AJSLP

Research Article

Development and Validation of a Probe Word List to Assess Speech Motor Skills in Children

Aravind Kumar Namasivayam,[a,b] Anna Huynh,[a] Rohan Bali,[a] Francesca Granata,[a] Vina Law,[a] Darshani Rampersaud,[a] Jennifer Hard,[a] Roslyn Ward,[c,d] Rena Helms-Park,[e] Pascal van Lieshout,[a,b,f] and Deborah Hayden[g]

Purpose: The aim of the study was to develop and validate a probe word list and scoring system to assess speech motor skills in preschool and school-age children with motor speech disorders.
Method: This article describes the development of a probe word list and scoring system using a modified word complexity measure and principles based on the hierarchical development of speech motor control known as the Motor Speech Hierarchy (MSH). The probe word list development accounted for factors related to word (i.e., motoric) complexity, linguistic variables, and content familiarity. The probe word list and scoring system was administered to 48 preschool and school-age children with moderate-to-severe speech motor delay at clinical centers in Ontario, Canada, and then evaluated for reliability and validity.
Results: One-way analyses of variance revealed that the motor complexity of the probe words increased significantly for each MSH stage, while no significant differences in the linguistic complexity were found for neighborhood density, mean biphone frequency, or log word frequency. The probe word list and scoring system yielded high reliability on measures of internal consistency and intrarater reliability. Interrater reliability indicated moderate agreement across the MSH stages, with the exception of MSH Stage V, which yielded substantial agreement. The probe word list and scoring system demonstrated high content, construct (unidimensionality, convergent validity, and discriminant validity), and criterion-related (concurrent and predictive) validity.
Conclusions: The probe word list and scoring system described in the current study provide a standardized method that speech-language pathologists can use in the assessment of speech motor control. It can support clinicians in identifying speech motor difficulties in preschool and school-age children, set appropriate goals, and potentially measure changes in these goals across time and/or after intervention.

Measuring outcomes following treatment in speech-language pathology is essential to evidence-informed practice. Outcome measurement allows for the assessment of treatment efficacy, evaluation of treatment progress, and planning for future courses of action (American Speech-Language-Hearing Association [ASHA], 2007; McCauley & Strand, 2008). However, for children with severe speech sound disorders (SSD), especially those with neuromotor or developmental speech motor control issues, measuring outcomes is challenging due to the complexity of their clinical presentation (Kearney et al., 2015). These children fall into four subtypes of motor speech disorders (MSD), namely, childhood dysarthria, childhood apraxia of speech (CAS), speech motor delay (SMD), and concurrent childhood dysarthria and CAS (Shriberg,

[a]Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, Ontario, Canada
[b]Toronto Rehabilitation Institute, Ontario, Canada
[c]Institute for Health Research, The University of Notre Dame Australia, Fremantle, Western Australia
[d]School of Allied Health, Curtin University, Bentley, Western Australia, Australia
[e]Linguistics, Department of Language Studies, University of Toronto Scarborough, Ontario, Canada
[f]Rehabilitation Sciences Institute, University of Toronto, Ontario, Canada
[g]The PROMPT Institute, Santa Fe, NM

Correspondence to Aravind Kumar Namasivayam: a.namasivayam@utoronto.ca

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Lisa Goffman

Received May 19, 2020
Revision received August 24, 2020
Accepted December 1, 2020
https://doi.org/10.1044/2020_AJSLP-20-00139

Disclosure: *This study was funded by an international competitive Clinical Trials Research Grant (2013–2019; NIH Trial Registration NCT02105402) awarded to the first author by The PROMPT Institute, Santa Fe, NM, USA. The authors will not receive any financial gains or royalties from the sale of the probe word scoring system and manuals. All proceeds go directly to the nonprofit organization, PROMPT Institute.*

Kwiatkowski, & Mabie, 2019). The subtypes are characterized by a range of speech motor issues, such as mandibular sliding, difficulty adjusting mandibular height for different vowels, undifferentiated tongue gestures, limited coordination between speech subsystems (e.g., between phonation and articulation), limited vocabularies, and unintelligible speech (Bernthal et al., 2009; Gibbon, 1999; Namasivayam, Coleman, et al., 2020; Namasivayam et al., 2013; Terband et al., 2013).

Probe word (PW) list and scoring systems (SS) are commonly used to measure treatment progress and generalization in this population. A PW list is composed of a customized set of words (i.e., a word list) and a scoring method that permits the measurement of intervention-related behavioral change (e.g., speech approximations) toward specific therapy targets. The PWs are customized carefully while being mindful of underlying constructs (what is being measured), task difficulty, information-processing load, and the client's needs and capabilities (Kearney et al., 2015; e.g., Strand et al., 2006).

PW List and SS: A Brief Overview

Different PW and SS have been extensively used in both single-subject (Maas et al., 2012; Square et al., 2014; Strand et al., 2006) and group designs (Murray, McCabe, & Ballard, 2015; Namasivayam et al., 2013). Earlier SS, such as the one presented by Hall et al. (1998), used a point deduction system. In this system, adult productions of items were given a score of 0, and discrepancies between the child's productions and those of adults were scored negatively. For instance, for every mismatched distinctive feature (i.e., voice, place, and manner), a point was deducted with distortions scored as a half point. The final score was then calculated based on the sum of mismatches from the adult form. Recent SS (e.g., see Maas et al., 2012; Maas & Farinella, 2012; Strand et al., 2006) use an auditory–perceptual 3-point scaling procedure that is based on scoring whole-word accuracy rather than individual sound productions (2 = *correct production*, 1 = *close approximation*, 0 = *incorrect production*). The scores are converted to a percentage based on the total possible points for a given set of words. This version involves the scoring of not only segmental-level information (place, voice, manner, and distortion errors) but also suprasegmental aspects of speech production (e.g., prosodic or stress errors), indices of speech timing (e.g., durational errors such as excessive vowel lengthening), and articulatory effort (e.g., excessive plosive release). Including both segmental and suprasegmental aspects into scoring is assumed to increase sensitivity to speech performance changes with time or intervention (Maas et al., 2012; Maas & Farinella, 2012; Strand et al., 2006).

For children with MSD, a combination of visual assessment of the accuracy of speech movements with linguistic transcription–based procedures is preferred (Square et al., 2014; Strand et al., 2006). One example is a 3-point scoring procedure by Strand et al. (2006), where 2 = *accurate movement gestures for correct production*, 1 = *intelligible production with minor errors* (mild vowel distortion, one distinctive feature off for consonant production, or close approximation of movement gesture), and 0 = *inaccurate production*. This allows for the scoring of both segmental (via auditory–perceptual linguistic transcription) and underlying speech motor control issues (via visual examination of speech movements). Such auditory–visual scoring procedures have been used successfully to study changes in speech performance following intervention in children with SSD and speech motor control issues (Dale & Hayden, 2013; Namasivayam et al., 2013; Square et al., 2014).

The PW list discussed in the current study is based on developmental speech motor research and a framework referred to as the Motor Speech Hierarchy (MSH; Hayden et al., 2010; Hayden & Square, 1994; see Figure 1). In the next few sections, we will briefly discuss the MSH and developmental evidence in support of this framework.

MSH

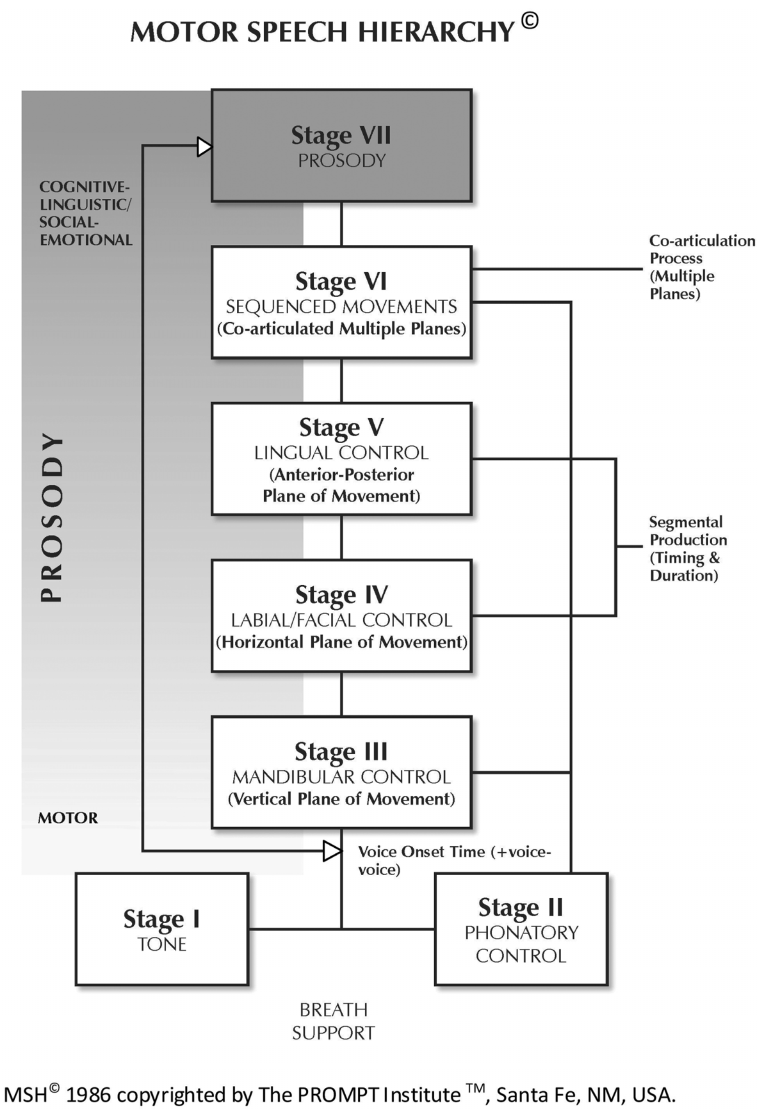
The PW list discussed in the current study is based on well-documented developmental speech motor research (Cheng et al., 2007; Green et al., 2000, 2002; Green & Nip, 2010; Iuzzini-Seigel et al., 2015; Namasivayam, Coleman, et al., 2020; Nip et al., 2009, 2011; Smith, 1992; Smith & Zelaznik, 2004) and the clinical and neurodevelopmental framework called the MSH (Hayden et al., 2010; Hayden & Square, 1994; see Figure 1). Henceforth, the PW list based on the MSH will be referred to as "MSH-PW." The MSH was originally conceptualized by Deborah Hayden in 1986 (Hayden & Square, 1994) to assist speech-language pathologists (SLPs) in understanding the processes involved in typical speech production and to serve as a guide for the assessment and intervention within the Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT; Hayden et al., 2010) approach.

The MSH reflects a hierarchical, nonlinear, and interactive development of speech motor control. The MSH posits seven key speech motor control stages: Stage I: tone, Stage II: phonatory control, Stage III: mandibular control, Stage IV: labial–facial control, Stage V: lingual control, Stage VI: sequenced movements, and Stage VII: prosody.

PROMPT intervention is based on the hierarchical establishment of movement parameters/control within each stage, including the refinement and integration of normalized movements from the preceding stage. The PROMPT approach assumes that, for children who are developing speech, intervention typically progresses in a systematic bottom-up fashion. That is, adequate physiological support for speech in the form of trunk, respiratory, and phonatory control (Stages I and II) must be present in order to organize and facilitate the supralaryngeal articulatory systems. Tone is defined as "resistance to passive stretch while a patient is attempting to maintain a relaxed state of muscle activity" (Goo et al., 2018, p. 661).

Within the MSH, speech production complexity is viewed in terms of the control of articulatory trajectories within and across planes of movement (i.e., vertical, horizontal, and anterior–posterior [transverse]). For example, Stage III: mandibular movements are said to occur in a vertical/midsagittal plane (Smith, 1992), Stage IV: speech-related labial–facial movements (e.g., lip rounding/spread)

Figure 1. Motor Speech Hierarchy (MSH). MSH reflects the hierarchical, nonlinear, and interactive development of speech motor control. The figure was used with permission from The PROMPT Institute, Santa Fe, NM. Copyright © 1986 The PROMPT Institute. This figure does not fall under the Creative Commons Open Access licensure.



are predominantly on the horizontal/coronal plane (with the exception of lip protrusion for a small sample of phonemes such as /u/ and /w/; Cosi & Caldognetto, 1996; Iuzzini-Seigel et al., 2015), Stage V: lingual movements are complex in nature and may occur in multiple planes (vertical, horizontal, and anterior–posterior; Hiiemae & Palmer, 2003; Sanders & Mu, 2013; Smith, 1992), and Stage VI: sequenced movements represent movements sequenced in time and co-articulated in multiple planes. Within the MSH framework, /ba/ can be considered a less complex production than /bu/, because /ba/ requires the mandible to move in an inferior direction (i.e., translate and rotate along the midsagittal plane) with minimal tongue movement. However, production of /bu/ involves not only the mandible moving in an inferior direction but also lip movement (rounding/protrusion) on two planes (Bunton, 2008; Forrest, 2002). A polysyllabic word such as "watermelon" or "rhinoceros" is relatively more complex, since it involves the interaction between multiple articulators and across multiple planes of movement carefully coordinated and sequenced through time. For a child to be accurately sequencing speech movements in complex contexts (Stage VI), it is a prerequisite that they have established adequate control of physiological support for speech (Stages I and II) and earlier stages of oral articulatory control (Stages III, IV, and V).

Stage VII relates to prosodic control. While prosodic control begins to develop during prespeech vocal play and is coordinated with mandibular movements in infancy

(Hübscher & Prieto, 2019; Prieto & Esteve-Gibert, 2018), it is on the top of the MSH as it represents the final stage of speech refinement within the PROMPT intervention. This is a paradoxical relationship in that prosody is not only fundamental but also the most complex system in communication, as it conveys more than words (i.e., the communicative intent of word use; Hayden & Square, 1994). Prosody not only provides the beginning or background melody for speech for infants to imitate and respond to but also requires the integration of all the earlier speech motor stages (to adequately control stress, intonation, pausing, and speech rate) to carry out complex communications (which encompass cognitive–linguistic and social–emotional context; Hayden & Square, 1994; Hübscher & Prieto, 2019; Prieto & Esteve-Gibert, 2018). This complex relationship is highlighted within the MSH framework in three ways: (a) a bidirectional connection (solid line) from the top of the MSH (Stage VII: prosody) to the bottom of the MSH (Stages I and II represent physiological support for speech), (b) a shaded box that encompasses all earlier stages, and (c) a connection of speech motor control (bottom of the shaded box) to social–linguistic and social–emotional context to express communicative intent (top of the shaded box; Hayden & Square, 1994).

Clinically, the MSH has been used in the assessment and treatment of speech motor disorders within the PROMPT approach. Each stage of the MSH is interactive with those above and below it. Each stage is considered as an interactive component within the context of the entire system; however, the importance of acquiring volitional control at each stage is emphasized in PROMPT treatment. A PROMPT clinician is trained to observe the child's level of control at each stage to determine the focus of intervention. The goal of intervention is to integrate speech motor actions learned at one stage, with previously learned action routines from other stages (for more details regarding the MSH, see Hayden et al., 2010).

Developmental Evidence for the MSH

Although the MSH was proposed in the mid-1980s solely based on clinical observation of speech movements (in typically developing children or children with neuromotor speech disorders), it aligns well with specific models of speech production related to articulatory phonology and task dynamics (Browman & Goldstein, 1992; Goldstein & Fowler, 2003; Saltzman & Kelso, 1987) and empirical developmental data (e.g., Green & Nip, 2010). A detailed timeline view of speech motor development in children based on a synthesis of research and observational data was published recently (please see Figure 1 in Namasivayam, Coleman, et al., 2020). Below, we provide an overview of this process based on data primarily from studies conducted on English-speaking children. We refer the reader to the article by Namasivayam, Coleman, et al. (2020) for a broader perspective on the relationship between specific theories of speech motor control, speech motor development, and SSD.

There are perceptual, acoustic, and kinematic data to suggest that different articulatory components (i.e., mandible, lips, tongue) of speech have distinct developmental schedules and the later developing labiofacial and lingual movements integrate into relatively stable and well-established mandibular movement patterns (e.g., Davis & MacNeilage, 1995; Green et al., 2000, 2002; Namasivayam, Coleman, et al., 2020; Nip et al., 2009; but see Diepstra et al., 2017; Giulivi et al., 2011). In infants less than 1 year old, fine force control required for the control of mandibular height is limited, and thus, mandibular movements are gross and limited to simple opening/closing movements (Green et al., 2000; Kent, 1992; Locke, 1983). Furthermore, vowels in the first year are characterized as low, nonfront, and nonrounded vowels, suggesting limited facial muscle (lip) interaction with the mandible and limited tongue elevation from the mandible (Buhr, 1980; Kent, 1992; Otomo & Stoel-Gammon, 1992). The coordination between the laryngeal and oral articulatory structures underlying voicing contrasts is acquired close to 2 years of age and follows the maturation and/or stabilization of mandibular movements (Green et al., 2002; Grigos et al., 2005; Yu et al., 2014).

By around 2 years of age, strong interlip spatial and temporal coupling is present in children (Green et al., 2000, 2002; Green & Nip, 2010; Nip et al., 2009) and undergoes a process of differentiation to gain independent control of the functionally linked upper and lower lips. This process has been observed between the ages of 2 and 3 years and supports the emergence of labiodental fricatives (/f/ and /v/; Green et al., 2000; Stoel-Gammon, 1985). Between the ages of 2 and 6 years, this process is further refined, and the upper and lower lip movements become adult-like with increasing contribution of the lower lip toward bilabial closure (Green et al., 2000, 2002). By around the age of 3 years, control of mandibular movements improves (emergence of /ɛ/ and /ɔ/ vowels), and tongue movements become relatively independent of the mandible resulting in reliable anterior–posterior lingual movements (emergence of diphthongs [e.g., /aʊ/, /ɔɪ/, /aɪ/] and coronal consonants [e.g., /t/ and /d/]; Donegan, 2013; Goldman & Fristoe, 2000; Kent, 1992; Otomo & Stoel-Gammon, 1992; Smit et al., 1990; Wellman et al., 1931).

By 4–5 years of age, a greater degree of control over mandibular height and improved tongue–mandible coordination is present and is characterized by the presence of all vowels and most consonants (Kent, 1992; McLeod & Crowe, 2018). Since the tongue is a hydrostatic organ with distinct functional segments, maturation and ambient language experience is required for the development of finer control and coordination of the tongue with neighboring articulators (Green & Wang, 2003; Kent, 1992; Nittrouer, 1993; Noiray et al., 2013). It has been noted that the coordination of the tongue's subcomponents follows different maturation schedules as well. Coordinated movements that use the back of the tongue to support the tongue tip during alveolar productions are adult-like by 4–5 years of age (Noiray et al., 2013), while those that relate to tongue tip release and tongue body backing mature later (Nittrouer, 1993). Thus, it is not surprising that children acquire rhotacized (retroflexed or bunched tongue) vowels (/ɝ/ and /ɚ/) much later and the production of complex fricatives last (McLeod & Crowe, 2018).

With regard to speech motor variability, studies have also demonstrated that earlier developing lip–mandible temporal coupling was similar to adults by 6 years of age (Green et al., 2000), whereas tongue tip to mandible temporal coupling is still more variable in 6- to 7-year-old children relative to adults (Cheng et al., 2007). Smith and Zelaznik (2004) have reported that variability of interlip coordination (and lower lip–mandible coordination) in 4- and 7-year-olds is greater than that in adults but decreases with age until it plateaus between 7 and 12 years of age. In children between 6 and 9 years of age, the extent and variability of lingual coarticulation is greater than in adults, supporting the notion that children are still fine-tuning speech motor control (Cheng et al., 2007; Nittrouer, 1993; Nittrouer et al., 2005, 1996; Zharkova et al., 2011). Data suggest that speech motor patterns become adultlike around 14 years but may take up to the age of 30 years to stabilize (Schötz et al., 2013; Smith & Zelaznik, 2004).

In summary, the above data suggest that there are varying schedules of development within the oral articulatory system. In general, these findings indicate that the lip–mandible coordination matures or develops earlier than tongue–mandible coordination or coordination within different subcomponents of the tongue (Cheng et al., 2007; Terband et al., 2009). Overall, these findings align with the concepts of MSH and the PROMPT intervention philosophy and imply that speech motor control development is hierarchical, sequential, nonuniform, interactive, and protracted (Smith & Zelaznik, 2004; Whiteside et al., 2003).

Rationale for the Current Study

PW lists are frequently reported in the literature for this population; however, they are often not standardized, and their content and administration procedures vary widely with very little information available on the validity (content, criterion, construct) and reliability (intra- and interrater) of the lists used (e.g., Square et al., 2014; Strand et al., 2006). Moreover, as these lists are individualized, the relative influence of cognitive, linguistic, and motor factors on the child's production is unknown (Green & Nip, 2010; Iuzzini-Seigel et al., 2015; Namasivayam, Coleman, et al., 2020). It has been suggested that, during speech development, speech motor skills act as a limitation with respect to the items that a child can produce (i.e., production constraint), while factors such as language and environment act as a drive (i.e., catalyst) that pushes a child to learn new speech motor control patterns, consequently facilitating the development of speech motor control (Green et al., 2000; Green & Nip, 2010; Iuzzini-Seigel et al., 2015). Given the well-known interactions between linguistic factors and speech motor control (e.g., Smith & Goffman, 2004; van Lieshout et al., 1995), it is critical to control for linguistic influences while investigating speech production in children with MSD (Namasivayam, Coleman, et al., 2020).

To address the limited availability of reliable and valid PW lists to assess intervention-related change for children with MSD (McCauley & Strand, 2008), we describe the development and validation of a PW list and SS for this population. The proposed PW list and SS is designed to be utilized as a criterion-based evaluative index to measure change in speech motor skill in children with MSD. The aims of the study are described in two phases:

- Phase I: We provide a description and analysis involved in the construction of the MSH-PW list based on both linguistic variables known to impact speech production (e.g., neighborhood density, mean biphone frequency, and log word frequency) and word (motoric) complexity factors (e.g., syllable structure, sound class, or complexity of underlying articulatory movements). We demonstrate that the MSH-PW list contains words with increasing word (i.e., motoric) complexity while minimizing the contribution of linguistic variables known to impact speech production.
- Phase II: We describe the validity and reliability of the PW list and SS procedure based on a sample of 45 preschool and school-age children with severe MSD.

Phase I: Development of MSH-PW List

Method

Word List

The PWs in this study were chosen to reflect the MSH stages, the speech motor development within each stage, and the content familiarity to children (Cheng et al., 2007; Green et al., 2000, 2002; Green & Nip, 2010; Hayden et al., 2010; Namasivayam, Coleman, et al., 2020; Smith & Zelaznik, 2004; Terband et al., 2011, 2013). In each stage, the main speech motor actions are embodied in the word forms. The MSH-PW list consists of 40 items (10 items per MSH stage) represented as pictures that are familiar to English-speaking preschool and school-age children. The items include the following categories: food (e.g., *cupcake, ham*), animals (e.g., *pup, owl*), body parts (e.g., *eye, feet*), objects (e.g., *ball, phone*), action verbs (e.g., *wash, show*), and common nouns and names (e.g., *papa, boy, Bob*). The following is a description and rationale for the selection of PWs for each MSH category (III to VI). Stages corresponding to tone (Stage I), phonatory control (Stage II), and prosody (Stage VII) are evaluated using the 40 items from the other stages.

Stage III: mandibular control. Words in this stage assess mandibular movements along the vertical plane and comprise primarily mid–low vowels (i.e., those requiring a moderate mandibular excursion range), bilabials (movements driven by the mandible as the main articulator at this stage), and an early-acquired diphthong. Syllable and word structures comprising consonants (C) and vowels (V) are VC, CV, CVC, or CVCV. This stage investigates the integration of mandibular opening/closing movements with the laryngeal (i.e., voicing) and velopharyngeal (i.e., nasalization) systems. Some examples of word forms chosen for probing at this stage are *pup*, *eye*, *map*, and *papa*.

Stage IV: labial–facial control. Words in this stage assess facial and specific lip movements (i.e., rounding and independent labial closure). At this stage, movement direction is predominantly on the horizontal/coronal plane.[1] Words are primarily composed of high–mid vowels (mainly rounded or retracted), diphthongs (e.g., /ɔɪ/), bilabials made with relatively independent action of the lips (i.e., lip rounding/retraction with high mandibular position as in *moon* and *peep*), labiodental fricatives made with relatively independent and individual lower lip control (e.g., *fish*), and other words with vowels requiring lip rounding and retraction (e.g., *bush*, *show*, *feet,* and *wash*). Syllable and word structures are CV or CVC. Some examples of word forms chosen for probing at this stage are *boy*, *feet*, *phone*, and *fish*.

Stage V: lingual control. Words in this stage assess lingual movements that are predominantly in the anterior–posterior and superior–inferior dimensions along the vertical/midsagittal plane. Words comprise vowels with differing heights (high–mid–low) paired with various combinations of lingual movements in the anterior–posterior and superior–inferior dimensions. These movement combinations are designed to examine the relative independence of tongue movements from mandible (e.g., *juice* where mandible is fixed at high position) and fine adjustments in tongue–mandibular coordination where lingual adjustments are required to adapt to varying mandibular angles/height (e.g., *dig* [high mandible position] vs. *log* [low mandible position]). At this stage, words also assess complex lingual fine motor control (e.g., rhotics/laterals) and lingual clusters (e.g., *clown, crib*, and *snake*). Syllable and word structures are VC, CVC, and CCVC. Word forms chosen for probing at this stage include *sun*, *ten*, *dig*, and *log*.

Stage VI: sequenced movement control. Words in this stage assess movements sequenced in time and are co-articulated in multiple planes (e.g., vertical, horizontal, anterior–posterior, and superior–inferior dimensions). Words vary between two and four syllables in length and are designed to examine the complex coordination between the mandibular, labial–facial, and lingual systems across multiple planes of movement. Examples of syllable and word structures include CVC-CVC, VC-CCVC, and CVC-Cr-Cr (where "r" indicates rhotacized vowel). Word forms chosen for probing at this stage consist of words such as *banana, umbrella, marshmallow*, and *rhinoceros*.

Linguistic complexity. There is a large volume of literature that has studied lexical and phonological factors affecting speech perception and production in children (e.g., Davis et al., 2018; Edwards et al., 2011; Kehoe et al., 2020; Munson et al., 2005). From these studies, three key variables have emerged, namely, word frequency, neighborhood density, and phonotactic probability. Variables such as word frequency and neighborhood density have been discussed from both a lexical and a phonological perspective, depending on the level of analysis, measurement methods, and context. The literature does not reveal a clear consensus on whether these variables should be categorized as lexical or phonological (Davis et al., 2018; Edwards et al., 2011; Kehoe et al., 2020; Munson et al., 2005). Thus, in the current study, we prefer to use the broader umbrella term "linguistic complexity" while referring to lexical and phonological complexity.

Word frequency refers to the number of times a specific word occurs in a spoken or written language corpus. General findings are that high-frequency words have facilitative effects on production, perception, and acquisition/learning in children and adults (e.g., Sosa, 2016). This effect is presumed to arise from the strengthening of neural access pathways and lexical representations as a result of repeated access in both production and perceptual domains (Bybee, 2003). Neighborhood density refers to the degree of phonological similarity between a specific word and other words. A word that differs from another word by substitution, addition, or deletion of a sound in any word position is considered a phonological neighbor (Luce & Pisoni, 1998). A word may contain many or only a few phonological neighbors; in the former case, the word is said to be in a dense neighborhood, while in the latter, the word belongs to a sparse neighborhood. Generally, children acquire words from dense neighborhoods earlier than sparse neighborhoods, possibly due to the decreased load on auditory–verbal short-term memory with words that share segments or phonological similarity (Kehoe et al., 2020; Stokes et al., 2012). However, high neighborhood density has been shown to inhibit word recognition in adults possibly due to lexical competition effects arising from phonological similarity (Luce & Pisoni, 1998).

Phonotactic probability is the relative frequency of occurrence of individual sounds or sound sequences in syllables and words in a given language (Jusczyk & Luce, 2002). Words with frequently occurring sound combinations are considered high-probability words, while words with less frequently occurring sound combinations are considered low-probability words. Phonotactic probability is positively correlated with neighborhood density since words with many neighbors generally contain sounds or sound sequences that occur frequently in each language (Sosa, 2016). Phonotactic probability is considered a phonological rather than lexical feature because it deals with characteristics of sounds or sound sequences and not whole words (Sosa, 2016). Studies have shown that children more accurately repeat and recall phonotactic high- rather than low-probability pseudowords (Storkel, 2001) and also learn novel words more easily when they contain frequent rather than infrequent sound sequences (Gonzalez-Gomez et al., 2013). From these studies, biphone frequency (i.e., the likelihood that two adjacent sounds co-occur in a given word position) has emerged as a key phonotactic probability measure that influences a child's speech production and learning (Storkel, 2001, 2004a, 2004b, 2004c).

Other factors such as word length have been discussed in the context of phonological or lexical complexity (e.g., Kehoe et al., 2020; Storkel, 2009). Several studies have found that word length affects word learning and that children generally learn short before long words (Maekawa & Storkel, 2006; Storkel, 2004a, 2004b, 2004c). Word length takes into account the number of phonemes in the target word, and longer words may also involve additional demands on speech

[1]For simplicity, lip protrusion is not differentiated from opening/spreading.

motor coordination. Thus, we and other researchers (e.g., Stoel-Gammon, 2010) have considered word length a measure of phonetic complexity (partially accounted for in the word complexity measure described by Stoel-Gammon, 2010; see Word Complexity section). Since the PWs were chosen to reflect the MSH stages, speech motor development, and content familiarity to children, the choice of words that could be included was limited in the current study. This prevented the investigation of the influence of other linguistic variables (e.g., grammatical class such as nouns vs. verbs) and their interactions with a child's age and vocabulary size. This should be considered as a study limitation and warrants further research. Additionally, discussions on the diverse theoretical perspectives that underlie mechanism(s) of interactions between phonological or lexical development and speech production is beyond the scope of this article (the reader is directed to excellent articles in this area by Dromey & Bates, 2005; Edwards et al., 2011; McAllister Byun & Tessier, 2016; Namasivayam, Coleman, et al., 2020; Smith, 2006).

In the current study, linguistic variables known to impact speech production such as neighborhood density, mean biphone frequency, and log word frequency (e.g., Bose et al., 2007; Edwards et al., 2011; Kehoe et al., 2020; Maekawa & Storkel, 2006; Sosa & Stoel-Gammon, 2012; Storkel, 2004a, 2004b, 2009) for the 10 PWs were calculated and averaged for each MSH stage (III, IV, V, and VI). The lexical characteristics of the PWs were determined from a child corpus of spoken American English (Storkel & Hoover, 2010). PWs that were not found in the child corpus were calculated from the English Lexicon Project (Balota et al., 2007). In addition, the number of phonemes, syllables, and morphemes in the MSH-PW list were noted. The overarching aim of Phase I is to demonstrate that linguistic (lexical and phonological) factors known to impact speech production showed minimal differences across the MSH stages. Therefore, linguistic complexity (i.e., neighborhood density, mean biphone frequency, and log word frequency) scores for items in each MSH stage should not be statistically different from one another.

Word (motoric) complexity. MSH-PW motoric complexity was determined using a modified version of the word complexity measure (mWCM) originally proposed by Stoel-Gammon (2010; see Table 1). For the mWCM, each PW within each MSH stage (except Stages I, II, and VII) was given points for number of syllables, the presence of certain word patterns, syllable structures, sound classes, or labial/lingual movements. The mWCM scores were then summed and averaged for each MSH stage from III to VI. We aimed to demonstrate that the MSH-PW list contains words with increasing word (i.e., motoric) complexity for more advanced stages. It is, therefore, expected that mWCM scores for each MSH stage will be statistically different from each other.

Content validity of MSH-PW list. The content validity of the PWs was determined by consulting a panel of 15 content experts. This panel consisted of clinicians, university professors, scientists, and postdoctoral students specializing in pediatric MSD. Panel members rated 13 underlying constructs (see Table 2) in the MSH-PW and SS for its relevance on a 4-point ordinal scale: 1 = *not relevant*, 2 = *somewhat relevant*, 3 = *quite relevant*, 4 = *highly relevant*. These content areas are based on empirical developmental information (e.g., Green & Nip 2010) and well-researched tasks in the literature (e.g., Shriberg et al., 2010). Lexical stress errors have been argued to reflect a child's speech motor control abilities (Kehoe et al., 1995; Shriberg et al., 2003; Strand et al., 2013). Therefore, lexical stress was included in this speech motor assessment to capture a child's difficulties to make subtle speech motor adjustments required to alter fundamental frequency, speech amplitude, and syllable durations (Kehoe et al., 1995).

Statistical Analysis

To demonstrate that the MSH-PW list contains words with increasing motoric word complexity (while minimizing the contribution of linguistic variables known to impact speech production), separate one-way analyses of variance (ANOVAs) were carried out with summed mWCM scores and linguistic complexity scores (neighborhood density, mean biphone frequency, and log word frequency) as the dependent variables and MSH stages as the group factor. As words in MSH Stage VI are multisyllabic (e.g., "marshmallow" and "rhinoceros") and arguably more linguistically complex (contains less frequently occurring words, sparse neighborhood density, etc.) than the other three MSH stages (III, IV, and V), we only provide descriptive statistics relating to linguistic complexity for this stage and did not include this in the ANOVA analysis. Tukey's honestly significant difference post hoc tests were performed, as necessary.

To assess content validity, we utilized the content validity index (CVI), which is the most widely utilized procedure of quantifying the content validity for a scale or instrument (Almanasreh, et al. 2019). We report both average-CVI (Ave-CVI) and by item-CVI scores (I-CVI) for relevance of items using standard formulas described in the literature (Almanasreh et al., 2019). Since chance agreement between the expert panelists is possible during the rating process, we adjusted each I-CVI for chance agreement using kappa scores (modified as κ^*; Polit et al., 2007). A scale or instrument can be considered as excellent in terms of content validity if it yields an I-CVI of $\geq .78$ and an Ave-CVI $\geq .90$. It should be noted that the probability of chance agreement decreases as the number of expert panelists increases. With 10 or more panelists, an I-CVI > .75 produces a $\kappa^* > .75$ (Almanasreh et al., 2019; Polit et al., 2007).

Results

For word (motoric) complexity, ANOVA results for mWCM scores (see Appendix A for raw scores) indicated a significant main effect, $F(3, 36) = 37.5$, $p < .001$, for MSH stages. Post hoc tests revealed that the mWCM scores increased significantly between each MSH stage (see Tables 3, 4, and 5). One-way ANOVAs for linguistic complexity indicated no significant main effects for neighborhood density,

Table 1. Parameters used for scoring Motor Speech Hierarchy probe word complexity (Stoel-Gammon, 2010).

Item description	Modified word complexity measure scoring
Word patterns	> 2 syllables = score 1
Word patterns	Stress on any syllable but the first = score 1
Syllable structure	Word-final consonant present = score 1
Syllable structure	Each consonant cluster present = score 1
Sound classes	Each velar consonant present = score 1
Sound classes	Each liquid, syllabic liquid, or rhotic vowel present = score 1
Sound classes	Each fricative or affricate consonant present = score 1
Sound classes	If voiced fricative or affricate consonant present = score 1
Movement trajectory	Labial rounding or retraction movements present = score 1
Movement trajectory	Each anterior to posterior (or vice versa) change (lingual) = score 1

$F(2, 27) = 2.64$, $p > .05$; mean biphone frequency, $F(2, 26) = 1.75$, $p > .05$; or log word frequency, $F(2, 27) = 3.22$, $p > .05$, indicating that words in MSH Stages III, IV, and V are not statistically different from each other with respect to selected linguistic factors. For content validity, Ave-CVI and I-CVI from a panel of pediatric motor speech experts are reported in Table 6 for relevance of items in the MSH-PW list. The I-CVI kappa scores ranged from .86 to 1, and the Ave-CVI was .95.

Phase II: Validity and Reliability of the MSH-PW List

Method

Participants

The current study included 48 children from a larger randomized controlled trial of speech motor intervention (Namasivayam, Huynh, et al., 2020), with a mean age of 48 months ($SD = 11$ months). Participant demographics are provided in Table 7. The children were recruited from community-based health care centers in Mississauga, Toronto, and Windsor, Ontario, Canada. We report baseline (i.e., preintervention) and postdelay/intervention data from these preschool/school-age children. The post delay/intervention data are specifically used to report predictive validity (see section on Criterion-Related Validity for details). Children were included in the study if they met the following criteria: (a) aged 3–10 years; (b) use of English as the primary language spoken at home; (c) diagnosed with moderate-to-severe SSD, an SMD subtype based on features reported in the precision stability index (Shriberg et al., 2010; Shriberg & Wren, 2019); (d) hearing and vision within normal limits; (e) Primary Test of Nonverbal Intelligence scores at or above the 25th percentile and standard score of ≥ 90 (within normal limits; Ehrler & McGhee, 2008); (f) receptive language skills at ≥ 78 standard score (i.e., age-appropriate or mildly delayed; no restrictions on expressive language scores) as assessed using Clinical Evaluation of Language Fundamentals assessment (preschool and school-age versions; Semel et al., 2003, 2004); (g) presented at least four out of nine indicators for motor speech involvement (e.g., lateral mandibular sliding, decreased lip rounding and retraction; as reported in Namasivayam, Huynh, et al., 2020; Namasivayam et al., 2013, 2019); and (h) had age-appropriate social and play skills. Due to slow participant recruitment, amendments to the inclusion criteria were made approximately 7.5 months following the start of the study/initial ethics approval. These amendments pertain to an increase in the age range from the original 3–6 to 3–10 years

Table 2. Content validity.

Item	Description
1	Assessment of mandibular movement range and stability (anterior and/or lateral mandibular slide)
2	Assessment of mandibular movement cycles: open-close (VC), close-open (CV), close-open-close (CVC), etc.
3	Assessment of lip symmetry
4	Assessment of lip rounding and lip retraction
5	Assessment of lip movements independent from mandible (e.g., lower lip movement [/f/, /v/])
6	Assessment of lingual movements (e.g., “dig” anterior to posterior; “ten” anterior to anterior)
7	Assessment of lingual-to-labial movement transitions and vice versa (e.g., “grape” lingual to labial)
8	Assessment of voicing transitions within and between syllables
9	Assessment of coordination and sequencing between multiple articulators
10	Assessment of syllable structure (e.g., CV, CV-CV, VC, CVC, CCVC, CVC-CVC)
11	Assessment of number of correct syllables in a word
12	Assessment of phonetic classes present (e.g., rhotics, fricatives, liquids/laterals)
13	Assessment of lexical stress

Note. Thirteen items relating to content domain of speech production and speech motor control (i.e., relevance) evaluated by a panel of 15 content experts. C = consonant; V = vowel.

Table 3. Descriptive statistics for the Motor Speech Hierarchy probe word (MSH-PW) list.

MSH stage	mWCM	Log word frequency	Mean biphone frequency	Neighborhood density	Phonemes	Syllables	Morphemes
III	0.8 (0.4)	2.5 (1.2)	0.004 (0.002)	15.9 (6.2)	2.6 (0.8)	1 (0)	1 (0)
IV	2.4 (0.9)	3.4 (0.6)	0.002 (0.001)	13.8 (7.9)	2.7 (0.4)	1 (0)	1 (0)
V	4.3 (1.3)	2.6 (0.4)	0.005 (0.003)	9.1 (5.9)	3.3 (0.6)	1 (0)	1 (0)
VI	6.3 (1.7)	2.3 (0.5)	0.002 (0.001)	0.5 (0.5)	6.8 (1.6)	2.7 (0.9)	1.5 (0.7)

Note. Means and standard deviations (in parentheses) for linguistic characteristics and modified word complexity measure (mWCM) for the MSH-PWs across each stage of the MSH.

old and the removal of restrictions in expressive language scores. Given the lack of restrictions on expressive language scores, there is a possibility that some of the children in the study may have developmental language disorder with co-occurring speech motor issues. In the current study, no attempt was made to rule out the co-occurrence of developmental language disorder. Children were excluded from the study if they presented with any of the following: (a) signs and symptoms suggesting global motor involvement (e.g., cerebral palsy), (b) more than seven out of 12 indicators on a CAS checklist (ASHA, 2007; cutoff used in study by Namasivayam, Pukonen, Goshulak, et al., 2015), (c) feeding/drooling or oral structural/resonance issues, and (d) autism spectrum disorder. All screening assessments related to inclusion and exclusion criteria were carried out by a licensed SLP. The study was approved by the research ethics board at the University of Toronto (Protocol 29142), and additional approvals were obtained from the three participating clinical sites.

Procedure

All children completed a test battery as part of the larger clinical trial (Namasivayam, Huynh, et al., 2020), which included speech assessments at body structures and functions level and activities-participation level as per the World Health Organization's International Classification of Functioning, Disability and Health for Children and Youth framework (Kearney et al., 2015; World Health Organization, 2007). The assessments were completed by two independent licensed SLPs who were blind to both group (whether children received intervention or not) and session (baseline or a 10-week follow-up) allocation as part of the larger clinical trial (Namasivayam, Huynh, et al., 2020). All assessments were audio- and video-recorded for the purposes of validity and reliability assessments. The assessment measures at the body structures and functions level included standardized testing of the following: speech motor control using the Verbal Motor Production Assessment for Children (VMPAC; Hayden & Square, 1999) and single word–level articulation and phonological performance using the Diagnostic Evaluation of Articulation and Phonology (DEAP) Test (Dodd et al., 2002) and Percentage of Consonants Correct (PCC; derived from the DEAP Test; Shriberg et al., 1997).

Table 4. One-way analysis of variance values for the Motor Speech Hierarchy probe words on the modified word complexity measure.

Variable	Sum of squares	*df*	M^2	*F*	Sig.
Between groups	169.7	3	56.5	37.5	.000
Within groups	54.2	36	1.5		
Total	223.9	39			

Listener ratings of speech intelligibility at word (Children's Speech Intelligibility Measure [CSIM]; Wilcox & Morris, 1999) and sentence (Beginner's Intelligibility Test [BIT]; Osberger et al., 1994) levels were obtained using naïve listeners (M_{age} = 22.61 years, SD = 3.94). The children's productions of speech intelligibility test items were audio-recorded using Zoom digital recorders (16 bits/sample at 44.1 kHz) and then edited to remove therapist instructions and extraneous noise and saved as .wav audio files.

Forty-eight participants were recruited in the larger randomized controlled trial (Namasivayam, Huynh, et al., 2020); however, complete data at both time points (baseline and 10-week follow-up) were only available from 45 participants for further analysis (i.e., data were lost to follow-up, not completing assessment following consent). There were four files per child (baseline and 10-week follow-up for CSIM and BIT) for a total of 180 audio files (45 children × 2 [CSIM and BIT] × 2 [baseline and 10-week follow-up]). The 180 audio files were divided into 45 playlists, comprising four randomly chosen files per playlist. Four files were selected to minimize excessive fatigue during testing and in response to participant's feedback recommending four to six audio files from previous studies (Namasivayam, Pukonen, Goshulak, et al., 2015; Namasivayam et al., 2019). In total, 135 naïve listeners (45 playlists × 3 listeners) participated in the speech intelligibility portion of the study. To avoid any bias, each playlist was played to a different group of three listeners. All listeners were blind to participant information, disorder classification, and session type. The playlists were created such that a group of listeners never heard the same child, word, or sentence list twice. All audio files were played in random order to three naïve listeners using a headphone amplifier (PreSonus HP60) and headphones (Sony MDR-XD10) at 70 dB SPL (e.g., Namasivayam, Pukonen, Goshulak, et al., 2015; Namasivayam et al., 2019).

PW Administration

In a quiet room, MSH-PWs were presented in a random order within each stage, and the order of stages

Table 5. Multiple comparisons of Motor Speech Hierarchy (MSH) probe word scores (from the modified word complexity measure) between MSH stages using the Tukey's honestly significant difference test.

MSH stage (I)	MSH stage (J)	Mean difference (I – J)	*SE*	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Stage III	Stage IV	−1.6*	.54	.030	−3.0	−0.1
	Stage V	−3.5*	.54	.000	−4.9	−2.0
	Stage VI	−5.5*	.54	.000	−6.9	−4.0
Stage IV	Stage III	1.6*	.54	.030	0.1	3.0
	Stage V	−1.9*	.54	.007	−3.3	−0.4
	Stage VI	−3.9*	.54	.000	−5.3	−2.4
Stage V	Stage III	3.5*	.54	.000	2.0	4.9
	Stage IV	1.9*	.54	.007	0.4	3.3
	Stage VI	−2.0*	.54	.004	−3.4	−0.5
Stage VI	Stage III	5.5*	.54	.000	4.0	6.9
	Stage IV	3.9*	.54	.000	2.4	5.3
	Stage V	2.0*	.54	.004	0.5	3.4

*The mean difference is significant at the .05 level.

was random both within and across participants. Direct imitation was used to elicit the items. For example, the SLP showed the picture of the word and said, "This is 'boy' … Say 'boy.'" The expected response from the child was "boy." The child was allowed one to two attempts to produce the word. No cues on the production of the word or feedback on the accuracy of the responses were given during MSH-PW administration (i.e., no knowledge of their performance or results). The interstimulus interval was about 3 s between each word presented. All MSH-PWs were audio- and video-recorded for off-line analysis. The examiner's scored sections were based on predetermined scoring rules (see section on MSH-PW Scoring).

MSH-PW Scoring

The MSH-PW scoring in the current study follows a modified version of the general auditory–visual scoring procedure described in the introduction, where both segmental and underlying speech motor control issues are scored via auditory–perceptual linguistic transcription and visual examination of speech movements (Dale & Hayden, 2013; Square et al., 2014; Strand et al., 2006). The scoring is weighted differently across each of the MSH stages and MSH-PWs. It is based on the presence or absence of certain features, such as appropriate mandibular movement range, correct syllable structure, and appropriate voicing transitions. The presence of each correct feature or appropriate movement parameter gets a score of 1, and if the feature or

Table 6. Content validity ratings from a panel of 15 content experts on 13 items (see Table 2) using a 4-point rating relevance scale.

Item	Panel of content experts (*N* = 15)															No. of agreements (*A*)	I-CVI (κ*)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	15	1.00
2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	15	1.00
3	4	4	2	4	3	4	4	3	4	4	4	4	2	4	3	13	.87
4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	3	15	1.00
5	2	4	4	4	3	4	4	4	4	4	4	4	4	4	3	14	.93
6	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	15	1.00
7	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	15	1.00
8	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	14	.93
9	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	15	1.00
10	4	4	4	4	4	4	2	4	4	4	4	4	2	4	4	13	.87
11	4	3	4	4	4	4	2	3	4	4	4	4	4	3	3	14	.93
12	4	3	4	2	4	4	4	4	4	4	4	4	4	1	3	13	.87
13	4	4	4	4	4	4	3	4	4	4	4	4	4	4	3	15	1.00
Average	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	.95

Note. Content validity index of an item (I-CVI) is the number of content experts rating the item either 3 or 4 divided by the total number of experts. Average-CVI (Ave-CVI) is the content validity index of the entire instrument calculated as the sum of the I-CVIs (I-CVI1 + I-CVI2 + … + I-CVI*n*) divided by the total number of items (Almanasreh et al., 2019). κ* is adjusted I-CVI for chance agreement (Polit et al., 2007). κ* is calculated as follows: κ* = (I-CVI − Pc) / (1 − Pc), where Pc is the probability of change agreement for each item. Pc = [(*N*!/*A*! (*N* − *A*)!]*.5*N*, where *N* = number of experts and *A* = number of panelists who agree that item is relevant.

Table 7. Participant demographics.

Variable	Mean (*SD*) or count (%)
Participants	*N* = 48
Age in months, *M* (*SD*)	48 (11)
Age range in months	36–93
Gender	Female = 19, Male = 29
Primary language (English) spoken at home	48 (100%)
Hearing and vision (within normal limits)	48 (100%)
History of speech and language intervention	32 (66%)
Primary Test of Nonverbal Intelligence[a]	
Nonverbal Index	105 (19)
Percentage of Consonants Correct[b]	42 (17)
Clinical Evaluation of Language Fundamentals[c]	
Receptive Language Index	96 (15)
Expressive Language Index	76 (15)
Mean number of indicators present:	
Motor speech involvement (max score = 9)[d]	8 (1)
Childhood apraxia of speech (max score = 12)[e]	4 (1)

Note. *SD* = standard deviation; % = proportion of children.

[a]Standard scores of the Primary Test of Nonverbal Intelligence (Ehrler & McGhee, 2008). [b]The Percentage of Consonants Correct (Shriberg et al., 1997) extracted from the Diagnostic Evaluation of Articulation and Phonology Test (Dodd et al., 2002). [c]Standard scores of Clinical Evaluation of Language Fundamentals (CELF-4, Semel et al., 2003; CELF Preschool-2, Semel et al., 2004). [d]Motor speech involvement checklist (Namasivayam et al., 2013). [e]Childhood apraxia of speech checklist (Namasivayam, Pukonen, Goshulak, et al., 2015).

movement is absent, it gets a score of 0 (see Table 8 for scoring examples). While all MSH-PWs in Stage III (mandibular) are scored out of five features, the scores for other stages are variable depending on the number of features (complexity) per word. For example, the five features assessed in Stage III consist of the following: appropriate mandibular range, appropriate mandibular stability and midline control, appropriate open/close (e.g., "ham") or close/open phase (e.g., "ba"), appropriate voicing transitions, and correct syllable structure. The MSH-PWs in Stage IV (labial–facial) are scored based on presence or absence of up to nine features, whereas words in Stage V (lingual) and Stage VI (sequenced) contain a maximum of seven and 14 features, respectively. For example, in Table 8, the left column demonstrates the scoring process for the word "map" (from Stage III: mandibular control) and the right column for the word "marshmallow" (from Stage VI: sequenced movements). While the word "map" has fewer content domains and thus a lower overall score (total of 5),

Table 8. Example of Motor Speech Hierarchy probe word (MSH-PW) scoring for MSH Stages III and IV.

MSH-PW scoring procedure	
Stage III: mandibular control	**Stage VI: sequenced movement control**
Example target word: *Map*	Example target word: *Marshmallow*
☒ Appropriate mandibular range	☒ Appropriate mandibular range
☒ Appropriate mandibular control (no anterior or lateral sliding)	☒ Appropriate mandibular control (no anterior or lateral sliding)
☒ Appropriate close/open/close phasing	☒ Appropriate lip symmetry
☒ Appropriate voicing transitions	☒ Appropriate independent bilabial control
☒ Correct syllable structure (CVC)	☒ Appropriate lip rounding
	☒ Appropriate labial–lingual transitions within and across syllables
	☒ Appropriate voicing transitions
	☒ Number of correct syllables (score on 3)
	☒ Appropriate lexical stress
	Phonetic classes present:
	☒ Rhotic
	☒ Fricative
	☒ Liquid/lateral
Total = 5	Total = 14

Note. Scoring involves the assessment of speech motor patterns, complex phonetic classes present, syllable-level features, and correct assignment of lexical stress. See text for more details. CVC = consonant–vowel–consonant.

the longer and more complex word "marshmallow" has not only more content domains but also a higher number of syllables and phonetic classes (e.g., rhotic, fricative, liquid/lateral) to be accounted for in scoring (total of 14).

Data Preparation and Management

All standardized speech motor control (VMPAC) and articulation and phonology assessments (DEAP) were scored on site using the hard copy of test forms and simultaneously audio- and video-recorded. These audio and video recordings were then reanalyzed by a second licensed SLP at the university laboratory for reliability and validity testing. Reliability and validity testing of the MSH-PWs was carried out on the recorded samples. To minimize potential risks to client confidentiality, all data were coded with a unique identifier, which included speaker number, session identification code, and date of recording. All data reported in the study are available in Appendix B.

Reliability Assessments

We analyzed three types of reliability for the MSH-PW list: intrarater reliability, interrater reliability, and internal consistency. Both intra- and interrater reliability were calculated on 20% of all data (18 randomly selected data files from 45 children × 2 sessions [baseline and 10-week follow-up]). For intrarater reliability, an SLP (Rater 1) reanalyzed the same child's recorded data after a time window of 2–3 weeks. This provided a measure of stability of the test scores across time. For interrater reliability, a second SLP rater (Rater 2) independently scored all of Rater 1's MSH-PW video samples. Both raters were licensed SLPs with over 20 years of specialization in pediatric MSD. The proportion of agreement (beyond that of expected chance) between the first and the second scoring for Rater 1 or between Raters 1 and 2 was calculated using Cohen's kappa on MedCalc statistical software (Version 19.2.1; MedCalc Software Ltd, 2020). Cohen's kappa is deemed appropriate to calculate intra- and interrater reliability for binary or categorical scores (Gianinazzi et al., 2015; Landis & Koch, 1977; Sim & Wright, 2005). Cohen's kappa benchmarks for categorical data proposed by Landis and Koch (1977) were used in the current study, where Cohen's kappa coefficient values ≥ .80, between .61 and .80, between .41 and .61, and < .41 represent excellent agreement, substantial agreement, moderate agreement, and fair to poor agreement, respectively (Gianinazzi et al., 2015).

Internal consistency reliability, also referred to as content sampling error, was calculated using Cronbach's coefficient alpha (Cronbach, 1951). Cronbach's coefficient alpha measures the interrelatedness of a sample of test items (i.e., the extent to which responses to items correlate with each other). When test items correlate to one another (i.e., demonstrate high shared covariances), the value of Cronbach's alpha is increased and approaches 1, whereas if items on a test or scale are completely independent from each other, then coefficient alpha approaches 0 (Tavakol & Dennick, 2011). As correlations between the items increase, it may suggest a greater degree of homogeneity among test items. When different test items measure different qualities, content, and constructs, the items will be unrelated to each other and the amount of error due to content sampling will be greater (i.e., Cronbach's alpha values are smaller; Tavakol & Dennick, 2011). However, high Cronbach's alpha values do not always imply homogeneity of the data set (Cortina, 1993; Vaske et al., 2017). High alpha values may arise from multidimensional data sets where separate clusters of items may intercorrelate highly (Vaske et al., 2017). Homogeneity denotes unidimensionality in a sample of test items and is recommended that it is established independently of internal consistency, for example, by using factor analysis (see Construct Validity section; Vaske et al., 2017).

The internal consistency levels of unacceptable, poor, questionable, acceptable, good, and excellent correspond to Cronbach's alpha coefficient values of < .5, .5–.6, .6–.7, .7–.8, .8–.9, and > .9, respectively (Tavakol & Dennick, 2011). All reliability scores (intrarater, interrater, and internal consistency) were assessed independently for each of the following stages of the MSH: Stage III: mandibular control, Stage IV: labial–facial control, Stage V: lingual control, and Stage VI: sequenced movements.

Validity Assessments

For Phase II of this study, we analyzed and report construct and criterion-related validity.

Construct validity. In this study, construct validity was measured in four ways: unidimensionality (demonstration of unidimensionality of the underlying data set), convergent validity (significant association of MSH-PW scores to similar tests assessing the same constructs), discriminant validity (nonsignificant association of MSH-PW scores to tests focusing on different constructs), and speech motor performance data from a sample of children with speech motor issues (SMD population; by analyzing the mean percent PW scores across each of the MSH stages).

Unidimensionality. Unidimensionality measurements on the MSH-PWs were carried out using factor analysis with varimax rotation across all items. Data sets that generate high loadings for only one factor indicate unidimensionality or the presence of a single construct or latent variable. To assess the suitability of data for factor analysis, a Kaiser–Meyer–Olkin measure (Kaiser, 1974) of sampling adequacy and Bartlett's test of sphericity (Bartlett, 1954) to assess multivariate normality were carried out. Kaiser–Meyer–Olkin scores > 0.8 and a significant ($p < .05$) Bartlett's test of sphericity implies that the sampling is adequate and meets assumption of multivariate normality (Hadi et al., 2016; Pallant, 2013).

Convergent validity. To assess convergent validity, Pearson correlation r was calculated between the MSH-PW scores and the speech motor control assessment scores from VMPAC (Hayden & Square, 1999). For this study, only the focal oromotor control (FOC) and sequencing (SEQ) subsections of VMPAC were used. VMPAC-FOC evaluates volitional oromotor control for mandibular, labial–facial, and lingual movements in both speech and nonspeech contexts. VMPAC-SEQ subsection evaluates a participant's ability to produce nonspeech and speech movements in the correct sequential order. For this study, only speech-related items from FOC and SEQ subsections were used in the estimation of convergent validity. Since the MSH-PW list only contains single words at mono-, bi-, and multisyllabic levels, other subsections of the VMPAC that pertain to global motor control (i.e., neurophysiological support for speech; e.g., head and neck control, postural control) and sentence-level connected speech and voice characteristics were deemed not relevant. High positive Pearson's r values between MSH-PW scores and VMPAC assessment would indicate that they are measuring similar constructs. In the current study, values of < .20 are considered very weak, .20–.39 indicate weak, .40–.59 indicate moderate, .60–.79 indicate strong, and ≥ .80 indicate very strong correlations (Evans, 1996).

Discriminant validity. Discriminant validity would be demonstrated if Pearson's r correlation coefficients between MSH-PW scores and a test focusing on a different construct yielded a nonsignificant association. We chose to test discriminant validity of MSH-PWs by estimating Pearson's r correlation coefficients between MSH-PW scores and single-word phonological error pattern scores from the standardized DEAP Test (Dodd et al., 2002). According to the DEAP Test manual, phonological error patterns are considered to reflect deficits in the child's knowledge of the linguistic–phonological system (Dodd et al., 2002), and thus, these scores should be distinct from those of MSH-PWs that assess speech motor performance in the current study.

Speech motor performance of children with SMD. During test construction and evaluation of psychometric properties of a standardized test, it is standard practice to establish construct validity via performance assessments of special populations (e.g., Hayden & Square, 1999; Kaufman, 1995; Strand et al., 2013). Children in the current study were classified as SMD subtype based on features reported in the precision stability index (Shriberg et al., 2010; Shriberg & Wren, 2019). The phenotype for SMD is consistent with a delay in the development of speech-related neuromotor precision and stability. That is, these children present with lower performance scores on measures of speech precision and stability (i.e., in the lower tail of speech motor skill development; Shriberg, Campbell, et al., 2019). Support for construct validity of the MSH-PWs can be tested with the following prediction: For children with a delay in speech motor skill development (i.e., SMD population), we predict that they would demonstrate a greater difficulty (lower mean percent MSH-PW scores) in controlling speech movements that are acquired and refined later such as lingual control and sequenced and coordinated movements across multiple planes (i.e., higher on the MSH Stages V and VI) relative to those that develop fairly early such as mandible and lip control (i.e., MSH Stages III and IV; Green & Nip, 2010; Namasivayam, Coleman, et al., 2020). To this end, we carried out a one-way ANOVA with percent MSH-PW scores as the dependent variable and MSH stages as the group factor. Tukey's honestly significant difference post hoc tests were performed, as necessary.

Criterion-related validity. We used two approaches to measure criterion-related validity (viz., concurrent validity and predictive validity). Criterion validity describes how effective the MSH-PW assessment is in estimating the child's performance on a standardized outcome measure such as speech intelligibility or speech severity. Concurrent and predictive validity differs depending on when the MSH-PW assessment and the speech intelligibility assessments are carried out (Miller & Lovler, 2018). If the MSH-PW scores and the criterion scores (e.g., speech intelligibility or speech severity scores) are assessed simultaneously, it refers to concurrent validity, as in the case when both assessments are compared during baseline (preintervention) testing. If MSH-PW scores at Time 1 (preintervention) are used to predict speech intelligibility or severity scores at Time 2 (post delay/intervention), this refers to predictive validity. Both concurrent and predictive validity are expressed statistically as Pearson's r correlation coefficients, and simple linear regression models were calculated between the MSH-PW scores and measures of word-level speech intelligibility (CSIM; Wilcox & Morris, 1999), sentence-level speech intelligibility (BIT; Osberger et al., 1994), and speech severity. Severity of SSD in children is commonly assessed using the PCC (Shriberg et al., 1997), which we extracted from the DEAP Test (Dodd et al., 2002). Previous studies have demonstrated a positive association between speech motor control and speech intelligibility (Namasivayam et al., 2013; Rong & Green, 2019; Sapir et al., 2010; Weismer, 2008; Weismer et al., 2012); thus, it is a reasonable expectation that we should observe large positive correlations between MSH-PW scores, speech intelligibility, and speech severity for both types of criterion-related validity.

Results

All data reported in the study are available in Appendix B.

Reliability Assessments

Results for the intrarater, interrater, and internal consistency reliability analyses are reported in Table 9. Intrarater reliability Cohen's kappa coefficient values between the first and the second scoring for Rater 1 were ≥ .80 across all stages of the MSH. For interrater reliability, Cohen's kappa coefficients were between .48 and .63 across the MSH stages. The Cronbach's alpha scores for internal consistency across different MSH stages ranged between .75 and .83.

Validity Assessments

Construct validity. In this study, construct validity was measured in three ways: unidimensionality (demonstration of unidimensionality of the underlying data set), convergent validity (significant association of MSH-PW scores

Table 9. Intrarater, interrater, and internal consistency reliability scores for the Motor Speech Hierarchy probe word (MSH-PW) list.

	Stages in MSH-PW list				
Reliability assessments	**Stage III: mandibular**	**Stage IV: labial–facial**	**Stage V: lingual**	**Stage VI: sequenced**	**Average**
Cohen's kappa coefficient					
Intrarater	.81	.81	.87	.84	.83
Interrater	.52	.57	.63	.48	.55
Cronbach's alpha coefficient					
Internal consistency	.75	.79	.78	.83	.78

to similar tests assessing the same constructs), and discriminant validity (nonsignificant association of MSH-PW scores to tests focusing on different constructs). *Unidimensionality* measurements on the MSH-PWs were carried out using a factor analysis across items. The data met the assumptions for factor analysis (see Table 10), and the results indicated that only one factor had an eigenvalue above 1 and accounted for approximately 82% of variance in data. *Convergent validity* was demonstrated by a strong and significant positive Pearson correlation between the MSH-PW scores and the VMPAC speech motor assessment, based on its speech items only (FOC + SEQ subsections: $r = .61$, $p < .01$). *Discriminant validity* was established from a low Pearson's r correlation ($r = .02$, $p > .05$) between MSH-PW scores and standard scores of phonological error patterns from the DEAP phonological assessment (Dodd et al., 2002).

ANOVA results for construct validity testing using a special group sample (SMD population) indicated a significant main effect, $F(3, 188) = 29.3$, $p < .001$, for MSH stages. Post hoc tests revealed that the percent MSH-PW scores only decreased significantly between earlier- (MSH Stages III and IV) and later-acquired MSH stages (Stages V and VI). Mean MSH-PW scores did not significantly differ within earlier MSH stages (III and IV) or within later-acquired MSH stages (V and VI). See Tables 11 and 12.

Figure 2 represents mean MSH-PW scores in percentage values across MSH stages based on a sample of children with SSD (SMD subtype). Figure 3 demonstrates individual differences in MSH-PW scoring profiles across five randomly selected participants across various clinics. At both individual (see Figure 3) and group (see Figure 2) levels, there is a clear pattern of decreasing speech motor performance (lower mean MSH-PW scores) moving from earlier- to later-acquired MSH stages.

Criterion-related validity. Criterion-related validity was demonstrated from positive and high Pearson's r correlation coefficients between the MSH-PW scores and measures of word-level speech intelligibility (CSIM scores), sentence-level speech intelligibility (BIT scores), and speech severity (PCC score; see Table 13). Pretreatment MSH-PW scores significantly predicted pre- and posttreatment speech intelligibility (CSIM and BIT) and speech severity (PCC). Pretreatment MSH-PW scores also explained a significant proportion of variance (42%–57%) in the speech intelligibility and speech severity scores (see Table 13).

Discussion

The proposed MSH-PW and SS is a criterion-based evaluative index designed to measure change in speech motor skills in children with SSD. The objective of Phase I was to provide a description of the construction and analysis of the MSH-PW's linguistic and word (motoric) complexity factors, and in Phase II, we provide evidence for the

Table 10. Factor analysis for Motor Speech Hierarchy (MSH) probe word items.

Factor analysis			
Tests			
Bartlett's test of sphericity	$\chi^2 = 195.17$; $p < .001$		
Kaiser–Meyer–Olkin	0.844		
Factor loadings with varimax rotation			
MSH stage	**Factor 1**		**Factor 1**
Stage III: mandibular	−0.912	Sum square loadings	3.316
Stage IV: labial–facial	−0.945	Proportional variance	0.829
Stage V: lingual control	−0.904	Cumulative variance	0.829
Stage VI: sequenced	−0.878		
Eigenvalues for 4 factors: 3.48532679, 0.21449434, 0.19457067, 0.1056082			

Table 11. Descriptive statistics and one-way analysis of variance (ANOVA) values for the mean Motor Speech Hierarchy (MSH) probe word scores from 48 children with speech motor delay across each stage of the MSH.

Descriptive statistics for MSH stages					
Variable	**Stage III**	**Stage IV**	**Stage V**	**Stage VI**	
M	71.8	68.0	43.3	40.2	
SD	21.6	22.2	17.7	21.7	
One-way ANOVA					
Variable	**Sum of squares**	***df***	***M*²**	***F***	**Sig.**
Between groups	38610	3	12870	29.3	.000
Within groups	82422.7	188	438.4	—	—
Total	121032.8	191	—	—	—

reliability and validity of the MSH-PW list for assessing speech motor skills in children with severe MSD.

Phase I

One-way ANOVAs revealed that the motor complexity of the MSH-PWs increased significantly for each MSH stage, and no significant differences in the linguistic complexity were found for neighborhood density, mean biphone frequency, or log word frequency. The MSH-PW list allows us to capture subtle changes in speech motor skills without undue influence of linguistic factors known to impact speech production.

Furthermore, content validity was established in Phase I using a systematic CVI process based on 15 expert raters (i.e., target respondents of the instrument). The experts rated how representative the items (e.g., mandibular stability, articulatory trajectories, and lexical stress) are of the content domain of speech production and speech motor control (i.e., relevance). Excellent content validity was established (I-CVI kappa scores ranged from .86 to 1.00; Ave-CVI of .95), where the expert raters agreed that the items in the MSH-PW list captured important physiological and behavioral indices of speech motor skills relevant to pediatric MSD.

Phase II

We provided evidence for the reliability and validity for the MSH-PW list for assessing speech motor skills in children with severe MSD.

Reliability

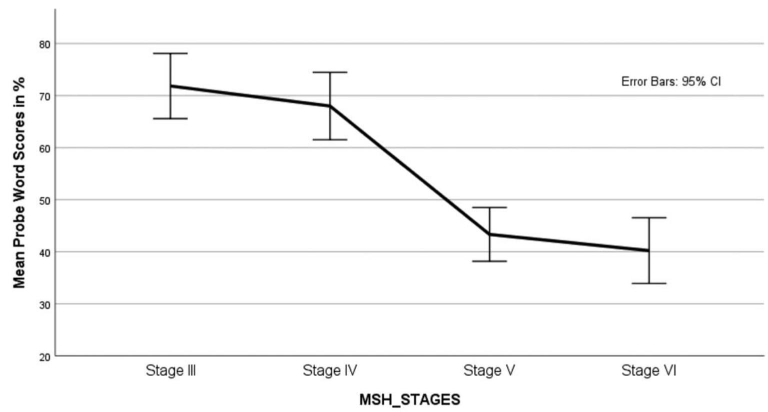
In the current study, intra- and interrater and internal consistency reliability were assessed. Intrarater reliability across all stages of the MSH indicated excellent agreement of scores between the first and the second scoring for Rater 1 (Cohen's kappa coefficient values ≥ .80). Intrarater reliability scores suggest a relatively low level of error. When interrater reliability was assessed between two independent raters, it indicated moderate agreement across the MSH stages (Cohen's kappa coefficients between .48 and .57), with the exception

Table 12. Multiple comparisons of mean Motor Speech Hierarchy (MSH) probe word scores (in percentage) between MSH stages using the Tukey's honestly significant difference test.

MSH stages (I)	MSH stages (J)	Mean difference (I – J)	*SE*	Sig.	95% Confidence interval: Lower bound	95% Confidence interval: Upper bound
Stage III	Stage IV	3.8	4.2	.805	−7.2	14.9
	Stage V	28.5*	4.2	.000	17.4	39.5
	Stage VI	31.6*	4.2	.000	20.5	42.7
Stage IV	Stage III	−3.8	4.2	.805	−14.9	7.2
	Stage V	24.6*	4.2	.000	13.5	35.7
	Stage VI	27.7*	4.2	.000	16.7	38.8
Stage V	Stage III	−28.5*	4.2	.000	−39.5	−17.4
	Stage IV	−24.6*	4.2	.000	−35.7	−13.5
	Stage VI	3.1	4.2	.885	−7.9	14.2
Stage VI	Stage III	−31.6*	4.2	.000	−42.7	−20.5
	Stage IV	−27.7*	4.2	.000	−38.8	−16.7
	Stage V	−3.1	4.2	.885	−14.2	7.9

*The mean difference is significant at the .001 level.

Figure 2. Mean Motor Speech Hierarchy (MSH) probe word scores in percentage across MSH stages based on a sample of children with speech sound disorders (a speech motor delay subtype). Error bars: 95% confidence intervals.



of MSH Stage V, which yielded a substantial agreement (Cohen's kappa coefficient value of .63). The moderate interrater reliability is not surprising since we used a multidimensional scoring approach in the study. In the current study, each MSH-PW was assessed on many parameters ranging from mandibular movement range and stability to accurate lexical stress and syllable structure (13 content dimensions listed in Table 2). Although scoring more dimensions is valuable in some aspects such as helping clinicians develop clinical observation skills and identifying critical problem areas for treatment planning and possibly tracking progress, it does have drawbacks (McNeil & Prescott, 1978; Odekar & Hallowell, 2005a, 2005b; Porch, 2007). It is well known that interrater reliability is higher when simple correct versus incorrect scoring is used (such as in standard articulation testing using perceptual transcription) and much lower as the number of scoring dimensions increase (e.g., see Odekar & Hallowell, 2005a, 2005b; Porch, 2007; Strand et al., 2013). With more dimensions added, there is a chance of confusion and varied interpretations between raters resulting in lower reliability scores (Odekar & Hallowell, 2005a, 2005b). To mitigate this, multidimensional scoring approaches require clinicians to spend additional time in training (e.g., learning scoring from recorded samples where reliable scores have been obtained from experienced raters) to develop observational skills and improve reliability. These may not be

Figure 3. Mean percent Motor Speech Hierarchy (MSH) probe word scores across MSH stages for five randomly selected participants to demonstrate individual differences in scoring profiles. S2, S4, E1, J2, and J26 are individual study participants from different clinics.

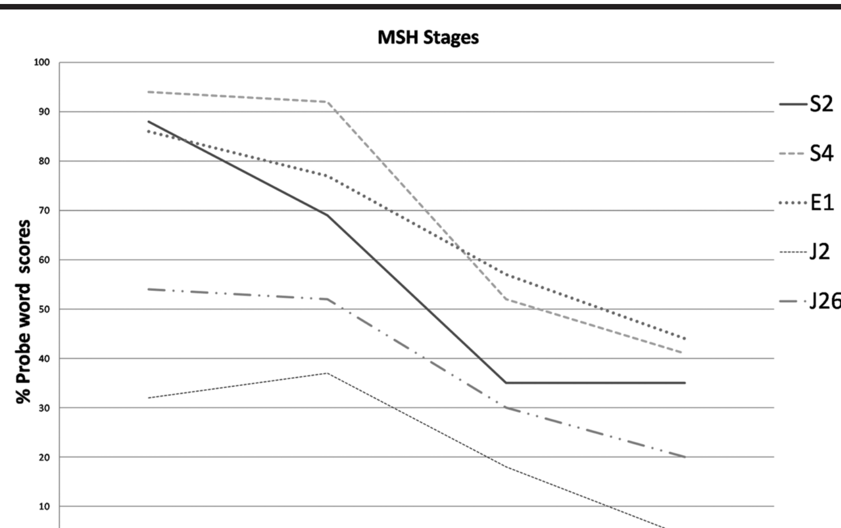


Table 13. Criterion-related validity.

Variable	α	β	*SE*	95% CI	Pearson's *r*	R^2	Adjusted R^2	Model	*t*	*p*
Concurrent validity										
Word-level speech intelligibility (CSIM)	–3.22	.18	0.02	[0.13, 0.23]	.76	.57	.57	$F(1, 43) = 59.24, p < .001$	7.69	< .001
Sentence-level speech intelligibility (BIT)	–23.49	.18	0.03	[0.12, 0.24]	.68	.47	.46	$F(1, 43) = 38.66, p < .001$	6.21	< .001
Speech severity (PCC)	–3.56	.19	0.02	[0.14, 0.25]	.74	.56	.55	$F(1, 44) = 56.11, p < .001$	7.49	< .001
Predictive validity										
Word-level speech intelligibility (CSIM)	10.57	.15	0.02	[0.09, 0.21]	.65	.42	.41	$F(1, 38) = 28.30, p < .001$	5.32	< .001
Sentence-level speech intelligibility (BIT)	–22.92	.23	0.03	[0.16, 0.30]	.74	.55	.54	$F(1, 39) = 48.55, p < .001$	6.96	< .001
Speech severity (PCC)	11.94	.17	0.02	[0.12, 0.22]	.72	.52	.51	$F(1, 42) = 46.87, p < .001$	6.84	< .001

Note. Concurrent and predictive validity expressed as correlation and linear regression coefficients between Motor Speech Hierarchy probe word scores and measures of word-level speech intelligibility (CSIM), sentence-level speech intelligibility (BIT), and speech severity (PCC). Predictive validity is determined with the pretreatment measures, and concurrent validity is determined with posttreatment measures. CI = confidence interval; CSIM = Children's Speech Intelligibility Measure; BIT = Beginner's Intelligibility Test; PCC = Percentage of Consonants Correct.

time and cost efficient in high-pressure service delivery contexts (Odekar & Hallowell, 2005a, 2005b).

Finally, the internal consistency of MSH-PW items demonstrated acceptable to good reliability (Cronbach's alpha ranged between .75 and .83). Generally, Cronbach's alpha scores in the range between .65 and .80 are considered "adequate" or "acceptable" for tests and scales in human social science and behavioral research (Vaske et al., 2017). We did not delete any of the items to increase the value of Cronbach's alpha scores for two reasons: It already has high reliability (Shelby, 2011), and because Cronbach's alpha represents the interrelatedness of a sample of test items, very high reliability scores (≥ .95) may suggest redundancy between items (Streiner, 2003; Tavakol & Dennick, 2011; Vaske et al., 2017).

Validity

The assessment of construct validity entailed measures of unidimensionality, convergent validity, and discriminant validity. The unidimensional structure of the MSH-PWs was assessed via factor analysis to evaluate the proportion of variance explained by different components of the data. The factor analysis yielded only one significant factor (e.g., only one factor with an eigenvalue > 1) that accounted for approximately 82% of variance in data. Such high loadings for a single factor when taken together with high internal consistency scores (average Cronbach's alpha of .78 across MSH levels) indicate unidimensionality or the presence of a single construct or latent variable (i.e., homogeneity in the data set; Cortina, 1993; Hadi et al., 2016; Pallant, 2013; Vaske et al., 2017). High construct validity was further demonstrated by appropriate correlations to other standardized test measures. A significant positive strong correlation was found when MSH-PW scores were compared to measures evaluating similar constructs as in the VMPAC speech motor assessment (i.e., convergent validity; also see the Limitations section for a discussion on the strength of association), and a low nonsignificant correlation was found between measures that do not assess speech motor control such as the phonological error patterns from the DEAP phonological assessment (Dodd et al., 2002), illustrating discriminant validity.

Construct validity was also supported by differences in mean MSH-PW across MSH stages in children with SMD. The SMD subtype is characterized by a delay in the development of speech-related neuromotor precision and stability (Shriberg, Campbell, et al., 2019, Shriberg et al., 2010; Shriberg & Wren, 2019). We predicted that children with SMD would demonstrate greater difficulty (lower mean percent MSH-PW scores) in controlling speech movements that are acquired and refined later (e.g., lingual control and sequenced movements across multiple planes; i.e., MSH Stages V and VI) relative to those that develop fairly early on (e.g., mandible and labial facial control; i.e., MSH Stages III and IV). Our data supported this construct validity prediction (see Figures 2 and 3). At an individual level, although there is variability (see Figure 3), the general pattern of increasing speech motor difficulty (lower MSH-PW scores) with increasing MSH stages is evident. At a group level, there is a clear and statistically significant distinction in speech motor performance between earlier- and later-acquired MSH stages (see Figure 2).

Finally, high criterion-related validity was demonstrated in the current study via significant and positive high correlation and regression coefficients between the MSH-PW scores and listener ratings of speech intelligibility (at both word and sentence levels) and a measure of speech severity (PCC scores). These findings are not surprising as previous studies have shown positive associations between speech motor skills, speech production, and speech intelligibility (Namasivayam et al., 2013; Namasivayam, Pukonen, Hard, et al., 2015; Rong & Green, 2019; Sapir et al., 2010; Weismer, 2008; Weismer et al., 2012). In the current study, MSH-PW scores accounted for approximately 42%–57% of the variance in speech intelligibility and speech severity when measured simultaneously (concurrent validity) or after a period (predictive validity). These findings are in line with those reported for dimensions of articulatory control (i.e., labial–facial control, tongue height, and front–back position of tongue) accounting for up to 50% of variance in speech intelligibility (Sapir et al., 2010; Weismer et al., 2012). The high concurrent and predictive validity of the MSH-PWs can give clinicians confidence that changes revealed in the MSH-PW scores will likely result in changes in speech severity and speech intelligibility at the single-word and sentence levels.

Clinical Significance

Importantly, the MSH-PWs may assist the SLPs in the formulation of the strengths and areas of focus (e.g., therapy goals) for intervention, as well as aid clinicians in describing the clinical presentation of severe SSD with motor origins. For example, if we consider a performance or motor skill criterion score of > 80% (based on age-matched normative data from the VMPAC manual; Hayden & Square, 1999) in Figure 3, it is clear that only participants S4, S2, and E1 have adequate mandibular control (high scores in Stage III) relative to participants J2 and J26. For labiofacial control (Stage IV), only participants S4 and E1 meet the required skill levels. For other MSH stages, no participant met the cutoff scores for adequate speech motor skills. A necessary next step for a clinician would be to examine the individual word scores within each MSH stage on the scoring form to see if a pattern emerges (see Table 8 for a scoring example). For example, appropriate mandibular range and stability, appropriate lip rounding, or tongue tip elevation may be absent across items. When taken together, the mean MSH-PW scores across each stage of the MSH and a careful analysis of error patterns within each MSH stage can indicate the level of breakdown in the speech motor system and the degree of differentiation for speech articulators (i.e., begun to develop independence with finely graded control; Namasivayam, Coleman, et al., 2020). Once these have been determined, developmentally appropriate speech motor patterns can be identified, and priorities within the speech motor system can be defined for intervention (see Hayden et al., 2010, for a detailed description).

Limitations

There are several limitations in the current study:

1. We were limited to finding words within each MSH stage that (a) contained certain speech motor parameters (e.g., movement trajectory directions), (b) followed certain speech motor developmental patterns, (c) met imageability requirements (for color picture stimuli cards), and (d) were familiar to children. This restricted our choice of words, and we were unable to control some aspects of the word list such as syllable-foot structure in multisyllabic words (some words contained weak initial [or unfooted] syllables and others did not), grammatical class words (both nouns and verbs were included, which may have been processed and produced differently for each child), and word familiarity (some words may have been more familiar [e.g., "ball" and "pup"] to children than others [e.g., "owl" and "map"]). These are potential confounds in the study and warrant further research.

2. We found a statistically significant ($p < .01$) and strong positive correlation supporting convergent validity between the MSH-PW scores and the VMPAC speech motor assessment ($r = .61$), based on speech items only for the latter. Given that there are no gold-standard speech motor skills assessments and VMPAC was the only standardized assessment with published norms (McCauley & Strand, 2008; Snyder, 2005), we had chosen to use this measure to assess convergent validity of the MSH-PWs. Very strong correlation scores ($r > .80$) were not obtained possibly due to the differences in the items and scoring procedures between the VMPAC and the MSH-PWs. For example, the VMPAC item scores are weighted based on the sensory cues (i.e., auditory/visual versus tactile) provided to elicit the responses (Hayden & Square, 1999). In contrast, no cues on the production of the word or feedback on the accuracy of the responses are provided during MSH-PW administration.

3. Another major limitation in the current study relates to the lack of fine-grained analysis of variables and items. In the current article, we only provide a broad view of development, validity, and reliability of the MSH-PW list and scoring process. We did not perform a fine-grained analysis on item reliability (e.g., how reliable is visual–perceptual scoring of mandibular movement range or lateral mandibular slide) or test construct validity in a control group of typically developing children. These follow-up studies are currently underway.

4. Test–retest data for multiple administrations of the MSH-PW within short periods were limited, as data were gathered from a larger clinical trial study (Namasivayam, Huynh, et al., 2020). Future studies may consider evaluating the test–retest reliability of this measure within a short time (e.g., retested within 7–10 days). It is anticipated that the results would remain consistent (i.e., no change should occur if no intervention were provided), and it would be minimally influenced by practice effects (as items are administered randomly).

5. Therapy for MSD in children is a specialized area within speech-language pathology, and clinicians may take many months to years to become skilled in observation, assessment, and intervention with this population (e.g., see Namasivayam, Pukonen, Goshulak, et al., 2015, for a clinician training protocol in pediatric MSD). In this context, it should be of interest to explore what effect differing levels of experience of SLPs would have on the administration and scoring of the MSH-PWs. In particular, given that the data demonstrated mostly moderate interrater reliability, it would be prudent to train SLPs to a validation criterion (e.g., using standardized materials and video samples). Further research with newer technologies such as markerless facial motion capture may one day replace such moderately reliable SLP scoring with more objective measures (e.g., see Bandini et al., 2017).

6. We were only able to provide two examples of the scoring process for the MSH-PWs (see Table 8). The full 20-page scoring manual and stimuli cards can be purchased for a fee directly from The PROMPT Institute.

7. Lastly, the findings reported in Phase II are based on speech from a very specific group of children with SMD; additional testing and validation with other populations are warranted (e.g., childhood dysarthria and CAS).

Conclusions

Through a systematic process of item development (Phase 1) and testing on a clinical population (Phase II), we have gathered evidence to support the preliminary use of the MSH-PW list in evaluating speech motor skills of children with severe MSD. The MSH-PW list described can assist clinicians by providing a standardized method to identify speech motor difficulties in a child, set appropriate goals, and potentially measure changes in these goals across time. Currently, large-scale normative data on typically developing children using the MSH-PWs are being collected to understand the process of speech motor development in greater detail.

Author Contributions

Aravind Kumar Namasivayam: Conceptualization (Lead), Formal Analysis (Lead), Methodology (Lead), Project administration (Lead), Writing – original draft (Lead), Writing – review & editing (Supporting), Funding acquisition (Equal), Resources (Lead), Supervision (Equal), Data curation (Lead). **Anna Huynh:** Data curation (Supporting), Formal analysis (Supporting), Methodology (Supporting), Project administration (Supporting), Validation (Supporting), Visualization (Supporting), Writing – original draft (Supporting), Writing – review & editing (Supporting). **Rohan Bali:** Formal analysis (Supporting), Methodology (Supporting), Project administration (Supporting), Resources (Lead), Software (Lead), Validation (Supporting), Visualization (Supporting). **Francesca Granata:** Data curation (Supporting), Formal analysis (Supporting), Resources (Supporting), Validation (Supporting), Writing – review & editing (Supporting). **Vina Law:** Data curation (Supporting), Formal analysis (Supporting), Methodology (Supporting), Writing – review

& editing (Supporting). **Darshani Rampersaud:** Data curation (Supporting), Formal analysis (Supporting), Project administration (Supporting), Resources (Supporting), Writing – original draft (Supporting), Writing – review & editing (Supporting). **Jennifer Hard:** Data curation (Lead), Formal analysis (Supporting), Methodology (Supporting), Resources (Supporting), Supervision (Supporting), Writing – original draft (Supporting). **Roslyn Ward:** Conceptualization (Supporting), Resources (Supporting), Software (Supporting), Validation (Supporting), Writing – review & editing (Supporting). **Rena Helms-Park:** Formal analysis (Supporting), Methodology (Supporting), Resources (Equal), Supervision (Equal), Validation (Supporting), Writing – review & editing (Supporting). **Pascal van Lieshout:** Conceptualization (Supporting), Data curation (Supporting), Formal analysis (Supporting), Funding acquisition (Equal), Investigation (Supporting), Methodology (Supporting), Project administration (Equal), Resources (Lead), Software (Supporting), Supervision (Supporting), Validation (Supporting), Writing – review & editing (Supporting). **Deborah Hayden:** Conceptualization (Lead), Methodology (Supporting), Resources (Lead), Software (Supporting), Supervision (Equal), Validation (Supporting), Writing – original draft (Equal).

Acknowledgments

The authors would like to thank all of the families who participated in the study and all of the support staff, including more than 40 research assistants and 15 speech-language pathologists from the University of Toronto who assisted with the study. Special thanks to the following speech-language pathologists for their assistance with data collection and scoring procedures: Donna Gaines, Terry Hall, Elizabeth Kuegeler Wolters, Gayle Merrefield, Renee Charlifue-Smith, Debra Goshulak, Michele Weerts, Inga Manuel, and Faye Petrie. We gratefully acknowledge the support and collaboration of the following clinical centers: The Speech & Stuttering Institute (Toronto, Ontario, Canada), John McGivney Children's Centre of Essex County (Windsor, Ontario, Canada), and ErinoakKids Centre for Treatment and Development (Mississauga, Ontario, Canada).

References

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy, 15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. https://www.asha.org/policy

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. https://doi.org/10.3758/BF03193014

Bandini, A., Namasivayam, A. K., & Yunusova, Y. (2017). Video-based tracking of jaw movements during speech: Preliminary results and future directions. In *Proceedings of the INTERSPEECH 2017 Conference, ISCA Archive, Stockholm, Sweden* (pp. 689–693). https://doi.org/10.21437/Interspeech.2017-1371

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximation. *Journal of the Royal Statistical Society: Series B (Methodological), 16*(2), 296–298. https://doi.org/10.1111/j.2517-6161.1954.tb00174.x

Bernthal, J. E., Bankson, N. W., & Flipsen, P. (2009). *Articulation and phonological disorders: Speech sound disorders in children* (6th ed.). Pearson.

Bose, A., van Lieshout, P., & Square, P. A. (2007). Word frequency and bigram frequency effects on linguistic processing and speech motor performance in individuals with aphasia and normal speakers. *Journal of Neurolinguistics, 20*(1), 65–88. https://doi.org/10.1016/j.jneuroling.2006.05.001

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155–180. https://doi.org/10.1159/000261913

Buhr, R. D. (1980). The emergence of vowels in an infant. *Journal of Speech and Hearing Research, 23*(1), 73–94. https://doi.org/10.1044/jshr.2301.73

Bunton, K. (2008). Speech versus nonspeech: Different tasks, different neural organization. *Seminars in Speech and Language, 29*(4), 267–275. https://doi.org/10.1055/s-0028-1103390

Bybee, J. (2003). *Phonology and language use*. Cambridge University Press.

Cheng, H. Y., Murdoch, B. E., Goozée, J. V., & Scott, D. (2007). Physiologic development of tongue–jaw coordination from childhood to adulthood. *Journal of Speech, Language, and Hearing Research, 50*(2), 352–360. https://doi.org/10.1044/1092-4388(2007/025)

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Cosi, P., & Caldognetto, E. M. (1996). Lips and jaw movements for vowels and consonants: Spatio-temporal characteristics and bimodal recognition applications. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 291–313). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-13015-5_23

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dale, P. S., & Hayden, D. A. (2013). Treating speech subsystems in childhood apraxia of speech with tactual input: The PROMPT approach. *American Journal of Speech-Language Pathology, 22*(4), 644–661. https://doi.org/10.1044/1058-0360(2013/12-0055)

Davis, B., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*(6), 1199–1211. https://doi.org/10.1044/jshr.3806.1199

Davis, B., Van der Feest, S., & Hoyoung, Y. I. (2018). Speech sound characteristics of early words: Influence of phonological factors across vocabulary development. *Journal of Child Language, 45*(3), 673–702. https://doi.org/10.1017/S0305000917000484

Diepstra, H., Trehub, S. E., Eriks-Brophy, A., & van Lieshout, P. H. H. M. (2017). Imitation of non-speech oral gestures by 8-month-old infants. *Language and Speech, 60*(1), 154–166. https://doi.org/10.1177/0023830916647080

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. Psychology Corporation.

Donegan, P. (2013). Normal vowel development. In M. J. Ball & F. E. Gibbon (Eds.), *Vowels and vowel disorders: A handbook* (pp. 24–60). Psychology Press.

Dromey, C., & Bates, E. (2005). Speech interactions with linguistic, cognitive, and visuomotor tasks. *Journal of Speech, Language, and Hearing Research, 48*(2), 295–305. https://doi.org/10.1044/1092-4388(2005/020)

Edwards, J., Munson, B., & Beckman, M. E. (2011). Lexicon–phonology relationships and dynamics of early language development—A commentary on Stoel-Gammon's "Relationships between lexical and phonological development in young children". *Journal of Child Language, 38*(1), 35–40. https://doi.org/10.1017/S0305000910000450

Ehrler, D. J., & McGhee, R. L. (2008). *Primary Test of Nonverbal Intelligence*. Pro-Ed.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole.

Forrest, K. (2002). Are oral-motor exercises useful in the treatment of phonological/articulatory disorders? *Seminars in Speech and Language, 23*(1), 15–26. https://doi.org/10.1055/s-2002-23508

Gianinazzi, M. E., Rueegg, C. S., Zimmerman, K., Kuehni, C. E., & Michel, G. (2015). Intra-rater and inter-rater reliability of a medical record abstraction study on transition of care after childhood cancer. *PLOS ONE, 10*(5), Article e0124290. https://doi.org/10.1371/journal.pone.0124290

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*(2), 382–397. https://doi.org/10.1044/jslhr.4202.382

Giulivi, S., Whalen, D. H., Goldstein, L. M., Nam, H., & Levitt, A. G. (2011). An articulatory phonology account of preferred consonant–vowel combinations. *Language Learning and Development, 7*(3), 202–225. https://doi.org/10.1080/15475441.2011.564569

Goldman, R., & Fristoe, M. (2000). *GFTA-2: Goldman-Fristoe Test of Articulation–Second Edition*. AGS. https://doi.org/10.1037/t15098-000

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Mouton de Gruyter.

Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A "bat" is easier to learn than a "tab": Effects of relative phonotactic frequency on infant word learning. *PLOS ONE, 8*(3), Article e59601. https://doi.org/10.1371/journal.pone.0059601

Goo, M., Tucker, K., & Johnston, L. M. (2018). Muscle tone assessments for children aged 0 to 12 years: A systematic review. *Developmental Medicine & Child Neurology, 60*(7), 660–671. https://doi.org/10.1111/dmcn.13668

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43*(1), 239–255. https://doi.org/10.1044/jslhr.4301.239

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research, 45*(1), 66–79. https://doi.org/10.1044/1092-4388(2002/005)

Green, J. R., & Nip, I. S. (2010). Some organization principles in early speech development. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 171–188). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199235797.003.0010

Green, J. R., & Wang, Y. T. (2003). Tongue-surface movement patterns during speech and swallowing. *The Journal of the Acoustical Society of America, 113*(5), 2820–2833. https://doi.org/10.1121/1.1562646

Grigos, M. I., Saxman, J. H., & Gordon, A. M. (2005). Speech motor development during acquisition of the voicing contrast. *Journal of Speech, Language, and Hearing Research, 48*(4), 739–752. https://doi.org/10.1044/1092-4388(2005/051)

Hadi, N. U., Abdullah, N., & Sentosa, I. (2016). An easy approach to exploratory factor analysis: Marketing perspective. *Journal of Educational and Social Research, 6*(1), 215–223.

Hall, R., Adams, C., Hesketh, A., & Nightingale, K. (1998). The measurement of intervention effects in developmental phonological disorders. *International Journal of Language & Communication Disorders, 33*(1), 445–450. https://doi.org/10.3109/13682829809179466

Hayden, D., Eigen, J., Walker, A., & Olsen, L. (2010). PROMPT: A tactually grounded model. In L. Williams, S. McLeod, & R. McCauley (Eds.), *Interventions for speech sound disorders in children* (pp. 453–474). Brookes.

Hayden, D., & Square, P. (1994). Motor speech treatment hierarchy: A systems approach. *Clinics in Communication Disorders, 4*(3), 162–174.

Hayden, D., & Square, P. (1999). *VMPAC: Verbal Motor Production Assessment for Children*. The Psychological Corporation. https://doi.org/10.1037/t15167-000

Hiiemae, K. M., & Palmer, J. B. (2003). Tongue movements in feeding and speech. *Critical Reviews in Oral Biology & Medicine, 14*(6), 413–429. https://doi.org/10.1177/154411130301400604

Hübscher, I., & Prieto, P. (2019). Gestural and prosodic development act as sister systems and jointly pave the way for children's sociopragmatic development. *Frontiers in Psychology, 10,* 1259. https://doi.org/10.3389/fpsyg.2019.01259

Iuzzini-Seigel, J., Hogan, T. P., Rong, P., & Green, J. R. (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learning and Development, 3*(1), 53–68. https://doi.org/10.1123/jmld.2014-0054

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing, 23*(1), 2–40. https://doi.org/10.1097/00003446-200202000-00002

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika, 39,* 31–36. https://doi.org/10.1007/BF02291575

Kaufman, N. (1995). *Kaufman Speech Praxis Test for Children*. Wayne State University Press.

Kearney, E., Granata, F., Yunusova, Y., van Lieshout, P., Hayden, D., & Namasivayam, A. (2015). Outcome measures in developmental speech sound disorders with a motor basis. *Current Developmental Disorders Reports, 2*(3), 253–272. https://doi.org/10.1007/s40474-015-0058-2

Kehoe, M., Patrucco-Nanchen, T., Friend, M., & Zesiger, P. (2020). The relationship between lexical and phonological development in French-speaking children: A longitudinal study. *Journal of Speech, Language, and Hearing Research, 63*(6), 1807–1821. https://doi.org/10.1044/2020_JSLHR-19-00011

Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research, 38*(2), 338–350. https://doi.org/10.1044/jshr.3802.338

Kent, R. D. (1992). The biology of phonological development. *Phonological development: Models, research, implications, 65,* 90.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Locke, J. L. (1983). *Phonological acquisition and change*. Academic Press.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Maas, E., Butalla, C. E., & Farinella, K. A. (2012). Feedback frequency in treatment for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 21*(3), 239–257. https://doi.org/10.1044/1058-0360(2012/11-0119)

Maas, E., & Farinella, K. A. (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(2), 561–578. https://doi.org/10.1044/1092-4388(2011/11-0120)

Maekawa, J., & Storkel, H. (2006). Individual differences in the influence of phonological characteristics on expressive vocabulary development by young children. *Journal of Child Language, 33*(3), 439–459. https://doi.org/10.1017/S0305000906007458

McAllister Byun, T., & Tessier, A. M. (2016). Motor influences on grammar in an emergentist model of phonology. *Language & Linguistics Compass, 10*(9), 431–452. https://doi.org/10.1111/lnc3.12205

McCauley, R. J., & Strand, E. A. (2008). A review of standardized tests of nonverbal oral and speech motor performance in children. *American Journal of Speech-Language Pathology, 17*(1), 81–91. https://doi.org/10.1044/1058-0360(2008/007)

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100

McNeil, M. R., & Prescott, T. E. (1978). *Revised Token Test*. Pro-Ed.

MedCalc Software Ltd. (2020). *MedCalc statistical software (Version 19.2.1)*. https://www.medcalc.org

Miller, L. A., & Lovler, R. L. (2018). *Foundations of psychological testing: A practical approach*. Sage.

Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children. *Journal of Speech, Language, and Hearing Research, 48*(1), 108–124. https://doi.org/10.1044/1092-4388(2005/009)

Murray, E., McCabe, P., & Ballard, K. J. (2015). A randomized controlled trial for children with childhood apraxia of speech comparing rapid syllable transition treatment and the Nuffield Dyspraxia Programme–Third Edition. *Journal of Speech, Language, and Hearing Research, 58*(3), 669–686. https://doi.org/10.1044/2015_JSLHR-S-13-0179

Namasivayam, A. K., Coleman, D., O'Dwyer, A., & van Lieshout, P. (2020). Speech sound disorders in children: An articulatory phonology perspective. *Frontiers in Psychology, 10*, 2998. https://doi.org/10.3389/fpsyg.2019.02998

Namasivayam, A. K., Huynh, A., Granata, F., Law, V., & van Lieshout, P. (2020). PROMPT intervention for children with severe speech motor delay: A randomized control trial. *Pediatric Research,* 1–9. https://doi.org/10.1038/s41390-020-0924-4

Namasivayam, A. K., Pukonen, M., Goshulak, D., Hard, J., Rudzicz, F., Rietveld, T., Maassen, B., Kroll, R., & van Lieshout, P. (2015). Treatment intensity and childhood apraxia of speech. *International Journal of Language & Communication Disorders, 50*(4), 529–546. https://doi.org/10.1111/1460-6984.12154

Namasivayam, A. K., Pukonen, M., Goshulak, D., Granata, F., Le, D. J., Kroll, R., & van Lieshout, P. (2019). Investigating intervention dose frequency for children with speech sound disorders and motor speech involvement. *International Journal of Language & Communication Disorders, 54*(4), 673–686. https://doi.org/10.1111/1460-6984.12472

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

Namasivayam, A. K., Pukonen, M., Hard, J., Jahnke, R., Kearney, E., Kroll, R., & van Lieshout, P. (2015). Motor speech treatment protocol for developmental motor speech disorders. *Developmental Neurorehabilitation, 18*(5), 296–303. https://doi.org/10.3109/17518423.2013.832431

Nip, I. S. B., Green, J. R., & Marx, D. B. (2009). Early speech motor development: Cognitive and linguistic considerations. *Journal of Communication Disorders, 42*(4), 286–298. https://doi.org/10.1016/j.jcomdis.2009.03.008

Nip, I. S. B., Green, J. R., & Marx, D. B. (2011). The co-emergence of cognition, language, and speech motor control in early development: A longitudinal correlation study. *Journal of Communication Disorders, 44*(2), 149–160. https://doi.org/10.1016/j.jcomdis.2010.08.002

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research, 36*(5), 959–971. https://doi.org/10.1044/jshr.3605.959

Nittrouer, S., Estee, S., Lowenstein, J. H., & Smith, J. (2005). The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops. *The Journal of the Acoustical Society of America, 117*(1), 351–364. https://doi.org/10.1121/1.1828474

Nittrouer, S., Neely, S. T., & Studdert-Kennedy, M. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research, 39*(2), 379–389. https://doi.org/10.1044/jshr.3902.379

Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *The Journal of the Acoustical Society of America, 133*(1), 444–452. https://doi.org/10.1121/1.4763983

Odekar, A., & Hallowell, B. (2005a). Comparison of alternatives to multidimensional scoring in the assessment of language comprehension in aphasia. *American Journal of Speech-Language Pathology, 14*(4), 337–345. https://doi.org/10.1044/1058-0360(2005/032)

Odekar, A., & Hallowell, B. (2005b). Exploring inter-rater agreement in scoring of the Revised Token Test. *Journal of Medical Speech-Language Pathology, 14*(3), 123–131. http://aphasiology.pitt.edu/id/eprint/1578

Osberger, M. J., Robbins, A. M., Todd, S. L., & Riley, A. I. (1994). Speech intelligibility of children with cochlear implants. *The Volta Review, 96*(5), 169–180.

Otomo, K., & Stoel-Gammon, C. (1992). The acquisition of unrounded vowels in English. *Journal of Speech and Hearing Research, 35*(3), 604–616. https://doi.org/10.1044/jshr.3503.604

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Allen & Unwin. http://www.allenandunwin.com/spss

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467. https://doi.org/10.1002/nur.20199

Porch, B. E. (2007). Comments on "Comparison of alternatives to multidimensional scoring in the assessment of language comprehension in aphasia" by Odekar and Hallowell (2005). *American Journal of Speech-Language Pathology, 16*(1), 84–86. https://doi.org/10.1044/1058-0360(2007/011)

Prieto, P., & Esteve-Gibert, N. (2018). *The development of prosody in first language acquisition* (Vol. 23). John Benjamins. https://doi.org/10.1075/tilar.23

Rong, P., & Green, J. R. (2019). Predicting speech intelligibility based on spatial tongue–jaw coupling in persons with amyotrophic lateral sclerosis: The impact of tongue weakness and jaw adaptation. *Journal of Speech, Language, and Hearing Research, 62*(8S), 3085–3103. https://doi.org/10.1044/2018_JSLHR-S-CSMC7-18-0116

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. *Psychological Review, 94*(1), 84–106. https://doi.org/10.1037/0033-295X.94.1.84

Sanders, I., & Mu, L. (2013). A three-dimensional atlas of human tongue muscles. *Oral Biology, 296*(7), 1102–1114. https://doi.org/10.1002/ar.22711

Sapir, S., Ramig, L., Spielman, J., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research, 53*(1), 114–125. https://doi.org/10.1044/1092-4388(2009/08-0184)

Schötz, S., Frid, J., & Löfqvist, A. (2013). Development of speech motor control: Lip movement variability. *The Journal of the Acoustical Society of America, 133*(6), 4210–4217. https://doi.org/10.1121/1.4802649

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2)*. The Psychological Corporation.

Semel, E. M., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. The Psychological Corporation.

Shelby, L. B. (2011). Beyond Cronbach's alpha: Considering confirmatory factor analysis and segmentation. *Human Dimensions of Wildlife, 16*(2), 142–148. https://doi.org/10.1080/10871209.2011.537302

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708–722. https://doi.org/10.1044/jslhr.4004.708

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., & Nadler, C. J. (2003). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics & Phonetics, 17*(7), 549–574. https://doi.org/10.1080/0269920031000138123

Shriberg, L. D., Campbell, T. F., Mabie, H. L., & McGlothlin, J. H. (2019). Initial studies of the phenotype and persistence of speech motor delay (SMD). *Clinical Linguistics & Phonetics, 33*(8), 737–756. https://doi.org/10.1080/02699206.2019.1595733

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilkens, C. M., & Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795–824. https://doi.org/10.3109/02699206.2010.503006

Shriberg, L. D., Kwiatkowski, J., & Mabie, H. L. (2019). Estimates of the prevalence of motor speech disorders in children with idiopathic speech delay. *Clinical Linguistics & Phonetics, 33*(8), 679–706. https://doi.org/10.1080/02699206.2019.1595731

Shriberg, L. D., & Wren, Y. E. (2019). A frequent acoustic sign of speech motor delay (SMD). *Clinical Linguistics & Phonetics, 33*(8), 757–771. https://doi.org/10.1080/02699206.2019.1595734

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

Smit, A. B., Hand, L., Freilinger, J., Bernthal, J., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*(4), 779–798. https://doi.org/10.1044/jshd.5504.779

Smith, A. (1992). The control of orofacial movements in speech. *Critical Reviews in Oral Biology & Medicine, 3*(3), 233–267. https://doi.org/10.1177/10454411920030030401

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders, 39*(5), 331–349. https://doi.org/10.1016/j.jcomdis.2006.06.017

Smith, A., & Goffman, L. (2004). Interaction of motor and language factors in the development of speech production. In B. Maasen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 225–252). Oxford University Press.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology, 45*(1), 22–33. https://doi.org/10.1002/dev.20009

Snyder, K. (2005). In Review of verbal motor production assessment for children. In **R. A. Spies, & B. S. Plake (Eds.),** *The sixteenth mental measurements yearbook,* (pp. 1080–1082). Buros Institute of Mental Measurements. http://marketplace.unl.edu/buros/

Sosa, A. (2016). Lexical considerations in the treatment of speech sound disorders in children. *Perspectives of the ASHA Special Interest Groups, 1*(1), 57–65. https://doi.org/10.1044/persp1.SIG1.57

Sosa, A., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research, 55*(2), 596–608. https://doi.org/10.1044/1092-4388(2011/10-0113)

Square, P. A., Namasivayam, A. K., Bose, A., Goshulak, D., & Hayden, D. (2014). Multi-sensory treatment for children with developmental motor speech disorders. *International Journal of Language & Communication Disorders, 49*(5), 527–542. https://doi.org/10.1111/1460-6984.12083

Stoel-Gammon, C. (1985). Phonetic inventories, 15–24 months: A longitudinal study. *Journal of Speech and Hearing Research, 28*(4), 505–512. https://doi.org/10.1044/jshr.2804.505

Stoel-Gammon, C. (2010). The word complexity measure: Description and application to developmental phonology and disorders. *Clinical Linguistics & Phonetics, 24*(4–5), 271–282. https://doi.org/10.3109/02699200903581059

Stokes, S., Kern, S., & dos Santos, C. (2012). Extended statistical learning as an account for slow vocabulary growth. *Journal of Child Language, 39*(1), 105–129. https://doi.org/10.1017/S0305000911000031

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337. https://doi.org/10.1044/1092-4388(2001/103)

Storkel, H. L. (2004a). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics, 25*(2), 201–221. https://doi.org/10.1017/S0142716404001109

Storkel, H. L. (2004b). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research, 47*(5), 1194–1212. https://doi.org/10.1044/1092-4388(2004/088)

Storkel, H. L. (2004c). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research, 47*(6), 1454–1468. https://doi.org/10.1044/1092-4388(2004/108)

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language, 36*(2), 291–321. https://doi.org/10.1017/S030500090800891X

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods, 42*(2), 497–506. https://doi.org/10.3758/BRM.42.2.497

Strand, E. A., McCauley, R. J., Weigand, S. D., Stoeckel, R. E., & Baas, B. S. (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech, Language, and Hearing Research, 56*(2), 505–520. https://doi.org/10.1044/1092-4388(2012/12-0094)

Strand, E. A., Stoeckel, R., & Baas, B. (2006). Treatment of severe childhood apraxia of speech: A treatment efficacy study. *Journal of Medical Speech-Language Pathology, 14*(4), 297–307.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Terband, H., Maassen, B., van Lieshout, P., & Nijland, L. (2011). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. *Journal of Communication Disorders, 44*(1), 59–74. https://doi.org/10.1016/j.jcomdis.2010.07.003

Terband, H., van Brenk, F., van Lieshout, P., Niljand, L., & Maassen, B. (2009). Stability and composition of functional synergies for speech movements in children and adults. In *Proceedings of the INTERSPEECH 2009 Conference, ISCA Archive, Brighton, United Kingdom* (pp. 788–791).

Terband, H. R., van Zaalen, Y., & Maassen, B. (2013). Lateral jaw stability in adults, children, and children with developmental speech disorders. *Journal of Medical Speech-Language Pathology, 20*(4), 112–118.

van Lieshout, P. H. H. M., Starkweather, C. W., Hulstijn, W., & Peters, H. F. (1995). Effects of linguistic correlates of stuttering on EMG activity in nonstuttering speakers. *Journal of Speech and Hearing Research, 38*(2), 360–372. https://doi.org/10.1044/jshr.3802.360

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences, 39*(2), 163–173. https://doi.org/10.1080/01490400.2015.1127189

Weismer, G. (2008). Speech intelligibility. In M. J. Ball, M. Perkins, N. Müller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 568–582). Blackwell. https://doi.org/10.1002/9781444301007.ch35

Weismer, G., Yunusova, Y., & Bunton, K. (2012). Measures to evaluate the effects of DBS on speech production. *Journal of Neurolinguistics, 25*(2), 74–94. https://doi.org/10.1016/j.jneuroling.2011.08.006

Wellman, B., Case, I., Mengert, I., & Bradbury, D. (1931). Speech sounds of young children. In G. D. Stoddard (Ed.), *University of Iowa Studies in Child Welfare* (Vol. 5, pp. 1–82).

Whiteside, S. P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: A developmental study of motor speech skills in humans. *Neuroscience Letters, 347*(1), 29–32. https://doi.org/10.1016/S0304-3940(03)00598-6

Wilcox, K. A., & Morris, S. (1999). *Children's Speech Intelligibility Measure: CSIM*. The Psychological Corporation.

World Health Organization. (2007). *International Classification of Functioning, Disability and Health-Children and Youth Version*.

Yu, V. Y., Kadis, D. S., Oh, A., Goshulak, D., Namasivayam, A., Pukonen, M., Kroll, R., De Nil, L. F., & Pang, E. W. (2014). Changes in voice onset time and motor speech skills in children following motor speech therapy: Evidence from /pa/ productions. *Clinical Linguistics & Phonetics, 28*(6), 396–412. https://doi.org/10.3109/02699206.2013.874040

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control, 15*(1), 118–140. https://doi.org/10.1123/mcj.15.1.118

Appendix A

Modified Word Complexity Measure Scores for Probe Words

MSH stage	Number	Probe words	Syllable/word shape	Word patterns A	Word patterns B	Syllable structure C	Syllable structure D	Sound classes E	Sound classes F	Sound classes G	Sound classes H	Movement trajectory: Labial: rounding/retraction I	Movement trajectory: Lingual: anterior/posterior J	Sum
III	1	Ba	CV	0	0	0	0	0	0	0	0	0	0	0
III	2	Eye	V	0	0	0	0	0	0	0	0	1	0	1
III	3	Map	CVC	0	0	1	0	0	0	0	0	0	0	1
III	4	Um	VC	0	0	1	0	0	0	0	0	0	0	1
III	5	Ham	CVC	0	0	1	0	0	0	0	0	0	0	1
III	6	Papa	CV-CV	0	0	0	0	0	0	0	0	0	0	0
III	7	Bob	CVC	0	0	1	0	0	0	0	0	0	0	1
III	8	Pam	CVC	0	0	1	0	0	0	0	0	0	0	1
III	9	Pup	CVC	0	0	1	0	0	0	0	0	0	0	1
III	10	Pie	CV	0	0	0	0	0	0	0	0	1	0	1
IV	1	Boy	CV	0	0	0	0	0	0	0	0	1	0	1
IV	2	Bee	CV	0	0	0	0	0	0	0	0	1	0	1
IV	3	Peep	CVC	0	0	1	0	0	0	0	0	1	0	2
IV	4	Bush	CVC	0	0	1	0	0	0	1	0	1	0	3
IV	5	Moon	CVC	0	0	1	0	0	0	0	0	1	0	2
IV	6	Phone	CVC	0	0	1	0	0	0	1	0	1	0	3
IV	7	Feet	CVC	0	0	1	0	0	0	1	0	1	0	3
IV	8	Fish	CVC	0	0	1	0	0	0	2	0	1	0	4
IV	9	Wash	CVC	0	0	1	0	0	0	1	0	1	0	3
IV	10	Show	CV	0	0	0	0	0	0	1	0	1	0	2
V	1	Ten	CVC	0	0	1	0	0	0	0	0	0	1	2
V	2	Dig	CVC	0	0	1	0	1	0	0	0	0	2	4
V	3	Log	CVC	0	0	1	0	1	1	0	0	0	2	5
V	4	Owl	VC	0	0	1	0	0	1	0	0	1	0	3
V	5	Sun	CVC	0	0	1	0	0	0	1	0	1	1	4
V	6	Snake	CCVC	0	0	1	1	1	0	1	0	1	2	7
V	7	Juice	CVC	0	0	1	0	0	0	1	1	1	1	5
V	8	Clown	CCVC	0	0	1	1	0	1	0	0	0	1	4
V	9	Crib	CCVC	0	0	1	1	0	1	0	0	0	2	5
V	10	Grape	CCVC	0	0	1	1	1	1	0	0	0	0	4
VI	1	Cup cake	CVC CVC	0	0	1	0	2	0	0	0	0	1	4
VI	2	Ice cream	VC CCVC	0	0	1	1	1	1	1	0	0	2	7
VI	3	Toothbrush	CVC-CCVC	0	0	1	1	0	1	2	0	1	1	7
VI	4	Robot	CV-CVC	0	0	1	0	0	1	0	0	1	2	5
VI	5	Banana	CV-CV-CV	1	1	0	0	0	0	0	0	0	1	3
VI	6	Marshmallow	CVrC-CV-CV	1	0	0	0	0	2	1	0	1	2	7
VI	7	Umbrella	VC-CCV-CV	1	1	0	1	0	2	0	0	0	2	7
VI	8	Hamburger	CVC-CR-CR	1	0	1	0	1	2	1	0	0	1	7
VI	9	Watermelon	CV-CR-CV-CVC	1	0	1	0	0	2	0	0	1	2	7
VI	10	Rhinoceros	CV-CV-CV(R)-CVC	1	1	1	0	0	2	2	0	0	2	9

Scoring key:

Word patterns (A)	Word patterns (B)	Syllable (C)	Syllable structure (D)	Sound classes (E)	Sound classes (F)	Sound classes (G)	Sound classes (H)	Rounding or retraction present = 1 (I)	Number of anterior to posterior changes—lingual (J)
> 2 syllables = 1	Stress on any syllable but the first = 1	Word-final cons yes = 1	Each cons cluster = 1	Each velar cons = 1	Each liquid, syllabic liquid, rhotic vowel = 1	Each fricative or affricate = 1	If voiced fricative or affricate additional = 1		

AQ19

Appendix B

Participant Age, Gender, and Pre/Post Raw Data Across Variables (p. 1 of 2)

ID	Gender	Age (months)	PCC PRE	PCC SEVERITY PRE	PCC POST	VMPAC-FOC-PRE	VMPAC-SEQ-PRE	Phono. STD.PRE	CSIM PRE	CSIM POST	BIT PRE	BIT POST	PROBE Stage III PRE	PROBE Stage IV PRE	PROBE Stage V PRE	PROBE Stage VI PRE
1	F	44	59.7	moderate–severe	65.7	75.7	73.9	65.0	56.0	68.7	37.7	78.4	96.0	82.7	59.3	64.3
2	F	39	29.9	severe	31.3	51.1	43.5	85.0	31.3	30.7	5.8	35.1	88.0	69.3	35.6	35.7
3	F	39	53.7	moderate severe	58.2	57.1	41.3	75.0	40.7	42.7	13.2	26.7	82.0	88.0	52.5	50.4
4	F	39	52.2	moderate–severe	62.7	65.3	19.6	70.0	46.0	43.5	10.8	20.0	94.0	92.0	52.5	41.7
5	M	44	47.8	severe	56.7	73.5	60.9	55.0	48.7	42.0	29.8	33.3	94.0	69.3	55.9	57.4
6	F	55	53.7	moderate–severe	71.6	83.6	67.4	55.0	62.0	74.0	55.0	70.2	96.0	93.3	50.8	68.7
7	F	48	49.3	severe	58.2	67.2	76.1	55.0	31.3	42.7	61.3	59.7	94.0	85.3	71.2	80.0
8	M	50	17.9	severe	19.4	57.8	19.6	55.0	22.7	14.7	4.2	8.8	56.0	65.3	30.5	26.1
9	M	45	61.2	moderate–severe	68.7	68.7	43.5	80.0	56.5	46.7	20.7	46.5	86.0	77.3	57.6	44.3
10	F	40	43.3	severe		61.2	30.4	65.0	34.0		13.2		72.0	68.0	30.5	27.0
11	M	44	43.3	severe	65.7	76.1	63.0	65.0	32.7	36.7	9.6	12.5	60.0	70.7	30.5	37.4
12	M	46	22.4	severe	19.4	63.1	32.6	70.0	14.0	25.3	0.9	3.3	32.0	37.3	18.6	4.3
13	M	63	43.3	severe	73.1	75.7	65.2	55.0	31.9	58.7	9.6	20.8	88.0	85.3	54.2	63.5
14	M	48		severe	6.0								12.0	0.0	0.0	0.0
15	M	48	41.8	severe	50.7	72.8	78.3	65.0	64.0	54.7	18.4	39.2	72.0	73.3	39.0	47.0
16	M	52	32.8	severe	29.9	75.0	78.3	60.0	42.7	50.0	26.3	20.0	84.0	73.3	35.8	34.8
17	F	58	52.2	moderate–severe	53.7	72.8	82.6	65.0	57.3	46.7	25.4	28.3	70.0	80.0	49.2	47.0
18	M	75	22.4	severe	41.8	65.7	58.7	55.0	46.0	58.7	45.6	51.7	82.0	62.7	44.1	44.3
19	M	48	26.9	severe	43.3	69.0	58.7	55.0	25.3	30.0	21.1	7.5	44.0	53.3	39.0	19.1
20	F	38	22.4	severe	22.4	48.9	32.6	65.0	34.0	44.2	10.5	29.2	68.0	62.7	40.7	40.0
21	F	38	19.2	severe	59.7	33.2	4.3			50.0		19.3	58.0	17.3	23.7	0.0
22	M	48	70.1	mild–moderate		71.3	37.0	75.0	62.0		15.8		78.0	88.0	61.0	56.5
23	F	37	59.7	moderate–severe	58.2	70.5	58.7	70.0	56.0	46.0	40.5	38.6	82.0	86.7	54.2	41.7
24	M	48	35.8	severe	44.8	68.3	39.1	65.0	49.3		11.7		80.0	64.0	25.4	44.3
25	M	36	50.7	moderate–severe		49.6	45.7	75.0	53.1		6.1		82.0	74.7	45.8	64.3
26	F	39	37.3	severe	49.3	47.0	19.6	55.0	22.9	36.7	3.3	8.8	80.0	76.0	37.3	17.4
27	M	59	49.3	severe	56.7	72.0	47.8	55.0	37.4		22.8	40.0	68.0	74.7	40.7	56.5
28	M	68	49.3	severe	65.7	75.7	65.2	55.0	38.8	61.3	1.7	28.9	36.0	44.0	25.4	17.4
29	M	66	68.7	mild–moderate	70.1	79.5	78.3	55.0	63.3	60.0	35.8	47.4	90.0	98.7	76.3	67.0
30	M	42	28.4	severe	35.8	39.6	8.7	65.0	28.0	27.9	5.3	11.7	62.0	70.7	50.8	20.0
31	M	37	40.3	severe	52.2	53.0	28.3	75.0	26.8	37.8	5.3	2.6	64.4	60.0	39.0	21.7
32	M	47	34.3	severe		74.3	71.7	55.0	38.0		11.4		94.0	73.3	52.5	53.0
33	F	38	37.3	severe	55.2	56.3	28.3	60.0	29.3	31.3	1.8	25.8	86.0	66.7	40.7	33.0
34	M	48	58.2	moderate–severe	58.2	66.8	56.5	60.0	34.0	41.7	34.2	6.7	64.0	60.0	44.1	43.5
35	M	41	46.3	severe	55.2	62.3	32.6	70.0	32.0	36.1	15.8	16.7	72.0	80.0	42.4	25.2
36	F	43	9.0	severe	26.9	66.4	26.1	55.0	17.3	17.3	3.5	3.6	54.0	52.0	30.5	20.9
37	M	93	68.7	mild–moderate	73.1	85.8	82.6	55.0	64.7	68.0	36.0	64.9	76.0	77.3	74.6	46.1
38	F	38	11.9	severe	19.4	55.2	17.4	65.0	13.9	17.3	0.0	0.9	60.0	52.0	27.1	7.8
39	M	42	11.9	severe	17.9	54.5	17.4	55.0	12.2	34.7	2.6	7.5	42.0	36.0	20.3	20.9
40	F	73	88.1	mild	88.1	76.1	54.3	60.0	63.9	61.3	28.1	44.7	96.0	96.0	84.7	79.1
41	M	48	41.8	severe	41.9	81.0	71.7	55.0	34.1	40.0	11.4	53.5	70.0	69.3	37.3	47.0
42	M	51	44.8	severe	68.7	88.1	84.8	65.0	62.6	66.7	31.6	48.2	86.0	73.3	49.2	65.2

(*table continues*)

Appendix B

Participant Age, Gender, and Pre/Post Raw Data Across Variables (p. 2 of 2)

ID	Gender	Age (months)	PCC PRE	PCC SEVERITY PRE	PCC POST	VMPAC-FOC-PRE	VMPAC-SEQ-PRE	Phono. STD.PRE	CSIM PRE	CSIM POST	BIT PRE	BIT POST	PROBE Stage III PRE	PROBE Stage IV PRE	PROBE Stage V PRE	PROBE Stage VI PRE
43	M	37			28.4								10.0	0.0	0.0	0.0
44	F	44	62.7	moderate–severe	65.7	81.3	39.1	70.0	50.0	58.7	30.6	43.0	86.0	78.7	44.1	57.4
45	F	59	50.7	moderate–severe	73.1	92.5	84.8	55.0	54.0	73.3	55.8	70.3	90.0	96.0	69.5	73.9
46	F	39	34.3	severe	40.3	77.6	56.5	70.0	55.3	63.3	27.0	46.8	88.0	80.0	49.2	54.8
47	M	45	19.4	severe	31.3	59.0	23.9	55.0	21.1	25.6	2.6	2.6	32.0	38.7	22.2	6.1
48	M	55	52.2	moderate–severe	65.7	81.7	60.9	55.0	53.5	70.7	59.5	43.3	92.0	90.7	64.4	56.5

Note. Blank cells are missing data. PRE = preintervention or baseline data; POST = postdelay or postintervention data (see Namasivayam, Huynh, et al., 2020); PCC and PCC SEVERITY = Percentage of Consonants Correct, derived from the Diagnostic Evaluation of Articulation and Phonology Test (Dodd et al., 2002); VMPAC FOC and SEQ = Focal Oromotor Control (FOC) and Sequencing (SEQ) subsections of Verbal Motor Production Assessment for Children (VMPAC; Hayden & Square, 1999); Phono.STD.PRE = standard score from the Diagnostic Evaluation of Articulation and Phonology phonological assessment (Dodd et al., 2002); CSIM = Children's Speech Intelligibility Measure (Wilcox & Morris, 1999); BIT = Beginner's Intelligibility Test (Osberger et al., 1994); PROBE (Stages III–VI) = mean probe word scores (in percentage) across MSH stages; F = female; M = male.