you cheered for my accomplishments like they were your own, and all the laughter over the past two years. I would like to thank my parents, Bradley and Delisa Bennett, for the never-ending support and for never doubting me, even when I doubted myself. I would also like to thank my boyfriend, Ford Luscy, for all the reassurance and for always being there for me throughout this process. I would like to give a special thanks to the American Speech-Language-Hearing Association (ASHA) community of members who participated in this study. I would also like to acknowledge the Auburn University College of Liberal Arts and the Department of Speech, Language, and Hearing Sciences support for graduate student research through research stipends that supported this work. Last but not least, thank you to Samantha Kent for generously giving her time to assist with the qualitative data analyses for this survey.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Intelligibility is a complex construct with multiple definitions aiming to capture the multifaceted nature of the phenomenon. Connolly (1986) defines intelligibility as context-dependent in that intelligibility can differ based on the listener's familiarity with the topic, the speaker's voice, and/or the speaker's dialect. Schiavetti (1992) defines speech intelligibility as the match between the intention of the speaker and the response of the listener to the speech passed through the transmission system. Munro and Derwing (1995) and Yorkston et al. (1996) similarly approach speech intelligibility as the extent to which an utterance is understood by the listener. Therefore, a valuable measure of pediatric speech intelligibility should capture: (1) the degree to which the produced acoustic signal is understood by the listener; (2) how well the listener understands the speaker's intended message; (3) the familiarity of the speaker to the listener; and (4) other context-dependent factors.

Speech-language pathologists (SLPs) generally agree that the measurement of speech intelligibility is a critical factor in speech assessment, however, a lack of consensus on how to measure speech intelligibility in the pediatric population remains (Hustad et al., 2019). This paper reviews a variety of methods described in Kent et al. (1994) for measuring speech intelligibility that clinicians can choose from based on the client's needs. It has been suggested that many clinicians commonly make impressionistic statements about intelligibility rather than using other available measurement approaches thought to be more time-consuming (Gordon-Brannan & Hodson, 2000). In the Forum on Phonology in American Journal of Speech Language

Pathology (AJSLP), Bleile (2002) and Khan (2002) provided their perspectives on how they approach assessment. Bleile (2002) stated that the level of intelligibility was observed by listening to clients' spontaneous speech samples. Khan (2002) reported that initial SSD assessment focused on establishing eligibility for speech-language services with additional testing at a later time to create more comprehensive goals. In Skahan et al. (2007), of 312 SLPs surveyed, it was found that 75.4% always subjectively estimate when measuring speech intelligibility. As mentioned previously, a variety of methods exist for measuring intelligibility (Kent et al., 1994), but Skahan and colleagues' investigation did not reveal which methods were being used (Skahan et al., 2007). Farquharson and Tambyraja (2019) survey of SLPs indicated that intelligibility assessment is not required for eligibility determination. This survey did not query clinicians' speech intelligibility assessment practices, therefore it is still unknown which methods were being used (Farquharson & Tambyraja, 2019).

The available tools for speech intelligibility evaluation have predominately served for the evaluation of adult motor speech deficits secondary to dysarthria (Kent et al., 1989; Yorkston et al., 1992). Few measures published for use in the pediatric populations have been standardized and many developed are no longer in print (Allison, 2020). In the absence of standardized measures, word recognition and scaling methods have been used with the pediatric population. The lack of readily available standardized measures of speech intelligibility for pediatric clients has often left clinicians in the position of using their best clinical judgment for diagnosis of speech intelligibility deficits.

The purpose of this paper is to (1) review available methods of speech intelligibility measurement and (2) report survey results to identify what methods are currently being used by SLPs serving the pediatric population.

Rationale

Intelligibility is the most important factor in communication. If a speaker is not intelligible, communication fails (Munro, 2010). While there is not a consensus on how to measure the intelligibility of pediatric speech, SLPs generally agree that measuring intelligibility is a critical factor in assessment (Hustad et al., 2019). In order to be successful when verbally communicating with others, an individual must be intelligible to that person. The central goal for SLPs working with any population is to ensure clients are able to communicate using intelligible speech and listeners are able to understand and correctly decipher what has been said (Miller, 2013). Level of intelligibility impacts multiple aspects of therapy, including need for services, intervention priorities, and evaluation of success of intervention (Skahan et al., 2007).

Intelligibility Measurements: A Review

The primary procedures for measuring speech intelligibility broadly fall into one of two categories: signal dependent or signal independent. The signal dependent measures occur when the listener judges the intelligibility of a speaker based upon the acoustic signal alone (audio recording) without the speaker being present. The listener's judgment of intelligibility is based on only the speech signal itself, using supra-segmental and segmental features, as well as phonotactic predictability and probability. The listener does not have access to context clues (e.g. facial expressions and gestures) available when communicating face to face or when observing a video of the speaker. Syntactic and semantic components are also not considered when judging intelligibility using a signal dependent measure. In contrast, when using a signal independent approach the speaker may be present when the listener is judging intelligibility. The listener has access to the contextual cues during their judgement of intelligibility. Listeners may observe the speaker directly, by viewing a video, or by making judgements based on prior experiences with the speaker.

Intelligibility Measurement Types

Direct approaches for measuring speech intelligibility generally fall into one of three categories: word recognition tasks, scaling tasks, and automated analysis. Each present with different advantages and pitfalls that could help to explain why some methods are chosen over others and the limitations of adopting either as a standard method chosen for speech intelligibility evaluation.

Word Recognition Approaches for Measuring Speech Intelligibility

Word recognition/transcription measures are one of the common measures used for speech intelligibility (Pascoe et al., 2006). Listeners orthographically transcribe or select the word thought heard in a signal dependent listening task. In the open-set variation of word recognition, listeners write down words thought heard (Lousada et al., 2014). A score of intelligibility is calculated by dividing the number of correctly identified words by the total number of target words yielding a percentage of intelligibility.

In the closed-set variation of word recognition, listeners are given a list of similar-sounding words, and they must choose (multiple choice) which word matches what they heard (Lousada et al., 2014). Response options of “other” with a write-in box, a “select all that apply”, and “no response” are provided in the multiple-choice format. Whereas forced choice measures require the listener to choose the word thought heard from a list only.

Open and closed set word recognition tasks have been employed in research and have served as the basis for intelligibility assessment tools. For example, the Test of Children’s Speech (Hodge et al., 2009), has both a word and sentence test, and listeners can measure intelligibility by employing either an open set or closed set measure (Hodge & Gotzke, 2014). Khwaileh and Flipsen (2010) studied the speech intelligibility of single words from *The Children's Speech Intelligibility Measure (CSIM)* and sentences from the *Beginners Intelligibility Test (BIT)* in children with cochlear implants. Judges listened to participants say the CSIM words and orthographically transcribed what they thought they heard (open-set) and also by use of multiple-choice (closed-set). The BIT sentences were orthographically transcribed. Orthographic transcription has been considered the "gold standard" for objective measurement on speech intelligibility and has been widely used in research (Beijer et al., 2012; Hodge & Gotzke, 2014; Khwaileh & Flipsen, 2010; Lagerberg et al., 2019).

Advantages & Pitfalls of Word Recognition Tasks

Word recognition tasks have been described as giving an objective measure of speech intelligibility, making them attractive to clinicians (Allison, 2020). Though this may appear to be the case, whether or not a word is recognized by a listener is influenced by many variables, which can result in poor inter- and intra-rater reliability. These influences include the listener's familiarity with the speaker, disordered speech in general, and with the test material. The extent of each of these influences does not have to be great in order to affect recognition of words (Miller, 2013). Word recognition tasks provide a percentage score which could be used to track change over time. However, these measures are time-consuming to complete. Speech samples must be recorded under ideal conditions and post processing of the recordings required for presentation to the listener. Further it may be assumed that listeners must complete the listening tasks in isolated sound treated rooms and with research grade listening equipment often not available to clinicians practicing outside of university settings. These barriers to implementation may dissuade some from using word recognition tasks in favor of faster methods (Allison, 2020).

Other considerations include the fact that while an individual may be intelligible in a sound-treated room, this method may lead to an overestimation of their intelligibility. In real-world scenarios, the individual is having to communicate with the presence of background noise, competing demands, and movement; similarly, the listener must also overcome these factors in order to decipher the speaker's message. When placed in conditions that are more similar to everyday situations, it is likely the speaker's intelligibility will decrease (Miller, 2013).

Scalar Approaches for Measuring Speech Intelligibility

Clinicians also frequently use rating scales to measure intelligibility because of the ease of administration and the quick results they provide (Miller, 2013). Commonly used rating scale types include ordinal, visual analog, and direct magnitude estimation scales.

Ordinal Scales

Ordinal scales typically employ category hierarchies, and the person completing the scale must choose where the client best fits along the hierarchy. A common example of an ordinal scale is a Likert scale. A typical Likert scale uses the hierarchies such as "strongly disagree, disagree, neutral, agree, strongly agree" for participants to choose from in response to questions or statements (Tastle & Wierman, 2006).

Ordinal scales can be found in current tools used to measure intelligibility. The Viking Speech Scale is a published measure to score the level to which children are understandable to strangers and unfamiliar conversational partners. (Pennington et al., 2011). The scale, designed for children with motor speech disorders secondary to cerebral palsy, is used to classify speech according to four different levels. Level I indicates that the child is not affected by a motor disorder with similarities of same age peers, Level II speech is imprecise but understandable, Level III speech is unclear and not understandable out of context, and Level IV indicates no understandable speech (Pennington et al., 2013). The Intelligibility in Context Scale (ICS) is another ordinal scale, developed to obtain ratings of how intelligible children are to parents and other listeners (McLeod et al., 2012). The ICS is an ordinal scale that can be given to parents to fill out. It utilizes a 5-point Likert scale, where 1 = never, 2 = rarely, 3 = sometimes, 4 = usually, and 5 = always. Parents are asked to fill out seven questions, which query the child's intelligibility to the parents and other communication partners. Hustad et al. (2012), adult participants listened to speech samples of 4-year-olds with and without cerebral palsy and rated how they thought each child would be understood by another adult. The ratings ranged from 1, which was *difficult or impossible to understand*, to 7, which was *very easy to understand.* A third example is the National Outcome Measurement System (NOMS) used by the American Speech-Language-Hearing Association (ASHA), which is a system designed to collect and aggregate outcomes data. NOMS uses Functional Communication Measures (FCMs) for each component or disorder. The FCMs are disorder-specific 7-point scales that enable a clinician to track the individual's progress over time by periodically rating their functional skills. A rating of 1 would be the least functional level, and a rating of 7 would be the most functional level (Mullen & Schooling, 2010).

Visual Analog Scales

Visual analog scales (VAS) involve the listener hearing speech from the speaker and then marking on a scale where they think that speaker's intelligibility would fall (Abur et al., 2019). VAS typically are 100mm in length and anchored on either end with labels or descriptions of the extremes of the characteristic that is being measured (Marsh-Richard et al., 2009). The scale may be marked using a pen-and-paper method or a digitized version of the scale. In the Abur et al. (2019) study, adult listeners judged intelligibility using a sliding scale ranging from 0% to 100%.

In another study by Chiu and Neel (2020), VAS was used to rate individual parameters of intelligibility instead of overall intelligibility. Using a computer, listeners rated ease of understanding, articulatory precision, voice quality, resonance balance, and prosodic adequacy. The overall goal of this study was to see if the ratings of the parameters and ease of understanding from VAS could predict speech intelligibility in quiet and in noise for speakers with Parkinson's disease. In Tjaden et al. (2014), a visual analog scale was used to compare ratings of intelligibility when speakers with multiple sclerosis and Parkinson's disease used the strategies of rate reduction, increased vocal intensity, and clear speech (Tjaden et al., 2014). The study used a vertical, continuous, 150-mm VAS described in Sussman and Tjaden (2012). Listeners used a computer and clicked along the scale where they thought each listener's intelligibility would fall. A computer program then converted the scores to a number between 0 and 1.0, with 0 being *no impairment* to 1.0 being *severely impaired.*

Direct Magnitude Estimation Scales

Direct magnitude estimates (DME) is a third scalar measure. In DME, there is no fixed minimum or maximum value on the scale that readers must stay between. There are two ways to employ DME – with a modulus (DME-M) or without a modulus (DME-WM; (Whitehill et al., 2002). In DME-M, the listener is given a modulus or standard to listen to that is selected by the experimenter. The speech sample can be low, middle, or high intelligibility, but is most often in the middle range. The modulus is assigned a modulus number, most commonly 10 or 100. The listener then rates all future speech samples relative to the modulus sample. In DME-WM, the listener reviews the first speech sample and assigns it any number they would like. They then rate all subsequent speech samples relative to the first speech sample (Schiavetti, 1992).

Kim and Kuo (2012) used DME in a study to determine if presentation level has an effect on the intelligibility of speakers with dysarthria. One experiment looked at DME-WM and one looked at DME-M. It was found that both DME-WM and DME-M ratings were lower for dysarthric speakers than healthy controls. In another study that used DME, researchers posed the questions of how dysarthric speakers rate other dysarthric speakers' speech as compared to speech-language therapists (SLTs), and then how dysarthric speakers rated their own speech as compared to SLTs. Each listener was first given the standard sample to listen to and told that the sample had a rating of 100. Listeners were then instructed to rate future samples based on the standard sample, making this an example of DME-M. The study found that people with dysarthria do rate dysarthric speakers differently than SLTs or naïve listeners. When it came to dysarthric speakers rate their own speech, however, there was no difference (Walshe et al., 2008).

Advantages & Pitfalls of Scalar Measurements

Scalar measures are simple to understand and quick to complete during an evaluation or therapy. Additionally, rating scales can be given to the parent or caregiver to gain their perspective on the child's intelligibility. Though rating scales are quick and easy to use in clinical situations, there are some drawbacks associated with them. Allison (2020) points out that not all listeners subjectively rate intelligibility the same, i.e., one listener's rating of 4 could be the same as another listener's rating of 6. Because of this, the reliability of rating scales can be lower than is clinically acceptable. Rating scales can also be affected by speech characteristics such as hypernasality, making it difficult for the listener to focus and accurately rate intelligibility (Konst et al., 2000).

Published Measures

Published standardized measures are common for many areas of speech-language pathology, such as articulation and language, but they are less common for intelligibility. Current published measures include the Beginner's Intelligibility Test (BIT), the ICS, and the Speech Intelligibility Test (SIT). Though these measures have been applied for research purposes, it is thought that they are not widely used in clinical practice. The BIT was administered by Castellanos et al. (2014) to examine if speech intelligibility and vocabulary skills in preschool-aged children who had cochlear implants could predict long-term speech and language outcomes for that population.

Advantages & Pitfalls of Published Measures

Published standardized measures present the ability to compare children to their age-matched peers. In addition, the fact that the stimulus material is the same allows for direct comparison of results to former studies using the same published measure, as was done in Chin et al. (2012). Administration and scoring of such measures is relatively simple and quick, adding to the appeal. But as Castellanos et al. (2014) points out, the BIT also tests a child's working memory and control processes. The child must remember the sentence the examiner read aloud and repeat it back, which requires the ability to encode the information, store it, and then retrieve it. If a child has impairments in these areas, the BIT is not testing only their intelligibility, but their working memory skills as well. Clinicians must be careful to ensure that they are truly testing intelligibility when administering published measures.

Automated Methods

Acoustic-based measures use parameters such as vowel space, speech rate, voice onset time, speech variability index, speech duration, number of syllables, syllable boundaries, and others to measure a speaker's intelligibility. Lee et al. (2014) used automated measures in their study of intelligibility and acoustic measures in children with cerebral palsy. Temporal, vowel spectral, nasality, and voice measures were the selected speech acoustic variables in the study. Their second research question was how different acoustic measures contribute to intelligibility in children that are diagnosed with cerebral palsy. They found that variables related to articulation have the largest effect on the intelligibility of the child. Another study used a Gaussian mixture model (GMM) to automatically evaluate speech intelligibility and hypernasality in children with cleft palate. Results showed that the algorithm used to evaluate speech gave a good speech classification (Ling et al., 2013).

Advantages & Pitfalls of Automated Methods

Traditional methods of measuring speech intelligibility can be time-consuming, and there is often a demand for fast and reliable assessment when communication disorders are present (Middag et al., 2009). Automated methods present an objective and quick way for clinicians to assess a patient's intelligibility (Van Nuffelen et al., 2009). In addition, automated analysis allows for easy repeatability and has positive implications for rural or remote areas where clinicians may be scarce (Paja & Falk, 2012). Though automated methods have the potential to be useful assessment tools, the final judge of intelligibility in real-world scenarios is the human ear. Therefore, automated methods should work with clinician's subjective ratings and not in place of them (Jiao et al., 2019). Clinicians also are comfortable with subjective and informal measures of intelligibility and value it highly, meaning that objective measures may be slow to be accepted in the field (Jiao et al., 2019; Lee & FitzGibbon, 2008; Miller, 2013).

Determining Clinical Use of Speech Intelligibility Measures

As discussed, there are several methods for measuring intelligibility, yet it is unknown as to which measures are used in clinical settings. It has been suggested that many clinicians frequently estimate (Gordon-Brannan & Hodson, 2000). On the other hand, it is thought that clinicians use a variety of methods (Skahan et al., 2007). Further, it has been claimed that the most widely-used objective clinical measures are word recognition and rating scales (Schiavetti, 1992), but no broad survey of clinical practices in pediatric intelligibility has been conducted. Without such evidence, the assumption that many clinicians commonly make impressionistic statements about intelligibility or use a variety of other available measurement approaches is contradictory and remains unproven. We will test the a priori hypothesis that SLPs use a variety of methods to evaluate intelligibility (Gordon-Brannan & Hodson, 2000; Schiavetti, 1992; Skahan et al., 2007). The purpose of this study is to answer three research questions:

1) What speech intelligibility measures are currently being used by pediatric SLPs?
2) What intelligibility measurement approach is being used most frequently?
3) What tools do speech pathologist think are needed to aid in the measurement of speech intelligibility?
4) What factors relative to both the speaker and the listener do SLPs consider when determining intelligibility?

It is hypothesized for research questions one and two that SLPs use different methods for measuring speech intelligibility with the most common being estimation from measures of speech severity or speech language samples. The third and fourth research questions are exploratory to gain understanding of the types of tools that are needed to establish standard measurement approaches for pediatric speech intelligibility.

Chapter 2

Manuscript

Survey of Pediatric SLP Methods for Measuring Speech Intelligibility

Introduction

Intelligibility is a complex construct with multiple definitions aiming to capture the multifaceted nature of the phenomenon. Connolly (1986) defines intelligibility as context-dependent in that intelligibility can differ based on the listener's familiarity with the topic, the speaker's voice, and/or the speaker's dialect. Schiavetti (1992) defines speech intelligibility as the match between the intention of the speaker and the response of the listener to the speech passed through the transmission system. Munro and Derwing (1995) and Yorkston et al. (1996) similarly approach speech intelligibility as the extent to which an utterance is understood by the listener. Therefore, a valuable measure of pediatric speech intelligibility should capture: (1) the degree to which the produced acoustic signal is understood by the listener; (2) how well the listener understands the speaker's intended message; (3) the familiarity of the speaker to the listener; and (4) other context-dependent factors.

Intelligibility is the most important factor in communication. If a speaker is not intelligible, communication fails (Munro, 2010). This paper reviews a variety of methods described in Kent et al. (1994) for measuring speech intelligibility that clinicians can choose from based on the client's needs. While there is not a consensus on how to measure the intelligibility of pediatric speech, SLPs generally agree that measuring intelligibility is a critical factor in assessment (Hustad et al., 2019). It has been suggested that many clinicians commonly make impressionistic statements about intelligibility rather than using other available measurement approaches thought to be more time-consuming (Gordon-Brannan & Hodson, 2000). In the Forum on Phonology in American Journal of Speech Language Pathology (AJSLP), Bleile (2002) and Khan (2002) provided their perspectives on how they approach assessment. Bleile (2002) stated that the level of intelligibility was observed by listening to clients' spontaneous speech samples. Khan (2002) reported that initial SSD assessment focused on establishing eligibility for speech-language services with additional testing at a later time to create more comprehensive goals. In Skahan et al. (2007), of 312 SLPs surveyed, it was found that 75.4% always subjectively estimate when measuring speech intelligibility. As mentioned previously, a variety of methods exist for measuring intelligibility (Kent et al., 1994), but Skahan and colleagues' investigation did not reveal which methods were being used (Skahan et al., 2007). Farquharson and Tambyraja (2019) survey of SLPs indicated that intelligibility assessment is not required for eligibility determination. This survey did not query clinicians' speech intelligibility assessment practices, therefore it is still unknown which methods were being used. The central goal for SLPs working with any population is to ensure clients are able to communicate using intelligible speech and listeners are able to understand and correctly decipher what has been said (Miller, 2013). Level of intelligibility impacts multiple aspects of therapy, including need for services, intervention priorities, and evaluation of success of intervention (Skahan et al., 2007).

The available tools for speech intelligibility evaluation have predominately served for the evaluation of adult motor speech deficits secondary to dysarthria (Kent et al., 1989; Yorkston et al., 1992). Few measures published for use in the pediatric populations have been standardized and many developed are no longer in print (Allison, 2020). In the absence of standardized measures, word recognition and scaling methods have been used with the pediatric population. The lack of readily available standardized measures of speech intelligibility for pediatric clients has often left clinicians in the position of using their best clinical judgment for diagnosis of speech intelligibility deficits.

The purpose of this paper is to (1) review available methods of speech intelligibility measurement and (2) report survey results to identify what methods are currently being used by SLPs serving the pediatric population.

Intelligibility Measurements: A Review

The primary procedures for measuring speech intelligibility broadly fall into one of two categories: signal dependent or signal independent. The signal dependent measures occur when the listener judges the intelligibility of a speaker based upon the acoustic signal alone (audio recording) without the speaker being present. The listener's judgment of intelligibility is based on only the speech signal itself, using supra-segmental and segmental features, as well as phonotactic predictability and probability. The listener does not have access to context clues (e.g. facial expressions and gestures) available when communicating face to face or when observing a video of the speaker. Syntactic and semantic components are also not considered when judging intelligibility using a signal dependent measure. In contrast, when using a signal independent approach, the speaker may be present when the listener is judging intelligibility. The listener has access to the contextual cues during their judgement of intelligibility. Listeners may observe the speaker directly, by viewing a video, or by making judgements based on prior experiences with the speaker.

Intelligibility Measurement Types

Direct approaches for measuring speech intelligibility generally fall into one of three categories: word recognition tasks, scaling tasks, and automated analysis. Each present with different advantages and pitfalls that could help to explain why some methods are chosen over others and the limitations of adopting either as a standard method chosen for speech intelligibility evaluation.

Word Recognition Approaches for Measuring Speech Intelligibility

Word recognition/transcription measures are one of the common measures used for speech intelligibility (Pascoe et al., 2006). Listeners orthographically transcribe or select the word thought heard in a signal dependent listening task. Open-set word recognition requires listeners to write down words thought heard (Lousada et al., 2014). A score of intelligibility is calculated by dividing the number of correctly identified words by the total number of target words yielding a percentage of intelligibility.

During closed-set word recognition tasks, listeners are given a list of similar-sounding words, and they must choose (multiple choice) which word matches what they heard (Lousada et al., 2014). Response options of "other" with a write-in box, a "select all that apply", and "no response" are provided in the multiple-choice format. Whereas forced choice measures require the listener to choose the word thought heard from a list only.

Open and closed set word recognition tasks have been employed in research and have served as the basis for intelligibility assessment tools. For example, the Test of Children's Speech (Hodge et al., 2009), has both a word and sentence test, and listeners can measure intelligibility by employing either an open set or closed set measure (Hodge & Gotzke, 2014). Khwaileh and Flipsen (2010) studied the speech intelligibility of single words from *The*

Children's Speech Intelligibility Measure (CSIM) and sentences from the *Beginners Intelligibility Test (BIT)* in children with cochlear implants. Judges listened to participants say the CSIM words and orthographically transcribed what they thought they heard (open-set) and also by use of multiple-choice (closed-set). The BIT sentences were orthographically transcribed. Orthographic transcription has been considered the "gold standard" for objective measurement on speech intelligibility and has been widely used in research (Beijer et al., 2012; Hodge & Gotzke, 2014; Hustad et al., 2020; Khwaileh & Flipsen, 2010; Lagerberg et al., 2019).

Advantages & Pitfalls of Word Recognition Tasks

Word recognition tasks have been described as giving an objective measure of speech intelligibility, making them attractive to clinicians (Allison, 2020). Though this may appear to be the case, whether or not a word is recognized by a listener is influenced by many variables, which can result in poor inter- and intra-rater reliability. These influences include the listener's familiarity with the speaker and their dialect, familiarity with the content of the conversation, the communicative environment, background noise, age and hearing status of the listener, shared cultural background with the speaker, physical/emotional state of the listener, disordered speech in general, and those introduced by the stimuli items presented during assessment. The extent of each of these influences does not have to be great in order to affect recognition of words (Miller, 2013). One advantage of word recognition tasks is that they provide a percentage score which could be used to quantify intelligibility and be used to track change over time. A limiting factor, however, is the need to recruit nonfamiliar listeners and the time needed to complete the listening tasks. Additionally, speech samples must be recorded under ideal conditions and post processing of the recordings required for presentation to the listeners. Further it may be assumed that listeners must complete the listening tasks in isolated sound treated rooms and with research grade listening equipment often not available to clinicians practicing outside of university settings. These barriers to implementation may dissuade some from using word recognition tasks in favor of perceived faster methods (Allison, 2020). Other considerations include the fact that while an individual may be intelligible in a sound-treated room, this method may lead to an overestimation of their intelligibility. In real-world scenarios, the individual is having to communicate with the presence of background noise, competing demands, and movement; similarly, the listener must also overcome these factors in order to decipher the speaker's message. When placed in conditions that are more similar to everyday situations, it is likely the speaker's intelligibility will decrease (Miller, 2013).

Scalar Approaches for Measuring Speech Intelligibility

Clinicians also frequently use rating scales to measure intelligibility because of the ease of administration and the quick results they provide (Miller, 2013). Commonly used rating scale types include ordinal, visual analog, and direct magnitude estimation scales.

Ordinal Scales

Ordinal scales typically employ category hierarchies, and the person completing the scale must choose where the client best fits along the hierarchy. Likert scales are most commonly used for these purposes, and prompt ratings using terms such as "strongly disagree, disagree, neutral, agree, strongly agree" for participants to choose from in response to questions or statements (Tastle & Wierman, 2006).

There are a few published ordinal scales available for measuring child speech intelligibility. The Viking Speech Scale is a published measure to score the level to which children are understandable to strangers and unfamiliar conversational partners. (Pennington et al., 2011). This scale, designed for children with motor speech disorders secondary to cerebral palsy, is used to classify speech according to four different levels. Level I indicates that the child is not affected by a motor disorder with similarities of same age peers, Level II speech is imprecise but understandable, Level III speech is unclear and not understandable out of context, and Level IV indicates no understandable speech. The Intelligibility in Context Scale (ICS) is another ordinal scale, developed to obtain ratings of how intelligible children are to parents and other listeners (McLeod et al., 2012). The ICS is an ordinal scale that can be given to parents to fill out. It utilizes a 5-point Likert scale, where 1 = never, 2 = rarely, 3 = sometimes, 4 = usually, and 5 = always. Parents and others familiar with the child's speech are asked to fill out seven questions, which query the child's intelligibility to the parents and other communication partners. This tool, having been validated for Australian English and other world languages, has limited normative data and has not been examined relative to "gold standard" (Hustad et al., 2021). Hustad et al. (2012) asked listeners to rate on a scale from 1, which was *difficult or impossible to understand,* to 7, which was *very easy to understand,* how understandable the speech samples of 4-year olds with and without cerebral palsy would be to another adult. The National Outcome Measurement System (NOMS) used by the American Speech-Language-Hearing Association (ASHA), which is a system designed to collect and aggregate outcomes data. NOMS uses Functional Communication Measures (FCMs) for each component or disorder. The FCMs are disorder-specific 7-point scales that enable a clinician to track the individual's progress over time by periodically rating their functional skills. A rating of 1 would be the least functional level, and a rating of 7 would be the most functional level (Mullen & Schooling, 2010).

Visual Analog Scales

Visual analog scales (VAS) involve the listener hearing speech from the speaker and then marking on a scale where they think that speaker's intelligibility would fall (Abur et al., 2019).

VAS typically are 100mm in length and anchored on either end with labels or descriptions of the extremes of the characteristic that is being measured (Marsh-Richard et al., 2009). The scale may be marked using a pen-and-paper method or a digitized version of the scale. In the Abur et al. (2019) study, adult listeners judged intelligibility using a sliding scale ranging from 0% to 100%. Chiu and Neel (2020) used VAS was used to rate individual parameters of intelligibility instead of overall intelligibility to determine if the ratings of the parameters and ease of understanding from VAS could predict speech intelligibility in quiet and in noise for speakers with Parkinson's disease. Using a computer, listeners rated ease of understanding, articulatory precision, voice quality, resonance balance, and prosodic adequacy. Tjaden et al. (2014) used a vertical, continuous, 150-mm VAS, as described in Sussman and Tjaden (2012), to compare ratings of intelligibility of speakers with multiple sclerosis and Parkinson's disease while using strategies of rate reduction, increased vocal intensity, and clear speech. Listeners used a digital VAS scale to select where they thought each listener's intelligibility would fall. The scores were converted to a number between 0 and 1.0, with 0 being *no impairment* to 1.0 being *severely impaired* using a computer program.

Direct Magnitude Estimation Scales

Direct magnitude estimates (DME) has also been used in several studies. There are two ways to employ DME – with a modulus (DME-M) or without a modulus (DME-WM; (Whitehill et al., 2002). When using DME-M, the listener is provided a modulus or standard to listen to that serves as a referent to which all future speech samples will be compared. The referent speech sample can be low, middle, or high intelligibility, but is most often representative of falling within the middle range. The modulus is assigned a modulus number, most commonly 10 or 100. The listener then rates the intelligibility of each following speech sample relative to the modulus sample. DME-WM requires the listener to rate the first speech sample and then rate all subsequent speech samples relative to the first speech sample (Schiavetti, 1992).

Kim and Kuo (2012) used DME to determine if presentation level has an effect on the intelligibility of speakers with dysarthria. DME-WM and DME-M ratings were lower for dysarthric speakers than healthy controls. In another study researchers posed the questions of how dysarthric speakers rate other dysarthric speakers' speech as compared to speech-language therapists (SLTs), and then how dysarthric speakers rated their own speech as compared to SLTs. Each listener was first given the standard sample to listen to and told that the sample had a rating of 100. Listeners were then instructed to rate future samples based on the standard sample. The study found that people with dysarthria do rate dysarthric speakers differently than SLTs or naïve listeners. When it came to dysarthric speakers rating their own speech, however, there was no j

Advantages & Pitfalls of Scalar Measurements

Scalar measures are simple to understand and quick to complete during an evaluation or therapy. Additionally, rating scales can be given to the parent or caregiver to gain their perspective on the child's intelligibility. Though rating scales are quick and easy to use in clinical situations, there are some drawbacks associated with them. Allison (2020), Miller (2013), and Schiavetti (1992) point out that not all listeners subjectively rate intelligibility the same, i.e., one listener's rating of 4 could be the same as another listener's rating of 6. Because of this, the reliability of rating scales can be lower than is clinically acceptable.

Published Measures

Published standardized measures are common for many areas of speech-language pathology, such as articulation and language, but they are less common for intelligibility. Current published measures include the Beginner's Intelligibility Test (BIT), the ICS, and the Speech

Intelligibility Test (SIT). Though these measures have been applied for research purposes, it is thought that they are not widely used in clinical practice. The BIT requires participants to repeat 10 short sentences that are spoken by the examiner (Castellanos et al., 2014; Osberger, 1994). As described previously, the ICS is a parent-reported measure of the child's speech when speaking in varying contexts (McLeod et al., 2012). The SIT is a computerized software that lets the clinician or researcher administer, score, and store the results of the test. The software prepares an 11-sentences test, with one sentence at each length from 5 to 15 words. The participant reads each sentence after the examiner and the audio is recorded. The audio is then played for different judges, and the judges type exactly what they hear into the program. The software then compares the judges' response to the sentence and scores the items (Dorsey et al., 2007).

Advantages & Pitfalls of Published Measures

Published standardized measures present the ability to compare children to their age-matched peers. In addition, the fact that the stimulus material is the same allows for direct comparison of results to former studies using the same published measure, as was done in Chin et al. (2012). Administration and scoring of such measures is relatively simple and quick, adding to the appeal. But as Castellanos et al. (2014) points out, the BIT also tests a child's working memory and control processes. The child must remember the sentence the examiner read aloud and repeat it back, which requires the ability to encode the information, store it, and then retrieve it. If a child has impairments in these areas, the BIT is not testing only their intelligibility, but their working memory skills as well. Clinicians must be careful to ensure that they are truly testing intelligibility when administering published measures. The ICS is a parent-reported measure, and clinicians must keep this in mind when considering the results. Parents often understand their child's speech best because they are familiar with them (Carley & Hustad, 2021). The SIT is a computer program and requires adequate technology to use; some clinicians may not have access or funds to be able to use it.

Automated Methods

Acoustic-based measures use parameters such as vowel space, speech rate, voice onset time, speech variability index, speech duration, number of syllables, syllable boundaries, and others to measure a speaker's intelligibility. Lee et al. (2014) used automated measures in their study of the speech intelligibility of children with cerebral palsy using acoustic measures (e.g., temporal, vowel spectral, nasality, and voice measures). Gaussian mixture models (GMM) have also been used to automatically evaluate speech intelligibility and hypernasality in children with cleft palate. Results showed that the algorithm used to evaluate speech gave a good speech classification (Ling et al., 2013).

Advantages & Pitfalls of Automated Methods

Traditional methods of measuring speech intelligibility can be time-consuming, and there is often a demand for fast and reliable assessment when communication disorders are present (Middag et al., 2009). Automated methods present an objective and quick way for clinicians to assess a patient's intelligibility (Van Nuffelen et al., 2009). In addition, automated analysis allows for easy repeatability and has positive implications for rural or remote areas where clinicians may be scarce (Paja & Falk, 2012). Though automated methods have the potential to be useful assessment tools, the final judge of intelligibility in real-world scenarios is the human ear. Therefore, automated methods should work with clinician's subjective ratings and not in place of them (Jiao et al., 2019). Clinicians also are comfortable with subjective and informal measures of intelligibility and value it highly, meaning that objective measures may be slow to be accepted in the field (Jiao et al., 2019; Lee & FitzGibbon, 2008; Miller, 2013).

Determining Clinical Use of Speech Intelligibility Measures

As discussed, there are several methods for measuring intelligibility, yet it is unknown as to which measures are used in clinical settings. A summary of the pros and cons of selected intelligibility measures can be found in Table 1.

Table 1.

Intelligibility Measures

Type	Signal Dependent	Signal Independent	Pros	Cons
Word Recognition	Yes	Yes	Objective Gives a percentage Can track change over time	Time-consuming Equipment needed may not be readily available
Ordinal Scales	Yes	Yes	Simple and quick to understand Can be given to parents or caregivers Easy to use	Interrater reliability is lower than what is clinically acceptable Can be affected by speech characteristics
Visual Analog Scales	Yes	Yes		
Direct Magnitude Estimation (DME)	Yes	Yes		
Published Measures	No	Yes	Comparison to age-matched peers Allows for direct comparison to former studies	Standardized tests sometimes test more than they are marketed for
Automated Measures	Yes	No	Objective and quick Easy repeatability Implications for rural or remote areas	The human component of assessment is important Clinician comfort

It has been suggested that many clinicians frequently estimate speech intelligibility (Gordon-Brannan & Hodson, 2000). On the other hand, it is thought that clinicians use a variety of methods for speech intelligibility assessment (Skahan et al., 2007). No broad survey of clinical practices in pediatric intelligibility has been conducted. Without such evidence, the assumption that many clinicians commonly make impressionistic statements about intelligibility or use a variety of other available measurement approaches is contradictory and remains unproven. We will test the a priori hypothesis that SLPs use a variety of methods to evaluate intelligibility (Gordon-Brannan & Hodson, 2000; Schiavetti, 1992; Skahan et al., 2007). The purpose of this study is to answer four research questions:

1) What speech intelligibility measures are currently being used by pediatric SLPs?
2) What intelligibility measurement approach is being used most frequently?
3) What tools do speech pathologist think are needed to aid in the measurement of speech intelligibility?
4) What factors relative to both the speaker and the listener do SLPs consider when determining intelligibility?

It is hypothesized for research questions one and two that SLPs use different methods for measuring speech intelligibility with the most common being estimation from measures of speech severity (articulation/phonology assessments) or speech language samples. The third and fourth research questions are exploratory to gain understanding of the types of tools that are needed to establish standard measurement approaches for pediatric speech intelligibility.

Methods

Participants

140 participants completed the survey sufficiently to address both the survey hypotheses and the exploratory questions. To meet inclusion criteria, participants were required to be 18 years of age or older, be currently practicing as an SLP who worked with the pediatric population, and currently reside in the United States, a territory of the United States, or a United States military base. If inclusion criteria were not met, participants were taken to the end of the survey and their answers were not included in order to maintain the validity of the results.

Materials

Data were collected by distributing an online survey (see Appendix A) through the platform Qualtrics (Qualtrics, Provo, Utah). The survey contained three qualification questions that participants were required to answer and 36 questions in four subject areas: (a) Personal demographics/background information, (b) Practices in assessment of speech intelligibility, (c) Consideration of communication dynamics, and (d) Considerations for COVID-19.

- Part A questions were demographic in nature to describe participants in terms of both personal variables (age, gender identification, race, etc.) and educational/work variables (degrees obtained, settings worked in, etc.).
- Part B questions queried the participants' current practices regarding assessment of intelligibility in the pediatric population, including training received, use of specific measures, and rationale for measuring intelligibility.
- Part C questions gathered information regarding the participants' considerations of factors that affect the communication dynamic relative to both the speaker and the listener.

- Part D contained one question that obtained information regarding how participants' approach to intelligibility measurement has been impacted by the COVID-19 pandemic.

Procedure

Prior to large-scale distribution of the survey, it was reviewed by two faculty members with experience in pediatric assessment, survey development, and qualitative analysis. Feedback was provided in order to improve the validity of the survey and ensure the questions would elicit responses that answered the research questions. Changes to question types (open response vs. multiple choice) and question order were made, and some questions were edited, added, or deleted in order to properly address the research questions. When the survey was finalized, it was disseminated via various online groups. A short description of the survey and a link to take interested participants to the survey was posted in three ASHA Special Interest Groups (SIGs), five ASHA Community groups, seven Facebook groups specific to speech-language pathology, and one Twitter group specific to speech-language pathology. The post was submitted to one additional Facebook group, but was not approved by the group's administrators and was not made public in the group. After approximately one week, a reminder to take the survey was posted in all three SIGs, all five ASHA Community groups, and two of the Facebook groups. The reminder post was submitted to two additional Facebook groups, but one was not approved by the group's administration and one was deleted from the group's page after posting. A reminder was also posted on the Twitter group. A reminder post was not attempted in the three remaining Facebook groups due to limited responses.

Data Analyses

Obtained responses were analyzed to test hypotheses. Not all questions were designated in Qualtrics as a forced response, therefore, not all participants completed all questions. In the case that some participants did not complete a question, the question was descriptively analyzed using the number of responses to that question. Likert responses were analyzed using frequencies and percentages.

Three open-ended questions were analyzed as follows: two questions were analyzed using descriptive analyses (percentages) and one survey question was analyzed using qualitative analyses. The qualitative analyses were conducted by the first author (A.B.), the last author (M.S.A) and another graduate research assistant (S.K., see acknowledgements). A.B. and M.S.A. read all the responses to mark meaningful segments in the data. Inductive codes were established, defined, and agreed upon by A.B. and M.S.A. Additional codes were added or removed iteratively. S.K. was asked to read through all the responses and then use the established set of first-stage codes to mark segments and to assign any additional codes, while blind to the other researchers' analysis. Segments determined to overlap more than one code were attached to each relevant code. S.K. did not assign any additional codes, and one code that was not used by either M.S.A., A.B., or S.K. was removed. Following independent coding and subsequent discussion to come to consensus, A.B. and S.K. combined similar codes into categories and categories were organized into overarching themes. A.B. and S.K. attained a 70% inter-coder reliability rating on categories and themes following independent organization and then engaged in discussion to reach 100% agreement.

Results

Demographics

A total of 217 participants initiated the survey. After responses were filtered based on inclusion criteria, 77 participants were excluded from the study. Two were excluded for not consenting to the survey. Seventy-five participants were excluded due to discontinuing the survey at some point: 17 after the qualification questions, 24 after the demographics section, and 34 at other points in the survey that addressed the study hypotheses. Because participants were informed that they could stop taking the survey at any time, if it was noted that they stopped answering questions at any point that was considered them withdrawing from the study and their response was not included. Therefore, results were calculated using the responses of 140 participants who completed the survey sufficiently to address the study hypotheses and exploratory questions.

Participant Background

Participants were queried on their ethnicity, race, gender, language(s), and age. They were also asked about their terminal degree, the year the terminal degree was obtained, work setting, state they practiced in at the time of the survey, and years practicing. Of the participants who indicated their ethnicity (n = 135), a total of four (2.96%) identified as Hispanic or Latino; the remaining 131 (97.04%) identified as not Hispanic or Latino. Of the 139 participants that reported their first language(s), 131 (94.24%) reported American English as their first language. The remaining eight (5.76%) reported a different language as their first language. Other first languages reported include Spanish, Marathi, Hindi, Kannada, Cantonese Chinese, Shanghainese Chinese, Mandarin Chinese, Bengali, Hindi, and a non-American English. Participant demographics details are reported in Table 2.

Table 2.

Participant Demographics

Demographic Variable	Response Count (Percentage %)		
Sex	**(n = 140)**	**Highest Level of**	
Female	137 (97.86%)	**Educational Achievement**	**(n = 140)**
Male	2 (1.43%)	Undergraduate degree	1 (0.71%)
Non-Binary	1 (0.71%)	Masters degree	109
		PhD	(77.86%)
Race/Ethnicity	**(n = 138)**	Clinical doctorate	16 (11.43%)
White	121 (87.68%)	Other	2 (1.43%)
Black/African American	9 (6.52%)		12 (8.57%)
Multiple Races	4 (2.9%)	**Years since graduate**	
Asian	3 (2.17%)	**highest SLP degree**	**(n = 115)**
Middle Eastern	1 (0.72%)	Less than 5 years	26 (22.61%)
		5-10 years	29 (25.22%)
Age	**(n = 132)**	11-15 years	8 (6.96%)
23-34	42 (31.82%)	16-20 years	14 (12.17%)
35-44	23 (17.42%)	21-30 years	25 (21.74%)
45-54	31 (23.48%)	More than 30 years	13 (11.3%)
55-64	33 (25%)		
65-74	3 (2.27%)	**Employment Setting***	**(n = 140)**
		Preschool	57 (40.71%)
Total years of SLP		Elementary School	58 (41.43%)
employment	**(n = 140)**	Middle School	24 (17.14%)
less than 2 years	9 (6.43%)	High School	13 (9.29%)
3-7 years	31 (22.14%)	Hospital	20 (14.29%)
8-14 years	25 (17.86%)	Private Practice	31 (22.14%)
15-24 years	38 (27.14%)	Birth to 3 Program	35 (25%)
25-34 years	23 (16.43%)	University Clinic	22 (15.71%)
35 years or more	14 (10%)	Headstart	4 (2.86%)
		Other	7 (5%)
Location	**(n = 139)**		
Southeast	40 (28.78%)	West	22 (15.83%)
(AL, AR, DE, FL, GA,		(CA, CO, ID, NV, OR, UT,	
KY, LA, MD, MS, NC,		WA)	
SC, TN, VA)		Southwest	13 (9.35%)
Midwest	34 (24.46%)	(AZ, TX)	
(IL, IN, IA, MI, MN, MO,		United States Military Base	1 (0.72%)
ND, OH,WI)			
Northeast	29 (20.86%)		
(CT, MA, NH, NJ, NY,			
PA, RI, VT)			

*N = total number of respondents; % = percentage of respondents based on number of respondents per question; *Employment setting question: check all that apply, will not sum to 100%.*

Practices in Assessment of Speech Sound Disorders (SSD)

Participants were queried regarding their assessment practices for SSD (see Appendix A, Question 10). Of the participants who reported the percentage of their caseload that was receiving services for SSD (n = 137): 12 (8.76%) reported 10% or less of their caseload; 23 participants (16.79%) reported up to a quarter of their caseload; 52 (37.96%) reported up to half of their caseloads; 29 (21.17%) reported up to three-quarters of their caseload; and 21 (15.33%) reported more than three-quarters of their caseload. Three participants did not respond to this question.

Participants were also questioned about the percentage of their caseload that required an articulation and/or phonology assessment (see Appendix A, Question 12). Of the participants who responded (n = 138): nine (6.52%) reported less than 10% of their caseload required the assessments; 14 (10.14%) reported that up to a quarter of their caseload required them; 40 (28.99%) of participants reported up to half of their caseload required articulation and/or phonology assessments; 35 participants (25.36%) reported the assessments for up to three-quarters of their caseload; and 40 participants (28.99%) reported the assessments were required for more than three-quarters of their caseload.

The use of computerized analysis for articulation/phonology assessment and intelligibility assessment was queried. For articulation/phonology assessment (see Appendix A, Question 13), the majority of participants (121; 86.43%) reported not using computerized analysis and 19 (13.57%) said they did use a form of computerized analysis. Those that answered yes were asked to provide the title of the computerized analysis they use. Responses included the Arizona Articulation and Phonology Scale, Fourth Revision (Arizona-4), the Hodson Computerized Analysis of Phonological Patterns (HCAPP), the Hodson Assessment of

Phonological Patterns – 3rd Edition (HAPP-3), the Goldman-Fristoe Test of Articulation 3 (GFTA-3), the Test of Children's Speech Plus (TOCS+), Q global, computerized scoring, Python scripts, and spectrogram analysis. Participants were then asked to indicate which, if any, computerized analysis procedures for intelligibility assessment were used (see Appendix A, Question 14). Of the participants that responded (n = 136): 131 (96.32%) reported not using it; five (3.68%) reported using some type of computerized analysis. The methods used included acoustic analysis (n = 3, 2.21%), automated counters of correct syllables (n = 2, 1.47%), automated counters of correct words (n = 1, 0.74%), automatic speech recognition (ASR) (n = 1, 0.74%), and Google voice captions (n = 1, 0.74%).

Participants were queried regarding typical parent involvement during the assessment (see Appendix A, Question 15). Participants were able to select more than one answer choice, resulting in percentages exceeding 100%. Of the respondents: 102 (72.86%) reported conducting a live interview with parents; 85 participants (60.71%) stated the parent accompanies the child into the assessment; 85 participants (60.71%) reported the parent completes developmental checklists; 64 (45.71%) stated that parents complete a mail-in case history form; 55 participants (39.29%) reported parent/child play during the assessment; 38 participants (27.14%) stated they conduct a phone interview with the parent; 34 (24.29%) stated the parent completes the Intelligibility in Context Scale; 25 participants (17.86%) stated the parent completes a survey; and seven participants (5%) reported other parental involvement. Other responses included IEP meetings, email correspondence, in-person case history, and variation in parent involvement depending on the student and parent. There were 11 participants (7.86%) that said there was no parent interaction during assessment.

Use of Estimation to Measure Intelligibility

Participants reported how often they use subjective estimation to measure speech intelligibility (see Appendix A, Question 16): five (3.57%) indicated never; 10 (7.14%) indicated infrequently; 30 (21.43%) indicated sometimes; 65 (46.43%) indicated often; and 30 (21.43%) indicated always. The participants that answered "never" were not asked the following question regarding the method used when estimating intelligibility.

Of the 135 participants who reported what method they use when they estimate intelligibility (see Appendix A, Question 17): 100 (74.1%) reported using speech samples and 17 (12.59%) reported estimating intelligibility from articulation/phonology assessment scores. The remaining 18 participants (13.33%) reported using another method to estimate intelligibility. Additional methods included using both speech samples and assessment scores, accuracy of Google Meet captions, overall judgement during the assessment (not including scores), parent interview, observations from previous sessions, percent consonants correct (PCC), Intelligibility in Context Scale (ICS), and Test of Children's Speech (TOCS+).

Practices in Assessment of Speech Intelligibility

Participants responded regarding their training in intelligibility assessment (see Appendix A, Question 11): 127 (90.71%) reported receiving training in their graduate courses and 73 (52.14%) reported receiving training in their undergraduate courses. Additional responses included workshops (n = 76, 54.29%), conventions (n = 71, 50.71%), and journals (n = 54, 38.57%). A total of 11 participants (7.86%) reported receiving other training. Other training methods included online professional development courses, research, PROMPT certification, faculty teaching position, conversations with colleagues, school district training, webinars, and practicums. A total of three participants (n = 2.14%) reported receiving training through none of the described methods.

Method of Collection

Participants reported what methods they use to collect intelligibility data (see Appendix A, Question 18): 113 (80.71%) reported using continuous speech sampling; 105 (75%) reported using observations in the child's natural environment; 59 participants (42.14%) reported using single words; 20 participants (14.29%) reported using a formal instrument of speech severity; 11 participants (7.86%) reported using a formal instrument of intelligibility; and 17 participants (12.14%) reported using other methods. Other methods reported included clinical observations, the Intelligibility in Context Scale, parent and teacher interviews, intelligibility ratings from parents and teachers, articulation assessments, percent consonants correct, structured conversations, syllable imitation tasks, and parental anecdotes.

Assessments. Participants were queried regarding the articulation and phonology assessments they used to measure intelligibility (see Appendix A, Question 19). Detailed responses can be found in Table 3. A total of nine participants (6.43%) reported using methods other than those listed in the table. Other methods reported included apps, tables from speech sound disorders texts, the Entire World of R and Entire World of S, the Fluharty screener, informal assessment of connected speech, a motor speech hierarchy, the Korean Standard Picture Articulation and Phonological Test (KS-PAPT), the Nuffield Dyspraxia Programme (NDP3), the CLEAR, and PROMPT motor speech system analysis. A total of 25 participants (17.86%) reported not using any of the given assessments.

Table 3.

Articulation and Phonology Assessments Used to Measure Intelligibility

Test	Number of Responses	Percentage
Goldman-Fristoe Test of Articulation: Sounds in Words	93	66.43%
Goldman-Fristoe Test of Articulation: Sounds in Sentences	73	52.14%
Khan-Lewis Phonological Analysis – 3^{rd} Edition	38	27.14%
Clinical Assessment of Articulation and Phonology	26	18.57%
Arizona Articulation and Phonology Scale 4	20	14.29%
Hodson Assessment of Phonological Patterns – 3^{rd} Edition	15	10.71%
Comprehensive Test of Phonological Processing	9	6.43%
Structured Photographic Articulation Test III	8	5.71%
Diagnostic Evaluation of Articulation and Phonology: Articulation Assessment	8	5.71%
Assessment of Phonological Processes – Revised	7	5%
Linguisystems Articulation Test	7	5%
Contextual Probes of Articulation Competence	4	2.86%
Photo Articulation Test – 3^{rd} Edition	4	2.86%
Test of Phonological Awareness Skills	4	2.86%
Bankson-Bernthal Test of Phonology	3	2.14%
Secord Contextual Articulation Tests	3	2.14%
Fisher-Logemann Test of Articulation Competence	2	1.43%
Deep Test of Articulation	1	0.71%
McDonald-Deep Test of Articulation	1	0.71%
Practical Test of Articulation and Phonology	1	0.71%
Templin-Darley Test of Articulation – 2^{nd} Edition	1	0.71%

Percentages calculated out of 140 responses; check all that apply format was applied for this question so percentages will not sum to 100%

Speech Intelligibility Measures. Participants were also queried regarding the measures of speech intelligibility used during assessment (see Appendix A, Question 20). A total of 136 participants responded to this question. Detailed results can be found in Table 4. A total of 16 participants (11.76%) reported using measures other than those listed in the table. Additional measures reported included percent of words intelligible, percent of words correct, percent of errors in connected speech, the ICS, the AAPS4, the GFTA-2, the KS-PAPT, parent and teacher questionnaires and surveys, district-provided rating scales, percent of speech intelligible to familiar vs unfamiliar listeners, and spontaneous language samples. A total of 42 participants (30.88%) reported using none of the listed measures.

Table 4.

Measures of Speech Intelligibility Used During Assessments

Measure	Number	Percentage
Goldman Fristoe Test of Articulation – 3rd Edition	70	51.47%
Percent Consonants Correct*	35	25.74%
ASHA National Outcomes Measurement System	22	16.18%
Preschool Speech Intelligibility Measure	3	2.21%
Assessment of Intelligibility in Dysarthria Speech	2	1.47%
Children's Speech Intelligibility Test	2	1.47%
Ling's Phonologic and Phonetic Level Speech Evaluation	2	1.47%
Meaningful Use of Speech Scale	2	1.47%
Test of Children's Speech	2	1.47%
Articulation Competence Index	1	0.74%
Quantitative Rating of Performance	1	0.74%
Speech Pattern Contrast Test	1	0.74%
Weiss Intelligibility Test	1	0.74%

*Percentages calculated out of 136 responses; check all that apply format was applied for this question so percentages will not sum to 100%. *For the purposes of this survey, as cited in Kent et al. (1994), the measure Percent Consonants Correct (Shriberg & Kwiatkowski, 1982) is included as a potential intelligibility index; however, it is noted that PCC is not necessarily a direct or recommended measure of speech intelligibility.*

Increment Use. The use of increments of speech to measure speech intelligibility was queried (see Appendix A, Question 21). For each increment, participants were asked to indicate their frequency of use by choosing one of the following options: never, infrequently, sometimes, often, or always. All participants responded to components of this question. Detailed results are summarized in Table 5. Seven participants reported using methods other than those listed in the table, which included environmental sounds, animal sounds, imitated syllables, PCC, peer responses, interviews with those that work with the child, SALT language samples, and the AAPS4.

Table 5.

Use of Increments of Speech for Measuring Speech Intelligibility

Increment (number of respondents)	Never		Infrequently		Sometimes		Often		Always	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Syllables (n = 134)	**50**	**30.71**	30	22.39	22	16.42	17	12.69	15	11.19
Words (n = 136)	4	2.94	12	8.82	25	18.38	44	32.35	**51**	**37.5**
Phrases (n = 134)	12	8.96	8	5.97	27	20.15	**56**	**41.79**	31	23.13
Sentences (n = 139)	7	5.04	9	6.47	23	16.55	**62**	**44.6**	38	27.34
Conversation (n = 140)	5	3.57	5	3.57	14	10	49	35	**67**	**47.86**
Reading speech and/or sentences (n = 128)	**41**	**32.03**	26	20.31	33	25.78	18	14.06	10	7.81
Other increments not listed (n = 28)	**20**	**71.43**	2	7.14	1	3.57	4	14.29	1	3.57

N = total number of respondents; % = percentage of respondents based on number of respondents per question; gray cell color/bold font indicates the most common response for that variable.

Speech Samples. Participants were asked to report how they use speech samples, if they use them (see Appendix A, Question 22). A total of 98 participants responded. Of those responses: 30 (30.61%) reported using words understood; 20 (20.41%) reported using utterances understood; 17 (17.35%) reported using estimation; eight (8.16%) reported using PCC; six (6.12%) reported using percentage or number of errors; four (4.08%) reported using words correct; and one (1.02%) reported using vowels correct. Two participants (2.04%) reported using each of the following methods: percent of words misunderstood, a form of technology, or mean length of utterance (MLU). A total of 16 (16.33%) participants gave answers not falling into these categories. Other answers included vague statements with no other context or explanation (e.g. "percentages"), sequence cards, barrier games, standardized assessment, PROMPT protocol, transcription, and rating scales.

Considerations of Communication Dynamics

Participants were asked to rate the extent to which variables relative to both the speaker and listener affected speech intelligibility. Participants were asked to choose one of the following for each variable: not at all, very little, little, a fair amount, and great. All 140 participants responded to the speaker variable question (see Appendix A, Question 23) and 139 participants responded to the listener variable question (see Appendix A, Question 25). Responses for variables relative to the speaker can be found in Table 6 and responses for variables relative to the listener can be found in Table 7. Participants were asked to state which other variables relative to the speaker they believed affected intelligibility (see Appendix A, Question 24). A total of 64 participants responded. Responses were grouped by category. Responses related to perceptual characteristics of speech (n = 18, 28.13%) included loudness, rate of speech, prosody, and pitch. Responses related to the temperament of the speaker (n = 12, 18.75%) included anxiety, mood, behavior, motivation to engage in services, level of comfort in certain situations, and experiences in the child's past such as trauma. Responses related to the child's physical features (n = 11, 17.19%) included hearing, velopharyngeal incompetence, motor control, motor planning, incomplete articulatory contact, muscle tone, and range of movement of articulators. Responses related to cognition (n = 8, 12.5%) included awareness of listener difficulty and correct productions, presence of another language/primary language, cognitive status, and social skills. Responses related to the environment (n = 5, 7.81%) included number of people around, opportunities to practice in living environment, items available for context, internet connectivity, background noise, and social aspects. Responses related to speech sound errors (n = 3, 4.69%) included consistency and type of speech sound errors. Other responses (n = 12, 18.75%) included parent models, parent support, type of task, vowel distortions, familiarity with targets, and etiology of disorder.

Table 6.

Speaker Variables

Variable (number of respondents)	Not At All		Very Little		Little		A Fair Amount		Great	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Resonance (hypernasality) (n = 140)	0	0	8	5.71	36	25.71	**71**	**50.71**	25	17.86
Voice quality (n = 139)	1	0.72	15	10.79	**58**	**41.73**	54	38.85	11	7.91
Age (n = 139)	4	2.88	15	10.79	45	32.37	**58**	**41.73**	17	12.23
Linguistic factors (n = 139)	1	0.72	16	11.51	49	35.25	**64**	**46.04**	9	6.47
Number of speech sound errors (n = 139)	0	0	1	0.72	1	0.72	36	25.9	**101**	**72.66**
Type of speech sound errors (n = 139)	0	0	2	1.44	1	0.72	29	20.86	**107**	**76.98**
Frequency of speech sound errors (n = 140)	0	0	1	0.71	2	1.43	22	15.71	**115**	**82.14**
Dialect spoken (n = 138)	5	3.62	23	16.67	**56**	**40.58**	47	34.06	7	5.07
Familiarity with listener (n = 140)	1	0.71	5	3.57	22	15.71	**75**	**53.57**	37	26.43
Familiarity with context of conversation (n = 139)	0	0	2	1.44	19	13.67	**77**	**55.40**	41	29.5
Gestures and facial expressions (n = 138)	2	1.44	9	6.52	44	31.88	**67**	**48.55**	16	11.59
Background noise (n = 138)	2	1.44	14	10.14	36	26.09	**66**	**47.83**	20	14.49
Shared cultural background with listener (n = 139)	5	3.6	17	12.23	52	37.41	**56**	**40.29**	9	6.47
Physical/emotional state (n = 140)	0	0	18	12.86	**57**	**40.71**	55	39.29	10	7.14

N = total number of respondents; % = percentage of respondents based on number of respondents per question; gray cell color/bold font indicates the most common response for that variable.

Table 7.

Listener Variables

Variable (number of respondents)	Not At All		Very Little		Little		A Fair Amount		Great	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Familiarity with speaker (n = 139)	0	0	0	0	9	6.47	64	46.04	**66**	**47.48**
Familiarity with dialect of speaker (n = 139)	1	0.72	4	2.88	24	17.27	**76**	**54.67**	34	24.46
Familiarity with content of conversation (n = 138)	0	0	2	1.45	6	4.35	**73**	**52.9**	57	41.3
Communication environment (n = 139)	0	0	5	3.6	18	12.95	**79**	**56.83**	37	26.62
Background noise (n = 139)	0	0	5	3.6	30	21.58	**69**	**49.64**	35	25.18
Age (n = 139)	6	4.32	13	9.35	**55**	**39.57**	46	33.1	19	13.67
Shared cultural background (n = 139)	3	2.16	12	8.63	**55**	**39.57**	53	38.13	16	11.51
Physical/emotional state (n = 139)	8	5.76	21	15.11	**62**	**44.6**	39	28.06	9	6.47
Hearing status (n = 139)	0	0	4	2.88	5	3.6	49	35.25	**81**	**58.27**

N = total number of respondents; % = percentage of respondents based on number of respondents per question; gray cell color/bold font indicates the most common response for that variable.

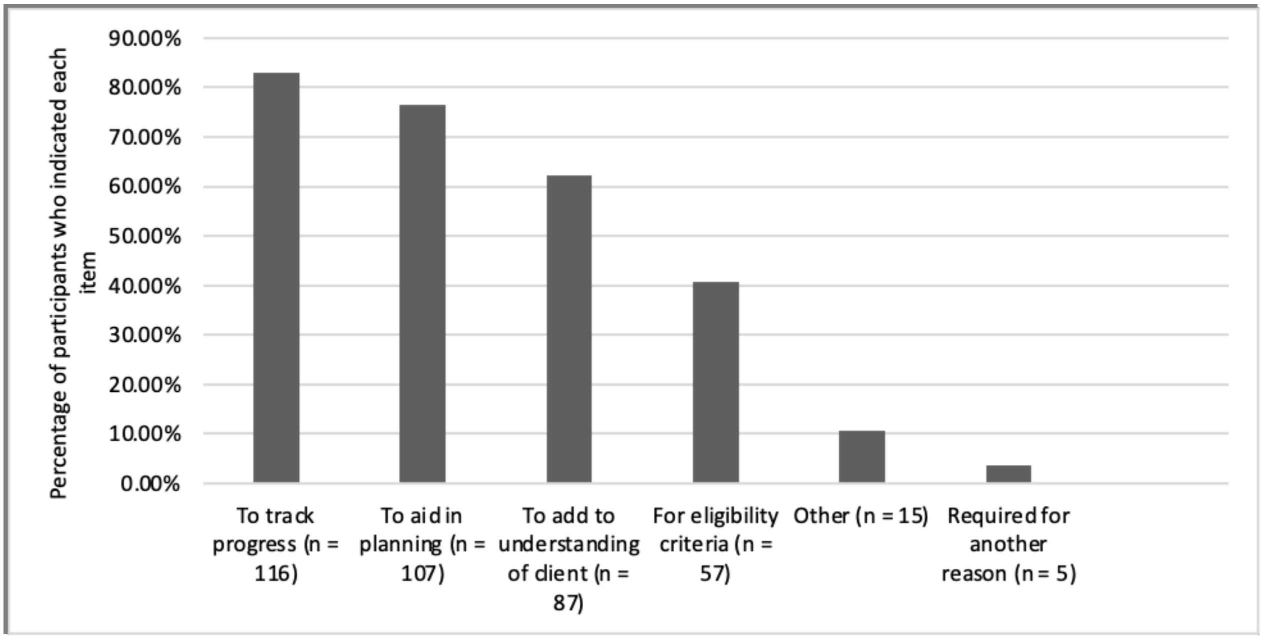
Participants were also asked to state which variables relative to the listener they believed affect intelligibility (see Appendix A, Question 26). A total of 36 participants responded. The responses were grouped by category. Responses related to the knowledge of the listener (n = 15, 41.67%) included: familiarity with topic, familiarity with children with SSD, familiarity with phonological processes, and occupation. Responses related to the focus of the listener (n = 8, 22.22%) included: attention, patience, concentration, eye contact, stress level, and interest. Other responses (n = 15, 41.67%) included cognitive status, background noise, positioning of speaker, ability to understand gestures, masks, and desire for child to receive support.

Rationale for Measuring Speech Intelligibility

Participants were queried regarding the reason they measure intelligibility (see Appendix A, Question 30). The reasons given were to track progress, to aid in planning, to add to understanding of the client, for eligibility criteria, required for another reason, or other. The percentage of SLPs who indicated each option is detailed in Figure 1. Other required reasons for intelligibility measurement included for initial or re-evaluations, for insurance, and for parent concern. Other reasons for intelligibility measurement include discrepancy in classroom functioning, because it’s functional, to aid in assessment, to provide descriptive information, to gain a more accurate picture of the child’s communication skills, and to assist teachers.

Figure 1.

Reasons for Measuring Intelligibility



**Percentages calculated out of 140 responses. Question format was check all that apply; percentages will not sum to 100%*

Participants were asked what they use intelligibility measures for (see Appendix A, Question 31). All participants indicated all prompts that applied: 122 (87.14%) responded for baseline measurements; 121 (86.43%) responded for treatment planning; 98 (70%) responded for intervention; 91 (65%) responded for eligibility for services; and two (1.43%) responded other. Other responses included discrepancy in perception of student functioning, and when a child is showing the need for services in speech in addition to language.

Use of Specific Measures

Participants were asked to rate how often they used specific measures by choosing one of the following: never, infrequently, sometimes, often, or always. (see Appendix A, Questions 27, 28, 29). The responses can be found in Table 8.

Table 8.

Use of Specific Measures

Measure (number of respondents)	Never		Infrequently		Sometimes		Often		Always	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Word recognition (n = 139)	**49**	**35.25**	18	12.95	27	19.42	27	19.42	18	12.95
Multiple choice (n = 139)	**108**	**77.7**	15	10.79	14	10.07	2	1.44	0	0
Forced choice (n = 138)	**110**	**79.71**	15	10.87	8	5.8	5	3.62	0	0
Ordinal scale (n = 140)	**67**	**47.86**	9	6.43	25	17.86	25	17.86	14	10
Visual analog scales (n = 139)	**96**	**69.06**	8	5.76	23	16.55	12	8.63	0	0
Direct magnitude estimation (n = 140)	**123**	**87.86**	7	5	8	5.71	2	1.43	0	0
Severity assessment measures (n = 140)	23	16.43	12	8.57	21	15	**51**	**36.43**	33	23.57
Acoustic based measures (n = 140)	**86**	**61.43**	21	15	22	15.71	8	5.71	3	2.14
Automated measures (n = 139)	**120**	**86.33**	13	9.35	6	4.32	0	0	0	0

N = total number of respondents; % = percentage of respondents based on number of respondents per question; gray cell color/bold font indicates the most common response for that variable.

Practices in Assessment of Speech Intelligibility

Of the 139 participants that reported the type of speech production they elicited for speech intelligibility assessment (see Appendix A, Question 32): 138 (99.28%) reported using spontaneous speech; 81 participants (58.27%) reported using imitative speech; 37 (26.62%) reported using cued speech; and five (3.60%) reported using other types. Other types of speech production elicited that were collected were reading and speech samples taken in a variety of contexts.

Participants were asked if they believed speech-language pathologists currently have adequate tools to assess speech intelligibility (see Appendix A, Question 33). Of 139 respondents: 74 (53.24%) responded yes, while 65 (46.76%) responded no. Participants that responded “yes” were asked to name the adequate tools that we possess (see Appendix A, Question 34). A total of 57 participants responded. The most common response (n = 35, 61.4%) was that clinician’s experience and training are sufficient to accurately measure speech intelligibility, closely followed by current formal and informal assessments/measures (n = 57.89%). A total of 6 participants (10.53%) responded that all had been listed in previous questions in the survey. Parent and family interviews or input were indicated by four participants (7.02%). Two participants (3.51%) responded with technology such as recordings and online tools. Other responses (n = 3, 5.26%) included speech science techniques, pattern analysis, and the sentiment that we do not have one, indicating that user erroneously marked “yes” on the question. Participants that responded “no” were asked to name what tools are needed to measure intelligibility (see Appendix A, Question 35). Those that responded “yes” were also asked to respond to this question. A total of 95 participants responded. The most common response (n = 39, 41.05%) was needing a new process or a set process to measure intelligibility that is clearly described. Eighteen of those participants (18.95%) specifically stated a standardized or objective measure of intelligibility was needed. Sixteen responses (16.84%) specifically stated a need for more time for assessments or a tool that is quick to use. The next most common response was needing more clinician training (n = 32, 33.68%). Six of those responses (6.32%) specified needing more training to understand the client's background (culture, community, dialect, race, diagnosis/disorder). A total of 16 participants (16.84%) responded that clinicians need more access to methods of measuring intelligibility. Twelve participants (12.63%) responded that more listeners or a broader context were needed to accurately measure intelligibility. A total of 17 (17.89%) participants gave other responses, including uncertainty about what is needed, recording devices, and a combination of different measures.

Participants were asked how the COVID-19 pandemic and shift to telepractice affected how they measure intelligibility (see Appendix A, Question 36). Qualitative analysis was used for responses. Responses (n = 125) were sorted into answers that identified measurement effects (n = 58, 46.4%) and those that did not identify measurement effects (n = 73, 58.4%). Measurement effects were defined as responses that directly stated a change that the participant had implemented due to the COVID-19 pandemic. Non-measurement effects were defined as responses that did not clearly state an action. Examples are responses that stated feelings (e.g. "Telepractice is harder.") Responses for measurement effects were organized into nine codes, then collapsed into four categories, and finally, two overarching themes emerged. The non-measurement effects were organized into seven codes, then collapsed into two categories, and then two overarching themes emerged. Codes, categories, and themes can be found in Figure 2 for measurement effects and Figure 3 for non-measurement effects. For the responses that identified measurement effects, the themes identified were: for some clinicians, nothing has changed; and some clinicians have had to adapt to continue to accurately measure intelligibility. For the responses that did not identify measurement effects, the themes identified were: ease and accuracy of assessment across modalities has been affected; and current technology available for telepractice is not sufficient to accurately measure intelligibility.

Figure 2.

Codes, Categories, and Themes of Measurement Effects

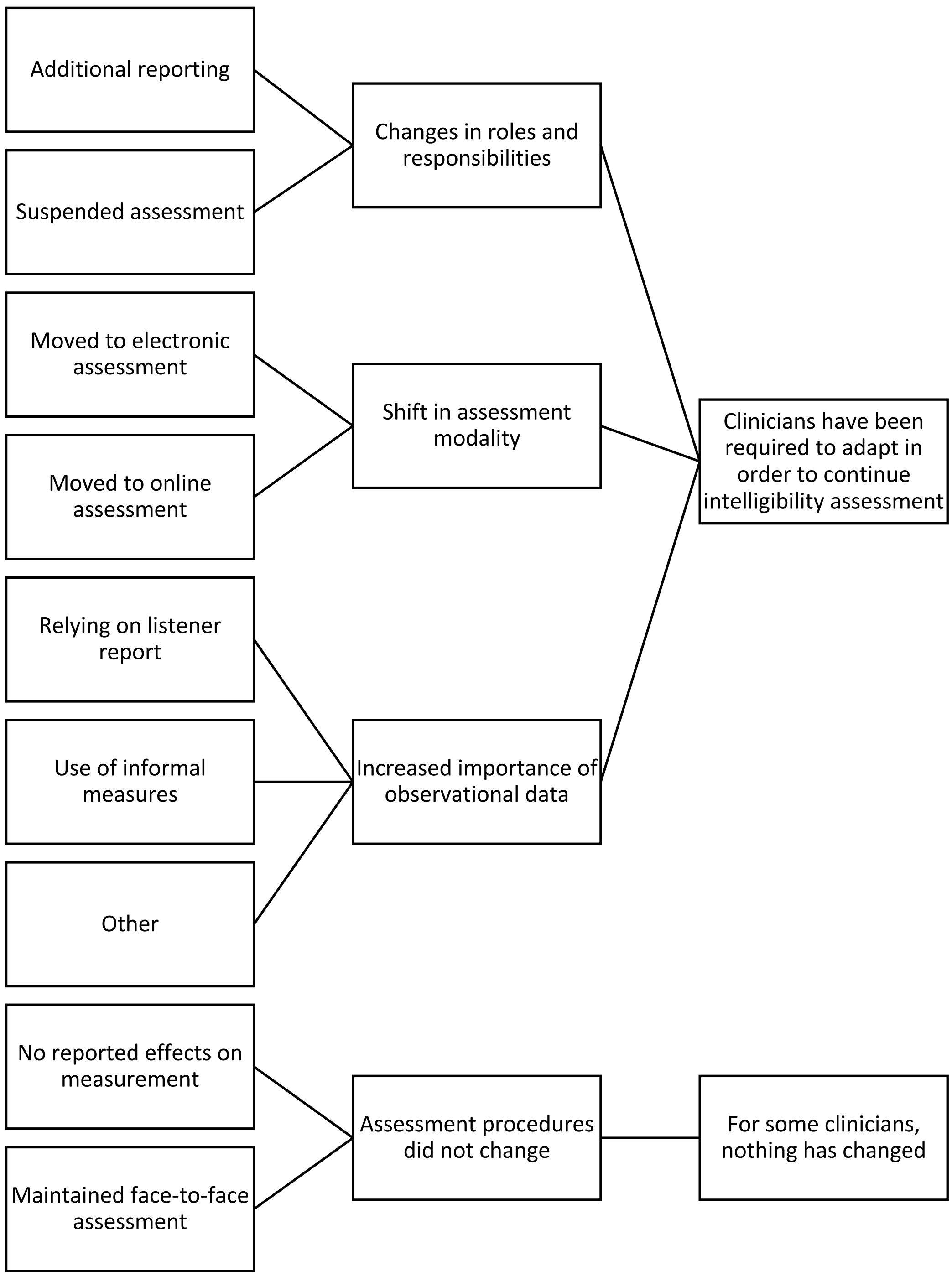
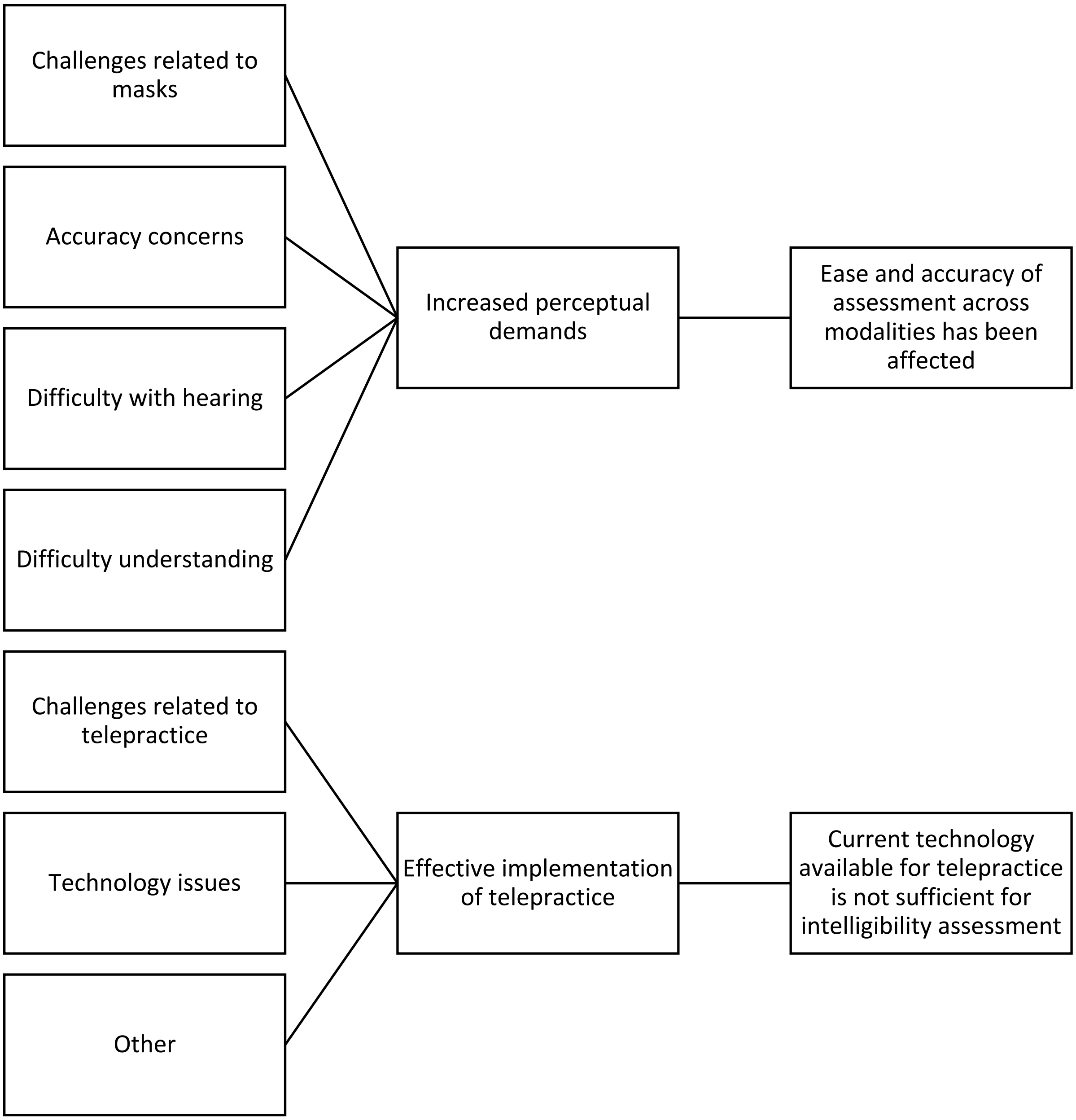


Figure 3.

Codes, Categories, and Themes of Non-Measurement Effects



Discussion

The aim of this investigation was to characterize current clinical practice patterns of pediatric SLPs for determining speech intelligibility, as well as the variables they consider and the tools they think are necessary to accomplish this assessment.

Current Use of Speech Intelligibility Measures

The first research question addressed the speech intelligibility measures currently being used by pediatric SLPs. Previous studies reported that SLPs do not use one specific method to measure intelligibility (Gordon-Brannan & Hodson, 2000; Schiavetti, 1992; Skahan et al., 2007). Therefore, the first study hypothesis was that SLPs would report using many different methods. Survey evidence supported this hypothesis for all levels of intelligibility assessment, including specific measures used, assessments used, reasons for measuring intelligibility, type of speech elicited to measure intelligibility, and how data was used once collected.

There were several questions for which a wide variety of answers was given. Each specific measure of speech intelligibility that was queried (subjective estimation, word recognition, multiple choice, forced choice, ordinal scales, visual analog scales, direct magnitude estimation [DME], severity assessment measures, acoustic based measures, and automated measures) had at least 10% of respondents report using that measure in some capacity. Articulation and phonology assessment used and increments of speech used to collect intelligibility data also showed variation. SLPs generally agree that intelligibility measurement is an important part of assessment (Hustad et al., 2019), but many different methods for measurement exist and are documented in the literature (Kent et al., 1994; Miller, 2013; Schiavetti, 1992). Participants were asked to indicate what they used from a list of measures detailed by Kent et al. (1994), and 13 unique measures were indicated by at least one participant.

Because the specific methods that SLPs are using to measure intelligibility have not previously been explored, it is hypothesized that SLPs choose a method based on multiple factors. These factors may include the methods they were taught in graduate school, methods that their clinical supervisors used, requirements specific to their workplace clinical practice culture, requirements specific for qualification of speech services or insurance reimbursement, and time constraints for evaluation. The reason for the wide variation in measurement procedures requires further investigation.

The second research question queried the most frequent measurement reported. The hypothesis that most responses would be estimation from measures of speech severity or speech language samples was supported by the responses. The overwhelming majority of clinicians (96.43%) reported using subjective estimation in some capacity, and the majority reported using a speech sample as the method through which to estimate intelligibility. The majority (83.75%) also reported using severity assessment measures in some capacity. It has previously been reported that estimation in general is the most common measure clinicians use (Bacon, 1995; Gordon-Brannan & Hodson, 2000; Weston & Shriberg, 1992). Some authors have cited lack of time to complete assessments as a reason that estimation is used so frequently (Bacon, 1995; Gordon-Brannan & Hodson, 2000). This conclusion is consistent with what was found in the current survey, as participants also reported needing more time to complete assessments. Time constraints appear to influence SLPs' decisions to estimate intelligibility instead of using a more in-depth approach. Additionally, estimating speech intelligibility has been shown to be variable between listeners. Yorkston and Beukelman (1978) found that the variability of estimates of intelligibility among individual clinicians limited their clinical use, and that estimates were most accurate when a large group of clinicians completed them and the mean estimate was used.

However, it is not realistic to expect a large number of clinicians to be available in most clinical settings. Intelligibility estimates taken in the clinical setting where background noise is relatively low may also overestimate the speaker's intelligibility (Miller, 2013). Time constraints continue to be reported as a barrier to more in-depth intelligibility assessment and flaws in relying on estimation for assessment point to the need for a change in how pediatric SLPs measure intelligibility.

Necessary Tools for Speech Intelligibility Measurement

The third research question asked SLPs to identify tools that are necessary to aid in intelligibility measurement. Over 45% of respondents indicated they did not believe the current tools for measuring intelligibility were adequate. When asked to specify what is needed, many participants specified a set process for measuring intelligibility, with some specifying the need for a standardized or objective measure of intelligibility. As stated previously, many SLPs reported not having enough time to accurately measure intelligibility or specified that a new tool for measuring intelligibility that was quick and could be done in real time would be ideal. Participants also felt that more clinician training and background knowledge was needed, as well as more contexts in which to assess the child. A standard measure of intelligibility would address many of these stated needs. The creation of a standardized training protocol for intelligibility measurement by experts in the field could not only be implemented clinically but also in graduate student training programs. This would address the need for a standardized tool while also giving clinicians more training, as the approach could be taught to any individual, whether they are in graduate school or have been practicing clinically for years. Once such a protocol is established, the intrarater and interrater reliability could be established, which could increase confidence in intelligibility judgements.

Consideration of Speaker and Listener Variables in Speech Intelligibility Measurement

The fourth research question explored the variables relative to the speaker and listener that pediatric SLPs consider when measuring intelligibility. The number, type, and frequency of speech sound errors was indicated to have the most effect on speech intelligibility. This is consistent with Weston and Shriberg (1992) and Hodson and Paden (1981), who discuss the role that phonological processes play in children's intelligibility. Hodson and Paden (1981) found that phonological processes, particularly less frequently occurring phonological processes, were more likely to be present in children that were unintelligible. Speaker variables that were indicated to affect intelligibility the least were voice quality, dialect spoken, and the physical/emotional state of the speaker. The fact that voice quality was one of the least indicated variables is consistent with the findings of Kim et al. (2015), who state that voice quality has been largely disregarded in research into speech intelligibility, and there is no agreement about the impact that voice quality has on intelligibility. Listener variables that were most indicated to affect intelligibility were familiarity with the speaker and the hearing status of the listener. Familiarity with the speaker has been found to be a contributing factor to intelligibility by others (Boonen et al., 2019; Borrie et al., 2012; D'Innocenzo et al., 2006; Flipsen Jr, 1995; Kent et al., 1994; McGarr, 1983; Miller, 2013; Munson et al., 2012; Tjaden & Liss, 1995). The variables that were indicated to have least effect were the age of the listener, having a shared cultural background with the speaker, and the physical/emotional state of the listener.

Overarching Themes

Evidence from this survey overwhelmingly supported the hypothesis that there is not overall agreement among pediatric SLPs on how to best measure speech intelligibility. While some methods of measurement are more popular than others, each measure presented had some portion of respondents indicate that they used it in clinical practice at least some percentage of the time. Questions about measurement procedures often included an option to add other methods that were not listed, and many questions received additional responses, demonstrating the variation among participants. There was disagreement between respondents regarding the adequacy of tools that are currently available to measure speech intelligibility, with 53.24% indicating the current tools are adequate and 46.76% indicating the current tools are not adequate. Clinical experience, clinical judgement, training, and current assessments were given as adequate measures by some, while others responded that subjective measures of intelligibility are not adequate and a more standard or objective process is needed.

One area that participants did agree on was the use of computerized analysis. The majority of participants (96.43%) generally agreed that they do not use computerized analysis for measuring speech intelligibility. Computerized assessment is a fairly new venture, though in recent years several studies regarding the topic have been done. However, many of these studies have developed computerized analyses for specific populations. Stelzle et al. (2010) developed a computerized analysis for edentulous patients, Kim et al. (2015) and Middag et al. (2009) developed analyses for those with pathological speech, and Ling et al. (2013) explored computerized analysis for children with cleft palate. This suggests that clinicians may not have access to computerized analyses for children that do not have specific disorders. There may be other barriers to implementation of computerized systems as well, such as cost and maintenance, and tools requiring the internet may not be feasible in rural areas with limited broadband access. The high number of clinicians that reported clinical judgement as a sufficient tools for intelligibility analysis may indicate overconfidence in perceptual abilities of clinicians. This may translate to a lack of motivation to explore new methods for measurement.

A few respondents commented that they did not believe objective measures of speech intelligibility were feasible, as it is inherently a subjective process that will be judged in real-world scenarios. It is acknowledged that the auditory-perceptual component of intelligibility assessment cannot be objective and free from bias; however, it is important to note that a standardized process to determine speech intelligibility could improve reliability of the procedure. Standardization can refer to a process that outlines how to go about measurement and can be taught to clinicians on a large scale so that similar findings are given across the board. This has been done with other perceptual measures successfully, such as with the Grade, Roughness, Breathiness, Asthenia, and Strain (GRBAS) scale (Hirano, 1981). The measurement of voice involves both subjective and objective measures, with auditory-perceptual judgement being a main part of clinical judgement (Barsties & De Bodt, 2015). However, tools such as the GRBAS scale give clinicians a standard instrument and is widely used to complete perceptual judgements (Hakkesteegt et al., 2010). The GRBAS scale has been found to have high levels of intra- and inter-judge consensus and its clinical use is supported by research (Dursun et al., 2008; Karnell et al., 2007; Nemr et al., 2012). The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) is another standard instrument developed for assessment of voice that uses conversational speech as well as standard sentences as samples for the clinician to assess (Kempster et al., 2009). Another example is the Modified Barium Swallow Impairment Profile (MBSImP), which was developed as a standard approach to interpreting modified barium swallow studies (MBSS) (Martin-Harris et al., 2008). Following training in the method and at least three years of experience post-certification, the method was found to have high inter- and intra-rater reliability. The MBSImP was also found to correlate with intake recommendations for modified diets made by SLPs, giving the approach clinical validity in that area (Martin-Harris et al., 2008).

The need for a standard approach for intelligibility measurement is evident in participant responses about the functional uses for intelligibility data. Regarding reasons for measuring intelligibility, more than half of respondents indicated three different reasons, and 40% indicated a fourth. The use of intelligibility data once collected also had varied responses, with four different uses being indicated by over 65% of respondents. It is clear that intelligibility measurement is done for specific purposes and has multiple clinical applications for pediatric SLPs. Baseline measurements, eligibility criteria, tracking of progress, and treatment planning were all indicated as reasons clinicians measure intelligibility. It is surprising that a measure with so many functional implications for a child has not yet been standardized with specific guidelines for SLPs to follow. The lack of a standard process for intelligibility measurement may lead to discrepancies within and between clinicians, decreased validity of assessment results, and a lower quality of care due to an insufficient treatment plan based on inaccurate assessment results (Martin-Harris et al., 2008). Additionally, it is difficult to appraise the sensitivity and specificity of intelligibility determination without a standard approach.

Also of note is the fact that there is disagreement among SLPs regarding the variables that are used to measure speech intelligibility. Responses to the speaker and listener variables ranged widely. Factors such as dialect, shared cultural background, voice quality, age of the listener, and physical/emotional state of both the speaker and listener were rated to have “little” effect on intelligibility. However, when asked to state what tools are needed in the field, more background knowledge of cultural background, dialect, and community was stated several times by respondents.

The most commonly reported measurement modality used in the survey was severity measures. While severity measures are easy to use and give clear results, they are not a true measure of the child's intelligibility. A commonly used severity measure is PCC, which many respondents stated using. Besides not being a true measure of intelligibility, research indicates that PCC is not well understood by pediatric clinicians. Dale et al. (2020) found that pediatric clinicians reported limited exposure to PCC both academically and clinically, inconsistent knowledge of PCC rules, and lack of confidence using PCC. Therefore, the fact that the most commonly reported measure being used by clinicians was severity measures highlights the need for a new process for measurement of speech intelligibility that is easy to use.

Unexpected Findings

One unexpected finding was the relatively even split between SLPs that believe we currently have adequate tools and SLPs that believe we do not currently have adequate tools. Due to the limited research in this area, it was not anticipated that such a high number of SLPs would feel that the tools that currently exist are sufficient to accurately measure intelligibility. Given the heavy emphasis for inclusion of measures of intelligibility in most pediatric SLP assessments, it may have been assumed that current practices are sufficient despite the lack of standardization within and between practice settings.

It was also unexpected that some SLPs would state that they feel objective measures would not be useful for intelligibility measurement. Some respondents directly stated the need for an objective way to measure intelligibility, while some stated that objective measures for intelligibility are not effective.

Strengths

Speech intelligibility is a functional outcome that is used regularly in clinical practice, yet an investigation into the specific measures that pediatric SLPs use for measurement has never been done before to our knowledge. While methods of measurement have been described frequently, the clinical use of those measures has not been investigated. Data collected from this survey contributes to our understanding of clinical practice patterns, which is necessary to eventually implement improved methods. This survey gives a new way to view speech intelligibility and lays groundwork for future investigations to inform new methods.

Limitations

Surveys are inherently biased due to being limited to the included questions. However, this was mitigated by the open response option on many questions, allowing participants to add their own input. Recruitment for the survey was also limited to those who happened to see the survey advertised in certain online groups. However, multiple groups were identified for survey dissemination. They ranged in platform (social media, ASHA community groups, ASHA Special Interest Groups), number of members, and main interest of members (population served, work setting, etc.). In order to reach as many participants as possible, the survey link was re-posted in the groups a week after the original post.

The respondents to the survey largely fell into the same demographic category – white women. 97.86% identified as female and 87.68% identified as white. However, about 95.5% of ASHA members are female and only 8.5% of ASHA members identify as racial minorities, so the overall demographics of the respondents roughly matched the demographics of ASHA members (Association, 2021).

Participants were asked if they believed current tools for intelligibility measurement were adequate or not. Participants that indicated "yes" were asked what current tools they believe are adequate, and participants that indicated “no” were asked what tools are needed. An error in survey design resulted in the participants that indicated “yes” being asked to list current tools they believe are adequate and what tools are needed. However, the error still allowed participants to respond to the appropriate question and did not significantly alter the survey.

There are a number of variables this survey did not address. Glossing of utterances assists the clinician in determining the pattern of the child’s speech and analyzing how far from the target the utterance truly was. This was not addressed in the survey. The role of context in evaluation and participants’ specific reasoning for using different measures were also not specifically addressed. These variables were beyond the scope of the current survey but should be addressed in future research.

Future Directions

Pediatric speech intelligibility currently lacks a standard process for measurement and SLPs currently use many different methods. The same client may receive different judgements of intelligibility based on the clinician that they see. Because intelligibility is such an integral part of effective communication, the variation in measurement techniques is beyond what is clinically acceptable. Intelligibility is used by clinicians for a variety of reasons that have functional implications for the client. Establishing a process that is simple, easy to teach, and can be implemented in many different settings would allow for more consistent approaches and measurements across the board.

Some SLPs also reported lack of awareness of available tests or measures that were given as answer choices throughout the survey. Many SLPs also stated that more training in intelligibility measurement was needed for SLPs. More comprehensive training in intelligibility and exposure to current tools is needed. The implementation of a standardized process would also add to clinician training, as the method could be taught to graduate student clinicians and practicing clinicians alike.

SLPs rated factors such as dialect, shared cultural background, and physical/emotional state as having little effect on intelligibility. Many also reported that the COVID-19 pandemic has resulted in heavier reliance on sources outside the clinician, such as family members, teachers, and observations of speech in different environments. These are valuable components of intelligibility measurement that may have not been considered a requirement in the past, but that have been necessitated by COVID-19. These factors can give the clinician rich information about the client and their intelligibility outside the clinic setting, and deeper understanding of the client's dialect, cultural background, and physical/emotional state can give context to any assessment that is performed. More research into how speaker and listener variables impact intelligibility is needed.

Conclusions

Results from this survey suggested that there is not agreement among pediatric SLPs on how to measure pediatric speech intelligibility, matching results from a previous investigation (Hustad et al., 2019). In addition, the most frequently reported measures were severity measures and estimation. Disagreement also exists regarding if adequate tools are available to SLPs or not; as well as what tools are needed. Lastly, there was variation in what variables relative to the speaker and listener pediatric SLPs believe affect intelligibility.

Results indicate that more research is needed in the area of pediatric speech intelligibility measurement. The creation of a standard approach to measuring speech intelligibility that is easy to implement and yields consistent result is needed. SLPs also specified that time is a limitation on effective intelligibility measurement, pointing to the need for a standard approach to be able to be completed quickly.

Many respondents also indicated that SLPs need more training in the broad area of intelligibility. Specific areas within intelligibility were also indicated, from more background knowledge about the client to increased knowledge about the current tools available.

Speech intelligibility is a crucial component of effective communication and is put to the test each time the individual attempts to verbally communicate. The development of a standard tool or approach to intelligibility measurement will create consistency across treatment settings and clinicians and add to the validity of intelligibility assessment results for individuals.

References

Abur, D., Enos, N. M., & Stepp, C. E. (2019). Visual analog scale ratings and orthographic transcription measures of sentence intelligibility in Parkinson's disease with variable listener exposure. *American Journal of Speech-Language Pathology, 28*(3), 1222-1232.

Allison, K. M. (2020). Measuring speech intelligibility in children with motor speech disorders. *Perspectives of the ASHA Special Interest Groups, 5*(4), 809-820.

Association, A. S.-L.-H. (2021). *Profile of ASHA members and affiliates, year-end 2020*. Retrieved 11-26-2021

Bacon, V. J. (1995). *Validity and efficiency of the check-slash transcription method for measuring intelligibility* (Publication Number 4987) Portland State University].

Barsties, B., & De Bodt, M. (2015). Assessment of voice quality: Current state-of-the-art. *Auris Nasus Larynx, 42*(3), 183-188.

Beijer, L., Clapham, R., & Rietveld, A. (2012). Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20*(2), 17-35.

Bleile, K. (2002). Evaluating articulation and phonological disorders when the clock is running. *American Journal of Speech-Language Pathology, 11*(3), 243-249.

Boonen, N., Kloots, H., Verhoeven, J., & Gillis, S. (2019). Can listeners hear the difference between children with normal hearing and children with a hearing impairment? *Clinical Linguistics & Phonetics, 33*(4), 316-333.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27*(7-8), 1039-1055.

Carley, B., & Hustad, K. (2021). Speech Intelligibility in Preschool Children: Is There Something Special about Parents as Listeners? American Speech-Language-Hearing Association National Convention, Washginton, D.C.

Castellanos, I., Kronenberger, W. G., Beer, J., Henning, S. C., Colson, B. G., & Pisoni, D. B. (2014). Preschool speech intelligibility and vocabulary skills predict long-term speech and language outcomes following cochlear implantation in early childhood. *Cochlear Implants International, 15*(4), 200-210. https://doi.org/10.1179/1754762813Y.0000000043

Chin, S. B., Bergeson, T. R., & Phan, J. (2012). Speech intelligibility and prosody production in children with cochlear implants. *Journal of Communication Disorders, 45*(5), 355-366.

Chiu, Y.-F., & Neel, A. (2020). Predicting intelligibility deficits in Parkinson's disease with perceptual speech ratings. *Journal of Speech, Language, and Hearing Research, 63*(2), 433-443.

Connolly, J. H. (1986). Intelligibility: A linguistic view. *British Journal of Disorders of Communication, 21*(3), 371-376.

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20*(9), 659-675.

Dale, E. W., Plumb, A. M., Sandage, M. J., & Plexico, L. W. (2020, 08/01/). Speech-language pathologists' knowledge and competence regarding Percentage of Consonants Correct. *Communication Disorders Quarterly, 41*(4), 222-230. http://dx.doi.org/10.1177/1525740119853806

Dorsey, M., Yorkston, K., Beukelman, D., & Hakel, M. (2007). *Speech Intelligibility Test for Windows.* In Institute for Rehabilitation Science and Engineering at Madonna.

Dursun, G., Boynukalin, S., Ozgursoy, O. B., & Coruh, I. (2008). Long-term results of different treatment modalities for glottic insufficiency. *American Journal of Otolaryngology, 29*(1), 7-12.

Farquharson, K., & Tambyraja, S. R. (2019). Describing how school-based SLPs determine eligibility for children with speech sound disorders. Seminars in speech and language,

Flipsen Jr, P. (1995). Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders, 28*(1), 3-19.

Gordon-Brannan, M., & Hodson, B. W. (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology, 9*(2), 141-150.

Hakkesteegt, M. M., Brocaar, M. P., & Wieringa, M. H. (2010). The applicability of the dysphonia severity index and the voice handicap index in evaluating effects of voice therapy and phonosurgery. *Journal of Voice, 24*(2), 199-205. https://www.jvoice.org/article/S0892-1997(08)00108-2/fulltext

Hirano, M. (1981). Psycho-acoustic evaluation of voice. In G. E. Arnold, F. Winckel, & B. D. Wyke (Eds.), *Disorders of Human Communication 5: Clinical Examination of Voice* (pp. 81-84). Springer-Verlag.

Hodge, M., Daniels, J., & Gotzke, C. (2009). TOCS+ intelligibility measures (Version 5.3)[computer software]. *Edmonton, Canada: University of Alberta*.

Hodge, M. M., & Gotzke, C. L. (2014). Construct-related validity of the TOCS measures: Comparison of intelligibility and speaking rate scores in children with and without speech disorders [Article]. *Journal of Communication Disorders, 51*, 51-63. https://doi.org/10.1016/j.jcomdis.2014.06.007

Hodson, B. W., & Paden, E. P. (1981). Phonological processes which characterize unintelligible and intelligible speech in early childhood. *Journal of Speech and Hearing Disorders, 46*(4), 369-373.

Hustad, K. C., Mahr, T., Natzke, P. E., & Rathouz, P. J. (2020). Development of speech intelligibility between 30 and 47 months in typically developing children: A cross-sectional study of growth. *Journal of Speech, Language, and Hearing Research, 63*(6), 1675-1687.

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

Hustad, K. C., Sakash, A., Natzke, P. E., Broman, A. T., & Rathouz, P. J. (2019). Longitudinal growth in single word intelligibility among children with cerebral palsy from 24 to 96 months of age: Predicting later outcomes from early speech production. *Journal of Speech, Language, and Hearing Research, 62*(6), 1599-1613.

Hustad, K. C., Schueler, B., Schultz, L., & Duhadway, C. (2012). Intelligibility of 4-year-old children with and without cerebral palsy [Article]. *Journal of Speech, Language & Hearing Research, 55*(4), 1177-1189. https://doi.org/10.1044/1092-4388(2011/11-0083)

Jiao, Y., LaCross, A., Berisha, V., & Liss, J. (2019). Objective intelligibility assessment by automated segmental and suprasegmental listening error analysis. *Journal of Speech, Language, and Hearing Research, 62*(9), 3359-3366. https://doi.org/10.1044/2019_JSLHR-S-19-0119

Karnell, M. P., Melton, S. D., Childes, J. M., Coleman, T. C., Dailey, S. A., & Hoffman, H. T. (2007). Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Journal of Voice, 21*(5), 576-590.

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*, 124-132.

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.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482-499.

Khan, L. M. (2002). The sixth view: Assessing preschoolers' articulation and phonology from the trenches. *American Journal of Speech-Language Pathology, 11*(3), 250-254.

Khwaileh, F. A., & Flipsen, P., Jr. (2010). Single word and sentence intelligibility in children with cochlear implants. *Clinical Linguistics & Phonetics, 24*(9), 722-733. https://doi.org/10.3109/02699206.2010.490003

Kim, J., Kumar, N., Tsiartas, A., Li, M., & Narayanan, S. S. (2015). Automatic intelligibility classification of sentence-level pathological speech. *Computer speech & language, 29*(1), 132-144.

Kim, Y., & Kuo, C. (2012). Effect of level of presentation to listeners on scaled speech intelligibility of speakers with dysarthria. *Folia Phoniatrica et Logopaedica, 64*(1), 26-33.

Lagerberg, T. B., Lam, J., Olsson, R., Abelin, Å., & Strömbergsson, S. (2019). Intelligibility of children with speech sound disorders evaluated by listeners with Swedish as a second language. *Journal of Speech, Language, and Hearing Research, 62*(10), 3714-3727.

Lee, A., & FitzGibbon, O. (2008). Speech-language pathologists' experience on perceptual evaluation of speech disorders: A survey. Ireland: ASHA Convention,

Lee, J., Hustad, K. C., & Weismer, G. (2014). Predicting speech intelligibility with a multiple speech subsystems approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 57*(5), 1666-1678.

Ling, H., Jing, Z., Qi, L., Heng, Y., & Lech, M. (2013). Automatic evaluation of hypernasality and speech intelligibility for children with cleft palate. *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)*, Melbourne, VIC, Australia.

Lousada, M., Jesus, L. M., Hall, A., & Joffe, V. (2014). Intelligibility as a clinical outcome measure following intervention with children with phonologically based speech–sound disorders. *International journal of language & communication disorders, 49*(5), 584-601.

Marsh-Richard, D. M., Hatzis, E. S., Mathias, C. W., Venditti, N., & Dougherty, D. M. (2009). Adaptive Visual Analog Scales (AVAS): a modifiable software program for the creation, administration, and scoring of visual analog scales. *Behavior Research Methods, 41*(1), 99-106.

Martin-Harris, B., Brodsky, M. B., Michel, Y., Castell, D. O., Schleicher, M., Sandidge, J., Maxwell, R., & Blair, J. (2008). MBS measurement tool for swallow impairment—MBSImp: Establishing a standard. *Dysphagia, 23*(4), 392-405.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech, Language, and Hearing Research, 26*(3), 451-458.

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The intelligibility in context scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research, 55*(2), 648-656.

Middag, C., Martens, J.-P., Van Nuffelen, G., & De Bodt, M. (2009). Automated intelligibility assessment of pathological speech using phonological features. *EURASIP Journal on Advances in Signal Processing, 2009*, 1-9.

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders, 48*(6), 601-612.

Mullen, R., & Schooling, T. (2010). The National Outcomes Measurement System for pediatric speech-language pathology. *Language, Speech, and Hearing Services in Schools, 41*(1), 44-60.

Munro, M. J. (2010). *Intelligibility: Buzzword or buzzworthy?* Pronunciation in Second Language Learning and Teaching Conference, Ames, IA: Iowa State University.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306.

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology, 21*(2), 124-139.

Nemr, K., Simões-Zenari, M., Cordeiro, G. F., Tsuji, D., Ogawa, A. I., Ubrig, M. T., & Menezes, M. H. M. (2012, 11/01/November 2012). GRBAS and Cape-V scales: High reliability and consensus when applied at different times [Article]. *Journal of Voice, 26*(6), 812-812. https://doi.org/10.1016/j.jvoice.2012.03.005

Osberger, M. J. (1994). Speech intelligibility of children with cochlear implants. *Volta Review, 96*(5), 169-180.

Paja, M. S., & Falk, T. H. (2012). Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech. Thirteenth Annual Conference of the International Speech Communication Association,

Pascoe, M., Stackhouse, J., & Wells, B. (2006). *Persisting speech difficulties in children: children's speech and literacy difficulties*. John Wiley & Sons.

Pennington, L., Mjøen, T., Andrada, M. d. G., & Murray, J. (2011). Viking speech scale.

Schiavetti, N. (1992). Scaling Procedures for the Measurement of Speech Intelligibility. In R. D. Kent (Ed.), *Intelligibility in Speech Disorders: Theory, Measurement and Management* (pp. 11-34). John Benjamins Publishing Company. https://doi.org/10.1075/sspcl.1.02sch

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders, III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47*(3), 256-270. https://doi.org/10.1044/jshd.4703.256

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.

Stelzle, F., Ugrinovic, B., Knipfer, C., Bocklet, T., NÖTh, E., Schuster, M., Eitner, S., Seiss, M., & Nkenke, E. (2010, 03/01/). Automatic, computer-based speech assessment on edentulous patients with and without complete dentures – preliminary results. *Journal of Oral Rehabilitation, 37*(3), 209-216. https://doi.org/10.1111/j.1365-2842.2009.02047.x

Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson’s disease and multiple sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research, 55*(4), 1208-1219. https://doi.org/10.1044/1092-4388(2011/11-0048)

Tastle, W. J., & Wierman, M. J. (2006). An information theoretic measure for the evaluation of ordinal scale data. *Behavior Research Methods, 38*(3), 487-494. https://doi.org/10.3758/BF03192803

Tjaden, K., & Liss, J. M. (1995). The influence of familiarity on judgments of treated speech. *American Journal of Speech-Language Pathology, 4*(1), 39-48.

Van Nuffelen, G., Middag, C., De Bodt, M., & Martens, J.-P. (2009). Speech technology-based assessment of phoneme intelligibility in dysarthria. *International Journal of Language & Communication Disorders, 44*(5), 716-730.

Weston, A. D., & Shriberg, L. D. (1992). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. *Journal of Speech, Language, and Hearing Research, 35*(6), 1316-1332.

Whitehill, T. L., Lee, A. S. Y., & Chun, J. C. (2002). Direct magnitude estimation and interval scaling of hypernasality [Article]. *Journal of Speech, Language & Hearing Research, 45*(1), 80-88. https://doi.org/10.1044/1092-4388(2002/006)

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*(6), 499-512.

Yorkston, K. M., Dowden, P. A., & Beukelman, D. R. (1992). Intelligibility measurement as a tool in the clinical management of dysarthric speakers. *Intelligibility in Speech Disorders: Theory, Measurement, and Management*, 265-286.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*(1), 55-66.

Appendix A

Survey Questions

Pediatric Intelligibility Assessment Survey

Start of Block: Informed Consent

Information letter
NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.

INFORMATION LETTER for a Research Study entitled "Survey of Pediatric Speech-Language Pathologist's (SLP's) Methods for Measuring Speech Intelligibility"

You are invited to participate in a research study to understand and identify which tools and methods are currently being used by practicing speech-language pathologists (SLPs) to measure speech intelligibility in children. The study is being conducted by Abigail Bennett, graduate research assistant, under the direction of Dr. Mary J. Sandage, Associate Professor in the Auburn University Department of Speech, Language, and Hearing Sciences. You are invited to participate because you currently practicing as an SLP who works with the pediatric population, you practice in the United States, United States territories, or United States military bases, and you are 18 years of age or older.

What will be involved if you participate? If you decide to participate in this study, you will be asked to answer some questions about which tools and methods you use. Your total time commitment will be approximately 15 minutes.

Are there any risks or discomforts? No.

Are there any benefits to yourself or others? Collecting data on what speech intelligibility measurements are currently being used by pediatric SLPs will help inform further research into this area. In addition, surveying the current methods used will create a deeper understanding of our field for speech-language pathologists and those outside the field alike.

Will you receive compensation for participating? No compensation will be offered, and your participation will be completely voluntary.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. You have the right to withdraw at any point during the survey, for any reason, and without any prejudice. In order to withdraw from this study, exit your web browser.

Any data obtained in connection with this study will remain anonymous. We will protect your privacy and the data you provide by not collecting any identifiable information. Your survey responses will be kept secure and completely confidential in the Auburn BOX and Auburn secure servers. Information collected through your participation may be used to fulfill an educational requirement, published in a professional journal, and/or presented at a professional meeting.

If you have questions about this research, please contact the Principal Investigator, Abigail Bennett, at alb0135@auburn.edu or one of the faculty advisors: Dr. Mary J. Sandage at sandamj@auburn.edu or Dr. Marisha Speights Atkins at marisha.speights@northwestern.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE IF YOU WANT TO PARTICIPATE IN THIS RESEARCH PROJECT. IF YOU DECIDE TO PARTICIPATE, THE DATA YOU PROVIDE WILL SERVE AS YOUR AGREEMENT TO DO SO. THIS LETTER IS YOURS TO KEEP.

By clicking the button below, you acknowledge:
-Your participation in the survey is voluntary.
-You are at least 18 years of age.
-You are currently practicing as an SLP who works with the pediatric population.
-You are aware that you may choose to terminate your participation at any time for any reason.

Please note that this survey will be best displayed on a laptop or desktop computer. Some features may be less compatible for use on a mobile device.

○ I consent, begin the survey

○ I do not consent, I do not wish to participate

Skip To: End of Survey If Information letter NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATE... = I do not consent, I do not wish to participate

End of Block: Informed Consent

Start of Block: Qualification Questions

Qualification Are you 18 years of age or older?

◯ Yes

◯ No

Skip To: End of Survey If Are you 18 years of age or older? = No

Qualification Are you a Speech-Language Pathologist who works with the **pediatric** population?

◯ Yes

◯ No

Skip To: End of Survey If Are you a Speech-Language Pathologist who works with the pediatric population? = No

Qualification Do you currently reside in the United States, a territory of the United States, or a United States military base?

◯ Yes

◯ No

Skip To: End of Survey If Do you currently reside in the United States, a territory of the United States, or a United State... = No

End of Block: Qualification Questions

Start of Block: Personal Demographics

Q1 What ethnicity do you identify as?

◯ Hispanic or Latino

◯ Not Hispanic or Latino

◯ Add another ____

Q2 Please select what race(s) you identify as. Indicate all that apply.

▢ White- A person having origins in any of the original peoples of Europe, the Middle East, or North Africa

▢ Black or African American - A person having origins in any of the Black racial groups of Africa.

▢ Native American or Alaska Native - A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.

▢ Asian - A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent

▢ Native Hawaiian or Other Pacific Islander -A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands

▢ Add as many as apply: ____

Q3 What gender do you identify as?

○ Male

○ Female

○ Non-Binary

○ Add another ____

○ Prefer not to answer

Q4 Is American English your first language or one of your first languages if you are bilingual or multilingual?

○ Yes

○ No (please list your language(s) in the order you acquired them)
____

Q5 What is your age?

____

End of Block: Personal Demographics

Start of Block: Demographics

Q6 Select your degree(s), and enter the year each was obtained (leave blank if it does not apply).

☐ Masters ____

☐ PhD ____

☐ Other ____

Q7 Check the location(s) that best describe you work setting. Select all that apply.

- ☐ Birth to 3 program
- ☐ Preschool
- ☐ Elementary School
- ☐ Middle School
- ☐ High School
- ☐ University clinic
- ☐ Private practice
- ☐ Headstart
- ☐ Hospital - Inpatient
- ☐ Hospital - Outpatient
- ☐ Other ____

Q8 In which state do you currently practice?

▼ Alabama ... United States Military Base

Q9 How many years have you practiced speech-language pathology? (Enter the number of years)

____

Q10 What percentage (%) of children currently on your caseload are receiving services for a speech sound disorder? (Enter the percentage)

____

End of Block: Demographics

Start of Block: Practices in Assessment of Speech Sound Disorders

Q11 Select the areas that describe the training you have received related to intelligibility assessment. (Select all that apply)

▢ Undergraduate level course(s)

▢ Graduate level course(s)

▢ Convention(s)

▢ Workshop(s)

▢ Independent Journal Study

▢ Other ____

▢ None

Q12 What percentage (%) of children on your caseload require an articulation or phonology assessment?

____

Q13 Do you use computerized analysis procedures for **articulation/phonological assessments**? If yes, please provide the title.

○ No

○ Yes ____

Q14 Do you use any of the following computerized analysis procedures for **intelligibility assessments**? Select all that apply.

▢ Acoustic

▢ Automated counters of correct syllables

▢ Automated counters of correct words

▢ Automatic speech recognition (ASR)

▢ Other ____

▢ I do not use computerized analysis procedures for intelligibility assessments.

Page Break

Q15 In your experience, what is the most typical parental involvement during the assessment process? Select all that apply.

- ▢ Accompanies child during assessment
- ▢ Completes developmental checklists
- ▢ Completes Intelligibility in Context Scale
- ▢ Completes mail-in case history form
- ▢ Completes survey
- ▢ Live interview
- ▢ Parent/child play during assessment
- ▢ Phone interview
- ▢ No parental interaction
- ▢ Other (please describe)

____

End of Block: Practices in Assessment of Speech Sound Disorders

---

Start of Block: Estimation of Speech Intelligibility

Q16 How often do you use subjective estimation to measure speech intelligibility?

- ◯ Never
- ◯ Infrequently
- ◯ Sometimes
- ◯ Often
- ◯ Always

Skip To: End of Block If How often do you use subjective estimation to measure speech intelligibility? = Never

Q17 When you estimate speech intelligibility, what method do you use?

- ◯ Estimate speech intelligibility from a speech sample
- ◯ Estimate from articulation/phonological assessment scores
- ◯ Another way ____

End of Block: Estimation of Speech Intelligibility

Start of Block: Assessment of Speech Intelligibility

Q18 Which of the following methods do you use to collect speech intelligibility data? Select all that apply.

- ☐ Formal Instrument of Speech Severity (ex. DEAP)
- ☐ Formal Instrument of Intelligibility (ex. TOCS+)
- ☐ Continuous speech sampling
- ☐ Single word sampling
- ☐ Observations in the child's natural environment
- ☐ Other ____
- ☐ None

Page Break

Q19 Choose the **articulation and phonology assessment(s)** that you use to determine **intelligibility**. Select all that apply.

- ☐ Assessment Link between Phonology and Articulation-Revised (ALPHA)
- ☐ Assessment of Phonological Processes-Revised (APP-R)
- ☐ Arizona Articulation and Phonology Scale 4 (AAPS4)
- ☐ Bankson-Bernthal Test of Phonology (BBTOP)
- ☐ Clinical Assessment of Articulation and Phonology (CAAP)
- ☐ Comprehensive Test of Phonological Processing (CTOPP)
- ☐ Contextual Probes of Articulation Competence (CPAC)
- ☐ Deep Test of Articulation (DTA)
- ☐ Diagnostic Evaluation of Articulation and Phonology: Articulation Assessment (DEAP)
- ☐ Fisher-Logemann Test of Articulation Competence (F-LTAC)
- ☐ Glaspey Dynamic Assessment of Phonology (GDAP)
- ☐ Goldman-Fristoe Test of Articulation-3rd Edition (GFTA-3): Sounds in Words
- ☐ Goldman-Fristoe Test of Articulation-3rd Edition (GFTA-3): Sounds in Sentences
- ☐ Hodson Assessment of Phonological Patterns-3rd Edition (HAPP-3)
- ☐ Khan-Lewis Phonological Analysis-3rd Edition (KLPA-3)

- [ ] Linguisystems Articulation Test (LAT)
- [ ] McDonald-Deep Test of Articulation
- [ ] Photo Articulation Test-3rd Edition (PAT-3)
- [ ] Practical Test of Articulation and Phonology (PTAP)
- [ ] Secord Contextual Articulation Tests (S-CAT)
- [ ] Smit-Hand Articulation and Phonology Evaluation (SHAPE)
- [ ] Structured Photographic Articulation Test III (SPAT-D III)
- [ ] Templin-Darley Tests of Articulation-2nd Edition
- [ ] Test of Phonological Awareness Skills
- [ ] Toddler Phonology Test (TPT)
- [ ] Weiss Comprehensive Articulation Test (WCAT)
- [ ] Other ____
- [ ] I do not use these assessments to determine intelligibility.

Page Break

Q20 Choose the **measure(s) of speech intelligibility** that you use during assessments. Select all that apply. (List from Kent et. al, 1994)

- [ ] Articulation Competence Index (ACI)
- [ ] ASHA NOMS
- [ ] Assessment of Intelligibility in Dysarthric Speech (AIDS)
- [ ] CID (Central Institute for the Deaf) Word SPINE (SPeech INtelligibility Evaluation)
- [ ] CID Picture SPINE
- [ ] Children's Speech Intelligibility Test (CSIT)
- [ ] Goldman-Fristoe Test of Articulation-3rd Edition (GFTA-3): Intelligibility Rating
- [ ] Ling's Phonologic and Phonetic Level Speech Evaluation (PPLSE)
- [ ] Meaningful Use of Speech Scale (MUSS)
- [ ] NTID Rating Scale
- [ ] Percentage of Consonants Correct (PCC)
- [ ] Preschool Speech Intelligibility Measure (P-SIM)
- [ ] Quantitative Rating of Performance (QRP)
- [ ] Speech Pattern Contrast (SPAC) Test
- [ ] Test of Children's Speech (TOCS)

☐ Weiss Intelligibility Test (WIT)

☐ Another measure not listed.

____

☐ I don't use any of these.

Q21 Indicate how often you use the following increments of speech to measure speech intelligibility.

	Never	Infrequently	Sometimes	Often	Always
Syllables	○	○	○	○	○
Words	○	○	○	○	○
Phrases	○	○	○	○	○
Sentences	○	○	○	○	○
Conversation	○	○	○	○	○
Read speech/Read sentences	○	○	○	○	○
Other	○	○	○	○	○

Q22 If you use speech samples, how do you calculate speech intelligibility using those speech samples?

____

End of Block: Assessment of Speech Intelligibility

Start of Block: Consideration of Communication Dynamics

Q23 To what extent do you believe each of the following factors relative to the **speaker** affect speech intelligibility?

	Not at all	Very Little	Little	A Fair Amount	Great
Resonance (Hypernasality)	○	○	○	○	○
Voice Quality (Roughness, Hoarseness)	○	○	○	○	○
Age	○	○	○	○	○
Linguistic factors (word choice and grammar)	○	○	○	○	○
Number of speech sound errors	○	○	○	○	○
Type of speech sound errors	○	○	○	○	○
Frequency of speech sound errors	○	○	○	○	○
Dialect Spoken	○	○	○	○	○
Familiarity with listener	○	○	○	○	○
Familiarity with content of conversation	○	○	○	○	○
Gestures and facial expressions	○	○	○	○	○
Background noise	○	○	○	○	○
Shared cultural background with listener	○	○	○	○	○
Physical/Emotional state (hunger, tiredness, anxiety,	○	○	○	○	○

etc.)

Q24 What other variables relative to the **speaker** do you believe affect intelligibility?

____

Q25 To what extent do you believe each of the following factors relative to the **listener** affect speech intelligibility?

	Not at all	Very little	Little	A Fair Amount	Great
Familiarity with speaker	○	○	○	○	○
Familiarity with dialect of speaker	○	○	○	○	○
Familiarity with content of conversation	○	○	○	○	○
Communication environment (one-on-one; group setting, etc.	○	○	○	○	○
Background noise	○	○	○	○	○
Age	○	○	○	○	○
Shared cultural background with speaker	○	○	○	○	○
Physical/Emotional state (hunger, tiredness, anxiety, etc.)	○	○	○	○	○
Hearing status	○	○	○	○	○

Q26 What other variables relative to the **listener** do you believe affect intelligibility?

____

End of Block: Consideration of Communication Dynamics

Start of Block: Use of Specific Measures

Instruction Refer to these definitions for the following question.

Word Recognition/Transcript: Listeners orthographically transcribe the words they thought they heard or choose them from a list. The listener is someone who is NOT familiar with the speaker.

Multiple Choice: A form of Word Recognition/Transcript where listeners are given a list of similar-sounding words and must choose which word they heard. There is an option for "other" with a write-in box, "all of the above," and "no response."

Forced Choice: A form of Word Recognition/Transcript where listeners are given a list of similar-sounding words and must choose which word they heard. They must choose a word that is on the list.

Q27 How often do you use each method to measure speech intelligibility?

	Never	Infrequently	Sometimes	Often	Always
Word Recognition/Transcript Measures	○	○	○	○	○
Multiple Choice	○	○	○	○	○
Forced Choice	○	○	○	○	○

Page Break

Instruction Refer to these definitions for the following question.

Ordinal Scales: Ordinal scales employ category hierarchies, and the person completing the scale must choose where the speaker best fits along the hierarchy. Likert scales are an example of ordinal scale measures. (ex. 1 = never, 2 = infrequently, 3 = sometimes, 4 = often, 5 = always)

Visual Analog Scales (VAS): The listener hears speech from the speaker and then marks on a scale where they think the speaker's intelligibility would fall. The scale is typically an 100mm line. VAS are usually anchored at either end by the extremes of the characteristic being measured (for example, "unintelligible" on one end and "completely intelligible" on the other end)

Direct Magnitude Estimation (DME): There is no fixed minimum or maximum value on the scale that users must stay between. Listeners are encourage to use the entire length of the scale from left to right with the rating value increasing or decreasing as the scale marker is moved. Another variation of DME requires listeners to listen to an example of a speaker who typically falls in the middle range of intelligibility. Each speech sample is then rated relative to the example provided.

Q28 How often do you use each method to measure speech intelligibility?

	Never	Infrequently	Sometimes	Often	Always
Ordinal Scales	○	○	○	○	○
Visual Analog Scales (VAS)	○	○	○	○	○
Direct Magnitude Estimation (DME)	○	○	○	○	○

Page Break

Instruction Refer to these definitions for the following question.

Severity Assessment Measures: Rate the severity of intelligibility. Percent Consonants Correct is a popular severity assessment measure.

Acoustic Based Measures: Uses speech parameters such as vowel space, speech rate, voice onset time, number of syllables, syllable boundaries, and speech duration to measure intelligibility.

Automated Measures: An objective, computerized tool used to measure speech intelligibility. An example of an automated measure is the Test of Children's Speech (TOCS+).

Q29 How often do you use each method to measure speech intelligibility?

	Never	Infrequently	Sometimes	Often	Always
Severity Assessment Measures	○	○	○	○	○
Acoustic Based Measures	○	○	○	○	○
Automated Measures	○	○	○	○	○

End of Block: Use of Specific Measures

Start of Block: Rationale for Measuring Speech Intelligibility

Q30 Why do you measure intelligibility? Select all that apply.

- ▢ It is required for eligibility criteria
- ▢ It is required for another reason (please enter reason) ____
- ▢ To aid in treatment planning
- ▢ To track progress over time
- ▢ To add to my understanding of the client
- ▢ Other ____

Q31 What do you use intelligibility measures for? Select all that apply.

- ▢ Treatment planning
- ▢ Intervention
- ▢ Baseline measurements
- ▢ Eligibility for services
- ▢ Other ____

End of Block: Rationale for Measuring Speech Intelligibility

Start of Block: Practices in the Assessment of Speech Intelligibility

Q32 What type of speech production do you elicit for intelligibility assessment? Select all that apply.

- ▢ Spontaneous speech
- ▢ Imitative speech
- ▢ Cued speech
- ▢ Other ____

Q33 Do you think we, as speech-language pathologists, have adequate tools to assess speech intelligibility?

- ◯ Yes
- ◯ No

Skip To: Q35 If Do you think we, as speech-language pathologists, have adequate tools to assess speech intelligib... = No

Skip To: Q34 If Do you think we, as speech-language pathologists, have adequate tools to assess speech intelligib... = Yes

---

Q34 What are the adequate tools that you think we have?

____

---

*

Q35 What tools do you believe are necessary for SLPs to successfully measure speech intelligibility?

____

____

____

____

____

End of Block: Practices in the Assessment of Speech Intelligibility

---

Start of Block: COVID Considerations

Q36 How has the COVID-19 pandemic and subsequent shift to telepractice affected how you measure speech intelligibility?

____

End of Block: COVID Considerations
