Accepted manuscript for:

Author/s: McLeod, S. and Masso, S.

Title: Screening children's speech: The impact of imitated elicitation and word position

Journal: Language, Speech, and Hearing Services in Schools **ISSN**: 1558-9129

Year: 2019 **Volume:** 50 **Issue:** 1 **Pages:** 71-82

DOI to published version: http://dx.doi.org/10.1044/2018_LSHSS-17-0141

Screening children's speech: The impact of imitated elicitation and word position

Sharynne McLeod

Charles Sturt University, Australia

Sarah Masso

Charles Sturt University, Australia

The University of Sydney, Australia

Correspondence: Professor Sharynne McLeod, PhD, Charles Sturt University, Panorama Ave, Bathurst, NSW, 2795, Australia. Email: smcleod@csu.edu.au

Key words: screening, assessment, imitated, spontaneous, syllable, speech, speech sound disorders

Accepted for publication in Language, Speech, and Hearing Services in School

Abstract

Purpose: Diagnostic decision-making is influenced by the attributes of assessments. In order to propose time-efficient protocols for screening children's speech this study aimed to determine whether eliciting imitated responses and analyzing productions in different word positions resulted in different levels of consonant accuracy.

Method: Participants were 267 English-speaking preschool-aged children in the Sound Start Study whose parents were concerned about their speech. They were assessed using the International Speech Screener (ISS, McLeod, 2013) using either imitated or spontaneous elicitation. Productions were compared with an established diagnostic assessment of speech accuracy (Diagnostic Evaluation of Articulation and Phonology, DEAP-Phonology, Dodd et al., 2002).

Results: Participants' performance on the ISS was significantly correlated with performance on the DEAP-Phonology. Eliciting imitated productions on the ISS (*M* = 2:18 mins, *SD* = 0:59) took significantly less time than spontaneous productions (*M* = 6:32 mins, *SD* = 2:34). There was no significant difference in accuracy of imitated versus spontaneous productions in word-initial position; however, consonants were significantly less accurate in spontaneous than imitated productions in other word positions. Overall, participants had significantly lower consonant accuracy in word-initial position than within word or word-final positions. Examination of the influence of word position on test discrimination, using receiver operating characteristic (ROC) analyses, revealed acceptable test discrimination for percentage of consonants correct across word positions.

Conclusions: This research supports using imitated elicitation and analysis of percentage of consonants correct in word-initial position as a time-efficient procedure when screening the speech of English-speaking preschool children.

Key words: child, screening, assessment, speech, speech sound disorders

Screening children's speech: The impact of imitated elicitation and word position

Children with speech sound disorders (SSD)[1] constitute one of the largest subgroups of children seen by speech-language pathologists (SLPs) throughout the world (Broomfield & Dodd, 2004; Brumbaugh & Smit, 2013; McLeod & Baker, 2014; Mullen & Schooling, 2010; Oliveira, Lousada, & Jesus, 2015; Priester, Post & Goorhuis-Brouwer, 2009). In order to diagnose SSD, SLPs use a variety of assessment tasks including connected speech sampling, stimulability, and intelligibility ratings; however, the most commonly used assessment task relies on single-word sampling of consonants within the ambient language (and some assessments also include vowels, consonant clusters, and tones). For example, Skahan, Watson and Lof (2007) found that 74.1% of 309 participants in the United States always used a single-word speech assessment to determine percentile rank and standard scores based on the frequency and accuracy of consonant use. McLeod and Baker (2014) found that 88.9% of 231 participants in Australia always used a single-word speech assessment to determine sounds in error. Other studies describe specific speech assessments commonly used by SLPs with children, each of which includes single-word sampling. For example, Joffe and Pring (2008) indicated the majority (84.7%) of their 98 participants in the UK used the South Tyneside Assessment of Phonology (STAP; Armstrong & Ainley 1988); Priester et al. (2009) indicated the majority (70%) of their 85 participants in the Netherlands used Logo-Art (Baarda, de Boer-Jongsma, & Haasjes-Jongsma, 2001), and Oliveria et al. (2015) indicated that 68% of their 88 Portuguese

[1] Children with speech sound disorders "can have any combination of difficulties with perception, articulation/motor production, and/or phonological representation of speech segments (consonants and vowels), phonotactics (syllable and word shapes), and prosody (lexical and grammatical tones, rhythm, stress, and intonation) that may impact speech intelligibility and acceptability" (International Expert Panel on Multilingual Children's Speech, 2012, p. 1).

SLPs typically used Curso Teórico-Prático Sobre Articulação Verbal (Guimarães & Grilo, 1996). In a review of 30 commercially available speech assessments in 19 languages other than English, each speech assessment used single-word picture naming (McLeod & Verdon, 2014).

In contrast to diagnostic testing screening assessments "are intended to separate out the children who need further investigation of their speech and language skills from those with normally developing speech and language … Resources can then be concentrated on the children failing the screen" (Law Boyle, Harris, Harkness, & Nye, 1998, p. 38). First-level (primary) screening typically is undertaken by non-SLPs and second-level (secondary) screening is undertaken by SLPs to determine which children may need a more comprehensive assessment (Sturner, Layton, Evans, Heller, Funk, & Machon, 1994). There are no universally accepted protocols for screening children's speech (Nelson, Nygren, Walker, & Panoscha, 2006); however, single-word assessments frequently have been used (McLeod & Baker, 2017). Sturner et al. (1994) suggested that many screening measures rely on subjective judgments and lack data to provide predictive validity.

When selecting appropriate assessments to use with children (whether for the purposes of screening or diagnosis) it is important to determine whether assessments are valid (McCauley & Swisher, 1984; McLeod, 2012). Construct validity involves "the degree to which a test maps onto the theoretical construct it is supposed to assess" (Flipsen & Ogiela, 2015, p. 167; McCauley & Swisher, 1984) and content validity involves "the degree to which a test measures the relevant behavior" (Flipsen & Ogiela, 2015, p. 167). The construct considered in most children's speech assessments is the accuracy of consonants (Flipsen & Ogiela, 2015) and this construct has been correlated with children's intelligibility – the "relevant behavior" (McLeod, Crowe, & Shahaeian, 2015; McLeod, Harrison & McCormack, 2012). Diagnostic accuracy can be influenced by test design, assessment attributes, and protocols (Friberg, 2010; Macrae, 2017). Frequently, diagnostic accuracy of an assessment is tested against an established gold standard assessment. Diagnostic accuracy can be determined using sensitivity and specificity or can be tested by comparing a child's score to a normative sample (Flipsen & Ogiela, 2015; Peña, Spaulding, & Plante, 2006).

Speech-language pathologists frequently describe lack of time and resources as impacting their professional practice, including during assessment (McLeod & Baker, 2014). Time is even more crucial when undertaking screening assessments. The issue of time, or "what to assess when the clock is ticking" was considered during a clinical forum published in the American Journal of Speech-Language Pathology (Williams, 2002, p. 211) and the authors discussed the selection of appropriate assessments and analyses. Two aspects of speech assessment that contribute to the time taken are elicitation and analysis of the sample.

Elicitation of single-word assessments can either occur spontaneously or via imitation. Typically, a spontaneous production of a word is sought during assessment; however, if this is not possible, cueing hierarchies are used (e.g., providing the child with a clue, binary choice, or requesting imitation of the target word) (McLeod & Baker, 2017; McLeod et al., 1994). Imitated productions are likely to take less time to elicit than spontaneous productions, and may be used during assessments because young children may not have the relevant vocabulary or may avoid producing target words in spontaneous speech tasks (Leonard, Schwartz, Folger, & Wilcox, 1978). A number of studies have examined the impact of imitated versus spontaneous elicitation of target words during speech assessments and researchers have shown that imitated elicitation of target words either: results in more consonants correct than spontaneous productions (DuBois & Bernthal, 1978; Johnson & Somers, 1978; Kresheck & Socolofsky, 1972), makes no significant difference (Andrews & Fey, 1986; Powell, 1997), or is variable (Goldstein, Fabiano, & Iglesias, 2004). For example, Kresheck and Socolofsky (1972) administered the Templin-Darley Test of Articulation twice to 45 4-year-old children and found higher scores using imitated elicitation compared with spontaneous elicitation. Andrews and Fey (1986) tested 14 children with phonological impairment and found that clinical decision-making was similar regardless of the elicitation mode. In contrast, individual differences were found by Goldstein et al. (2004) who compared spontaneous and imitated productions of single words by 12 Spanish-speaking children with SSD: 62% of productions of consonants were identical, 25% of consonants were correct in spontaneous productions only, and 13% of consonants were correct in imitated productions only. To date, a large-scale study of the impact of elicitation modality (accuracy and time) has not been undertaken with children with SSD.

During a diagnostic assessment, it is essential to elicit and analyze all phonemes within the ambient language in all word positions (Flipsen & Ogiela, 2015; Friberg, 2010). Researchers recommend the provision of at least two opportunities to produce the target sound in words with a variety of phonotactic, syllable, and stress contexts (e.g., polysyllabic words, consonant clusters) (Bernhardt & Stemberger, 2000; Masso, McLeod, & Baker, 2018). In contrast, during screening assessments, it is important to elicit salient words that enable a high level of diagnostic accuracy, but also enable time-efficient analysis and diagnostic decision-making. Word-initial syllables have been described as more salient than within word and word-final syllables since they provide a large number of contrasts and are less likely to be reduced (Beckman, 1998). Support for the “word beginning saliency principle” (Cilibrasi, Stojanovik, & Riddell, 2015, p. 51) has been demonstrated with children with typical development, as well as those with language and reading disorders (Cilibrasi et al., 2015; Marshall & Van der Lely, 2009). When considering children's speech, typically developing English-speaking children are more likely to delete final consonants than initial consonants as they acquire adult-like speech (Smit, 1993). Rvachew and Andrews (2002) considered consonant accuracy across word positions for 10 preschool-aged children with SSD. Approximately half (46%) were correct in all three word positions, 25% produced different word-initial productions compared with word-final and within word productions, and 13% produced different within word productions compared with word-initial and word-final productions. Edwards, Fox, and Rogers (2002) assessed perception of final consonants by 105 children (aged 3-4, 5-6, and 7-8 years) and 22 adults and found articulatory-acoustic representations of final consonants were impacted by receptive vocabulary ability and articulatory accuracy. To date, there has not been a large-scale study of children with SSD considering the impact of word position, or the impact of imitated versus spontaneous productions on children's production accuracy in different word positions.

Context for the Current Study

The International Speech Screener: Research Version (ISS, McLeod, 2013) is being developed as a screening tool that can be used internationally. The current study sought to establish the criterion validity of the ISS by comparing the diagnostic accuracy and salient modality features of the ISS with the Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd et al., 2002), an established and valid measure of speech accuracy for English-speaking children. DEAP-Phonology, a single-word assessment, has been used in many research studies considering children with SSD, including the Sound Effects Study (McLeod, Harrison, McAllister, & McCormack, 2013) and the Sound Start Study (McLeod, Baker, et al., 2017).

Aims

1. To determine whether children's performance on the ISS (McLeod, 2013) was significantly correlated with performance on the DEAP-Phonology (Dodd et al., 2002) and to identify a suitable screening threshold for the ISS with optimum sensitivity and specificity. It was hypothesized that performance on the ISS would be correlated with the DEAP-Phonology.
2. To determine whether there was a significant difference in assessment time and consonant accuracy for children who produced the ISS word list spontaneously and those who produced it following an imitated model (PCC, PCC-initial, PCC-within word, PCC-final). It was hypothesized that imitated productions would take less time and be more accurate than spontaneous productions.
3. To determine whether there was a significant difference between total consonant accuracy (PCC) and consonants in different word positions (PCC-initial, PCC-within word, PCC-final) on the ISS and to identify suitable screening thresholds with optimum sensitivity and specificity for each of these measures. It was hypothesized that word-initial consonants would be more accurate.

METHOD

Participants

Participants in the current study were 267 children aged 48-66 months (Mean = 54.41; SD = 4.28) who were assessed as a part of the Sound Start Study (McLeod, Baker et al., 2017; McLeod, Crowe et al., 2017). There were 127 (47.57%) participants aged 4;0-4;5, 103 (38.58%) aged 4;6-4;11, and 37 (13.86%) aged 5;0-5;6. All the participants' parents and/or educators expressed concern about their speech skills. Of the total participants, 164 (61.4%) were male. The majority of the participants were monolingual English speakers (*n* = 201, 75.3%), 62 (23.2%) spoke English and one additional language, three (1.1%) spoke English and two additional languages, and one (0.4%) participant spoke English and three additional languages. All children who spoke English as an additional language were reported by their parents to have English language proficiency as good as, or better, than their other languages. Participants' socioeconomic status was considered according to the Australian Bureau of Statistics Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD, Australian Bureau of Statistics, 2008) where social and economic variables are considered to produce a final index score (1 – 10). An index score of 1 indicates an area of most disadvantage and an area of 10 indicates an area of most advantage. The participants in the current study were from a range of IRSAD deciles ranging from 1 to 10 with a mean of 5.72 (*SD* = 2.99).

Participant Recruitment

Early childhood education centers (*n* = 79) in Sydney, Australia were invited to participate in the Sound Start Study and consent for participation was obtained from the directors of 45 centers. A total of 1,920 children aged 4- to 5-years old was attending the participating centers when recruited. The parents of all children at participating centers were invited to complete a screening questionnaire and were completed by the parents of 1,205 (62.7%) children. Parents were invited to provide consent for their child's educator to complete a similar screening questionnaire and 1,123 (93.20%) consented; educators completed screening questionnaires for 1,064 (88.3%) children. Children were eligible for further assessment if parents and/or educators indicated that they had concerns about how the child "talks and makes speech sounds" (Glascoe, 2000) and also indicated that the child's speech was not clear. Children were also required to speak English as well as or better than any additional languages and not have an identified history of hearing loss, cleft lip and/or palate, or developmental delay.

A total of 327 children were eligible for direct screening by a speech-language pathologist (SLP) and participation in the current study. Of those eligible, 275 (84.10%) were assessed (reasons for attrition included a lack of parental consent, child assent, or a parent-reported diagnosis of childhood apraxia of speech. These factors are explained further in Figure 1 in McLeod, Baker et al., 2017). Although 275 children were assessed in the Sound Start Study, only 267 children completed the ISS (McLeod, 2013) with attrition primarily due to the children's ability to attend to and complete the full sequence of assessments administered over approximately 30-45 minutes (see McLeod, Crowe et al., 2017 for the full sequence of assessments).

Participants' nonverbal intelligence, language, hearing, and oromusculature were screened as part of the Sound Start Study. In the current study, all participants completed the Primary Test of Nonverbal Intelligence (PTONI, Ehrler & McGhee, 2008): 11 (4.1%) demonstrated very superior nonverbal intelligence, 23 (8.6%) superior, 38 (14.2%) above average, 85 (31.8%) average, 32 (12.0%) below average, 36 (13.5%) low, and 42 (15.7%) very low nonverbal intelligence. Most participants also completed the Preschool Language Scales – 5th Edition Screening Test (PLS-5S, Zimmerman, Steiner, & Pond, 2013) (n = 262). Based on participants' performance on the PLS-5S, 94 (35.9%) passed for language, 177 (68.0%) passed for articulation, 145 (55.3%) passed for connected speech, 249 (95.0%) passed for social skills, 258 (98.5%) passed for fluency and 245 (93.5%) passed for voice. Most participants (n = 253) underwent a hearing screening with 227 (89.72%) passing the screening tasks at 40dB across all frequencies (tested in noisy early childhood education settings) and 26 (10.27%) referred to audiology services due to failing the hearing-screening task. As with the other screening tasks, most (n = 258) completed an oromuscular assessment (Robbins & Klee, 1987). Based on participants' performance on the oromuscular assessment, 218 (84.5%) demonstrated oral structure within normal limits and 48 (18.6%) demonstrated oral function within normal limits.

Instruments

For the current study, participants' speech was evaluated using two assessment tools: the ISS (McLeod, 2013) and the DEAP-Phonology (Dodd et al., 2002).

International Speech Screen (ISS)

The ISS (McLeod, 2013) is a 54-item single-word picture-naming task that samples all English consonants and six English clusters (/dɹ, kl, ɡɹ, st, spl, skw/) on two occasions in word-initial position. Words vary in length and stress. The ISS was designed around ensuring that there were two examples of each word-initial consonant. Most pairs of words for each word-initial consonant contained a monosyllabic word and a bisyllabic or polysyllabic word (e.g., *dog, dinosaur*). Within these words most English vowels, within-word consonants, word-final consonants, and some word-final consonant clusters (i.e., /lz, mp, ŋk/) are sampled. Slightly more consonant clusters were elicited in word-initial position than in word-final position since word-initial consonant clusters are acquired later than word-final consonant clusters (e.g., McLeod, van Doorn & Reed, 2001). The ISS was presented via Powerpoint™ with one to two pictures/stimuli per slide. An audio-voice recording of the target word was available on each slide and played during the imitation condition. The ISS currently has not been normed on children with typically developing speech and language.

Diagnostic Evaluation of Articulation and Phonology

The DEAP consists of a Diagnostic Screen and four subtests: Articulation, Phonology, Inconsistency, and Oro-motor Ability. The DEAP-Phonology (Dodd et al., 2002) is a 50-item single-word picture-naming task that samples all English consonants and vowels as well as a number of consonant clusters, word lengths, shapes, and stress patterns. The DEAP-Phonology has been normed on a sample of children aged 3;0-6;11 from Australia (*n* = 144) and Britain (*n* = 684). Children's percentage of consonants correct (PCC), percentage of vowels correct (PVC), and percentage of phonemes correct (PPC) were calculated and converted to scaled scores and percentile ranks based on participants' ages. Participants who demonstrated a standard score equal to or less than 6 were classified as having speech accuracy outside the typical range based on the norms provided in the DEAP manual.

Procedure

This study was approved by the Charles Sturt University Human Research Ethics Committee (Approval number – 2013/070) and the NSW Department of Education and Communities State Education Research Applications Process (SERAP) (Approval number – 2013267). The director of each early childhood education center provided center-level consent and all parents provided direct consent for their child's participation. All children provided assent to participate in the assessments.

All assessments were conducted by one of two experienced speech-language pathologists (SLPs). These same SLPs were involved in the initial contact with each center as well as distribution and collection of all screening questionnaires. During the direct assessment session, a number of assessments were administered, including the ISS (McLeod, 2013) and the DEAP-Phonology (Dodd et al., 2002). Administration of the ISS was conducted following direct instruction from the test developer and the DEAP-Phonology was conducted according to the published assessment manual. Typically, the DEAP-Phonology was administered before the ISS. Participants were randomly allocated to complete the ISS in either one of two conditions: (1) direct imitation of a recorded voice, (2) spontaneously with a hierarchy of cueing as required. The cueing hierarchy was as follows: (1) asking "what's this?" (2) providing the child with a semantic clue, (3) providing a binary choice with the required response being the first item, then (4) requesting imitation of the target word (McLeod et al., 1994).

All assessments were administered in a quiet room at the participant's early childhood center. Each direct assessment was conducted over a period of 30-45 minutes and breaks were provided as often as required. Each assessment session was recorded using a Panasonic HC-V700 video camera and a Zoom H1 audio recorder. All speech samples were transcribed online using the International Phonetic Alphabet and then checked using audio recordings. Participants' total PCC, PVC, and PPC values were calculated using the same criteria for the ISS and DEAP. Standard scores were calculated for the DEAP (in order to determine the diagnostic classification) but were not calculated for the ISS. Raw scores of each measure (PCC, PVC, PPC, PCC-initial, PCC-within word, PCC-final) were used in the analysis.

Reliability

Reliability for each of the assessments was completed on a randomly selected sample of approximately 10% of the participants (ISS inter-judge n = 27, ISS intra-judge n = 21, DEAP-Phonology inter-judge n = 30, DEAP-Phonology intra-judge n = 30). ISS intra-judge agreement was 96.7% based on 4,494 phonemes and inter-judge agreement was 95.4% based on 5,561 phonemes. DEAP-Phonology intra-judge agreement was 91.5% and inter-judge agreement was 90.1% based on 6,629 phonemes for both calculations. A threshold greater than 85% was considered acceptable (Shriberg & Lof, 1991).

Data Analysis

Children's speech accuracy on the ISS and DEAP-Phonology were analyzed using the PROPH+ module of Computerized Profiling (Long, Fey, & Fletcher, 2008). Measures of percentage of consonants correct (PCC) and percentage of vowels correct (PVC) were extracted from the PROPH+ analysis for each test. Children's percentage of phonemes correct (PPC) was calculated for each test based on the raw PROPH+ output for consonant and vowel accuracy. Further analysis of the ISS was conducted to explore consonant accuracy at different word positions. Children's percentage of initial consonants correct on the ISS (PCC-initial) was calculated based on the PROPH+ consonant target analysis output (including consonant errors and accurate initial consonants). Similar analyses were undertaken for consonants that occurred within words (PCC-within word), and in word-final position (PCC-final).

A research assistant who was blinded to the purpose of the research and the condition (imitated versus spontaneous elicitation) calculated the time taken to complete the ISS for each participant. Administration time was calculated in minutes and seconds starting with each participant's production of the first item on the ISS and ending with production of the last item on the ISS.

All data were entered into, and analyzed, using SPSS version 24.0 (IBM, 2016). Descriptive statistics were extracted for the whole sample and tests of normality were conducted for all variables. Spearman's correlations were conducted to examine the correlation between PCC, PVC, and PPC between the two speech accuracy tasks (Aim 1). Mann Whitney U tests were conducted to determine whether there was a significant difference between the imitation group and the spontaneous group on: (1) administration time and (2) consonant accuracy across different word positions (Aim 2). Receiver Operating Characteristic (ROC) analyses were conducted to determine the overall discriminatory power of the ISS as a test of consonant accuracy (PCC, PCC-initial, PCC-within word, PCC-final) and determine a suitable screening threshold for the test based on an overall measure of PCC. The premise for the ROC curve analysis is that participants demonstrate typical or atypical performance on a reliable test of the target measure. In this case, participants' performance on the DEAP-Phonology subtest was the reference point for the ROC analysis. Participants were classified according to their performance on the DEAP-Phonology as being within normal limits (WNL, indicating a standard score of 7 or more for PCC on the Phonology subtest) or, not within normal limits (not-WNL, less than or equal to a standard score of 6). A measure of sensitivity indicates the true positive rate (i.e., not-WNL) and a measure of specificity indicates the true negative rate (i.e., WNL). The overall discriminatory power of a test is determined based on the value of the area under the ROC curve where a larger value (closer to 1) indicates better screening test accuracy. Hosmer, Lemeshow, and Sturdivant (2013) recommended that values of .50 indicate no test accuracy greater than chance, between .50 and .70 indicate poor test discrimination, and values greater than .70 indicate acceptable test discrimination. A value under the ROC curve of .90 or greater indicates outstanding test discrimination (Hosmeret al., 2013). Optimum sensitivity and specificity were identified using the Youden Index that is designed to calculate the maximum distance between specificity and sensitivity values (Hajian-Tilaki, 2013). The Youden index (*J*) is used to measure overall diagnostic effectiveness and is a function of specificity and sensitivity (Schisterman, Perkins, Liu, & Bondell, 2005) (Aims 1 and 3). Wilcoxon Signed Rank Tests were conducted to determine whether there were significant differences in consonant accuracy at different word positions (PCC, PCC-initial, PCC-within word, PCC-final) (Aim 3).

RESULTS

All 267 participants completed the ISS (McLeod, 2013) and DEAP-Phonology (Dodd et al., 2002). On the DEAP-Phonology, the average score for the participants' PCC was 72.57 (*SD* = 14.16) and ranged from 29.10 to 98.60 (see Table 1). All participants' parents and/or educators were concerned about their talking and making speech sounds. A total of 191 (71.54%)

demonstrated a standard score (based on PCC for the DEAP-Phonology) outside the typical range for their age; whereas, 76 (28.46%) participants demonstrated a standard score within normal limits.

Content Validity of the ISS

The content validity of the ISS was examined by determining whether there was a correlation between children's speech accuracy on the ISS and the DEAP-Phonology as measured by PCC, PVC, and PPC. Tests of normality indicated that none of the variables being examined were normally distributed. The measures calculated based on performance on the ISS indicated PCC skewness of -.873 (SE = .149) and kurtosis of .805 (SE = .297), PVC skewness of -1.907 (SE = .149) and kurtosis of 4.566 (SE = .297), PPC skewness of -.887 (SE = .149) and kurtosis of 12.726 (SE = .297). Similarly, the measures calculated based on performance on the DEAP-Phonology indicated PCC skewness of -.597 (SE = .149) and kurtosis of .174 (SE = .297), PVC skewness of -1.440 (SE = .149) and kurtosis of 2.588 (SE = .297), and PPC skewness of -.524 (SE = .149) and kurtosis of -.046 (SE = .297). See Table 1 for median values of all variables. Although median is typically a standard reported measure for non-normally-distributed data, mean has been used in this instance to remain consistent with previous reporting of speech accuracy utilising measures of PCC, PVC, and PPC. The total sample mean PCC on the ISS was 77.12 (*SD* = 12.76) and the DEAP-Phonology was 72.57 (*SD* = 14.16). The total sample mean PVC on the ISS was 94.55 (*SD* = 4.70) and the DEAP-Phonology was 92.75 (*SD* = 5.87). The total sample mean PPC on the ISS was 83.76 (*SD* = 10.47) and the DEAP-Phonology was 78.76 (*SD* = 10.96). Each measure was significantly correlated between two tests: PCC, $r_s = .881$, PVC, $r_s = .512$, and PPC, $r_s = .873$ (all *p*s <.001) (see Table 1).

Insert Table 1 here

Test Discrimination of the ISS

The second component of Aim 1 was to determine the overall test discrimination of PCC on the ISS compared with the established measure of speech accuracy (DEAP-Phonology: diagnostic threshold of WNL/not-WNL based on PCC standard scores). A ROC curve analysis was conducted (see Figure 1a) resulting in a total area of .912 (SE = .019) under the ROC curve indicating outstanding test discrimination (Hosmeret al., 2013). Sensitivity (.836) and specificity (.900) were identified with a Youden Index of .736 (Table 2). The corresponding PCC score at the optimum cutoff was 84.65%. Thus, the number of children correctly identified as having delayed or typical speech sound development was most accurate at a PCC threshold of 84.65% on the ISS. Using the identified cutoff of 84.65%, the number of true positive, true negative, false positive, and false negative diagnoses based on the ISS were extracted (see Table 3).

Insert Table 2 and Table 3 here

Impact of Spontaneous versus Imitated Elicitation

The second aim was to determine whether there was a significant difference between children in the spontaneous group and the imitated group based on: (1) administration time, and (2) consonant accuracy. In order to be sure that any differences across groups were due to the administration modality the groups were compared based on their performance on the DEAP (PCC and PVC scores). No significant difference was found between participants' performance on the DEAP in the imitated and spontaneous groups based on PCC (Spontaneous M = 71.97 SD = 14.21, Imitated M = 73.23 SD = 14.12, Z = -.666, p = .506) and PVC (Spontaneous M = 92.47 SD = 6.26, Imitated M = 93.07 SD = 5.40, Z = -.377, p = .706).

There was a significant difference in administration time between the two conditions with the spontaneous condition taking significantly more time than the imitated condition

(Spontaneous *M* = 6:32, *SD* = 2:34 [minutes:seconds], Imitated *M* = 2:18, *SD* = 0:59, *Z* = -12.616, *p* = .000).

Mann-Whitney tests were conducted to compare participants' ISS PCC, PCC-initial, PCC-within word, and PCC-final between spontaneous (*n* = 140) and imitated (*n* = 127) group conditions (see Table 4). Mean values are reported here and median values are reported in Table 4. There was a significant difference in PCC between the two conditions (Spontaneous *M* = 75.39 *SD* = 11.68, Imitated *M* = 79.02 *SD* = 13.49, *Z* = -2.251, *p* = .024). A significant difference was not present between participants in the imitated and spontaneous groups based on PCC-initial (Spontaneous *M* = 71.76 *SD* = 14.85, Imitated *M* = 73.10 *SD* = 14.47, *Z* = -.893, *p* = .372), but was present between groups based on PCC-within word (Spontaneous *M* = 71.43 *SD* = 14.89, Imitated *M* = 76.73 *SD* = 19.00, *Z* = -2.071, *p* = .038) and PCC-final (Spontaneous *M* = 81.60 *SD* = 11.21, Imitated *M* = 85.13 *SD* = 13.20, *Z* = -2.342, *p* = .019). There was a significant difference in participants' total PVC between the two conditions (Spontaneous *M* = 93.15 *SD* = 3.13, Imitated *M* = 96.10 *SD* = 5.41, *Z* = -5.094, *p* = .000). There was a significant difference in participants' total PPC in the two conditions (Spontaneous *M* = 81.83 *SD* = 10.53, Imitated *M* = 85.2 *SD* = 10.07, *Z* = -2.768, *p* = .006).

Insert Table 4 here

Impact of Word Position

The measures calculated based on performance on the ISS, accounting for consonant word position, indicated that none of the word position variables were normally distributed with PCC-initial skewness of -.725 (SE = .149) and kurtosis of .327 (SE = .297), PCC-within word skewness of -1.020 (SE = .149) and kurtosis of 1.196 (SE = .297), and PCC-final skewness of -1.415 (SE = .149) and kurtosis of 3.042 (SE = .297).

The third aim was to determine whether there was a significant difference in children's speech accuracy between PCC measures (PCC-initial, PCC-within word, PCC-final) in different word positions on the ISS. The mean PCC-initial was 72.41 (*SD* = 14.64), mean PCC-within word score was 73.95 (*SD* = 17.34), and mean PCC-final score was 83.28 (*SD* = 12.40). Wilcoxon Signed Rank Tests indicated that PCC-initial was significantly lower than PCC-within word, $z = -2.91$, $p = .004$ and PCC-final, $z = -12.10$, $p = .000$ respectively. Similarly, PCC-within word was significantly lower than PCC-final, $z = -10.42$, $p = .000$.

Insert Figures 1a-d here

The second component of the Aim 3 was to determine whether consonants produced in different word positions (ISS PCC, PCC-initial, PCC-within word, PCC-final) provide acceptable sensitivity and specificity when correlated with an established measure of speech accuracy (DEAP-Phonology PCC). Three additional ROC analyses were undertaken based on participants' ISS PCC-initial score (i.e., the percentage of word-initial consonants produced correctly) when compared to total DEAP-Phonology PCC score (see Figure 1b-d). The measure of PCC-initial indicated a total area of .896 (SE = .019) under the ROC curve indicating acceptable test discrimination. Sensitivity (.850) and specificity (.883) were identified with a Youden Index of .734. The corresponding PCC-initial score at the optimum cutoff was 81.65%. The measure of PCC-within word indicated a total area of .853 (SE = .025) under the ROC curve indicating acceptable test discrimination. Sensitivity (.791) and specificity (.829) were identified with a Youden Index of .620. The corresponding PCC-within word score at the optimum cutoff was 82.21%. The measure of PCC-final indicated a total area of .838 (SE = .027) under the ROC curve indicating acceptable test discrimination. Sensitivity (.894) and specificity (.650) were identified with a Youden Index of .544. The corresponding PCC-final score at the optimum cutoff was 94.37%. Thus, the number of children correctly identified as having delayed or typical speech sound development based on the ISS will be most accurate at a threshold of PCC-initial of 81.65%, PCC within word of 82.21%, and/or PCC-final of 94.37%. Using these identified cutoffs for each measure, the number of true positive, true negative, false positive, and false negative diagnoses based on the ISS was extracted (see Table 3).

DISCUSSION

The aims of this study were to determine whether children's performance on the ISS was correlated to performance on the DEAP-Phonology, to consider the impact of imitated elicitations, and word position on the outcomes and administration time of speech assessment using the ISS (McLeod, 2013). The results were: (1) participants' PCC, PVC, and PPC on the ISS was significantly correlated with their performance on an established speech assessment (DEAP-Phonology, Dodd et al., 2002), (2) all consonant measures on the ISS demonstrated acceptable (PCC-initial, PCC-within word, and PCC-final) to outstanding (PCC) test discrimination when compared to their DEAP-Phonology PCC standard score, (3) participants' consonant accuracy varied significantly depending on the word position measured with word-initial consonants identified as the least accurate and word-final consonants the most accurate, (4) there was a significant difference in accuracy of spontaneous versus imitated productions for PCC, PCC-within word, and PCC-final, but not PCC-initial, and (5) the imitated test condition was significantly faster to administer than the spontaneous condition. Each of these findings will be discussed in the following sections.

Criterion Validity of the ISS

The results of this study demonstrate that the ISS has good criterion validity when compared to the established measure of speech accuracy, the DEAP-Phonology (Dodd et al., 2002). The significant correlation between measures of PCC, PVC, and PPC between the ISS and the DEAP-Phonology indicated that the ISS is measuring behaviors relevant for the accurate diagnosis between speech sound disorders and typical speech development.

In determining the test discrimination of the ISS, this study has highlighted that different measures of consonant accuracy extracted from the ISS will generate similar confidence in the identification of speech sound disorders in a large sample of preschool children. The screening accuracy of the ISS was established against the DEAP-Phonology across the measures of PCC, PCC-initial, PCC-within word, and PCC-final. It is acknowledged that diagnostic accuracy also may be impacted by comparative normative sampling (Peña et al., 2006).

Impact of Imitated versus Spontaneous Elicitation

The results of this study demonstrated an administration effect when children were assessed using imitated stimuli. On the whole, the results of this study support previous investigations that have found shorter administration time for imitated elicitation and consonants to be less accurate in spontaneous administration conditions than imitated conditions (DuBois & Bernthal, 1978; Johnson & Somers, 1978; Kresheck & Socolofsky, 1972). The only measure identified in the current study that was resistant to the imitated versus spontaneous administration effect was PCC-initial. This result may suggest that either, the administration effect is variable (cf. Goldstein et al., 2004) or, the relative salience of consonants during an imitation task make phonological representation retrieval more accurate (cf. Edwards et al., 2004). These findings indicate that consonant accuracy in word-initial position on the ISS are correlated with PCC on the DEAP (an established measure of speech accuracy).

Impact of Word Position

In addition to demonstrating the sound psychometric properties of the ISS, this study has also revealed information regarding the impact of word position on consonant accuracy. Previous studies have identified that final consonants are more prone to deletion than consonants in other word positions (e.g., Rvachew & Andrews, 2002; Smit, 1993). The results of the current study suggest that the specific word-position errors may be masked when using broad measures of PCC (whether PCC, PCC-initial, PCC-within word, or PCC-final). However, this study also confirmed that the word-position most prone to consonant error on the ISS was the word-initial position. Children's consonant accuracy in initial position was significantly poorer than consonant accuracy across all other word positions, and the word as a whole. Interestingly, this study also revealed that PCC-initial was the only measure not sensitive to the administration effect between spontaneous and imitated administration conditions.

Clinical Implications

This study has identified that the International Speech Screener (ISS, McLeod, 2013) is a valid screening task to identify children with SSD using different measures of consonant accuracy and is sensitive to word-position and administration condition effects. There was no significant difference for PCC-initial between imitated and spontaneous conditions; however, children who produced the ISS using imitated elicitation demonstrated significantly higher consonant accuracy across PCC, PCC-within word, and PCC-final compared with children who produced words from the ISS words spontaneously. The results of this research support the use of PCC-initial as a time-efficient measure to screen the speech of English-speaking preschool children. As a broad measure of speech accuracy PCC-initial demonstrated adequate sensitivity and specificity at a clinical threshold of 81.65% to identify children with SSD and was resistant to changes in administration condition (i.e., spontaneous or imitated). This study also provides evidence to support the use of imitated elicitation procedures and analysis of word-initial consonants in the development of efficient screening protocols of children's speech.

Limitations

There were three primary limitations to this research. The first was that the sample of children represented in this research were not typically developing. Instead, they were identified as a likely clinical sample based on parent and/or early childhood educator concerns about communication development using the Parents' Evaluation of Developmental Status (Glascoe, 2000). Although this is unlikely to impact some elements of the included analysis (i.e., the ROC analyses), the inclusion of a clinically-skewed sample may impact the interpretation of children's speech accuracy across word positions. It would be valuable for future research to replicate this study with children who have typically-developing speech and language, particularly regarding the similarities and differences between spontaneous and imitated production accuracy and word-position effects.

The second and third limitations of this study lies with the comparative assessment used as an established measure of speech accuracy to identify children with speech sound disorders, the DEAP. The DEAP was standardized on a sample of 144 children from Australia (including only 25 children aged 4;0-5;5) and compared with a sample of 684 children from the UK (Dodd et al., 2002). The sample of children present in the current study represent a larger number of children, within a smaller age band than the DEAP Australian sample which may impact the quality of the diagnostic comparison. The third limitation lies in the comparison between a primarily spontaneous speech production task (DEAP) and the ISS that was administered to half of the participants as an imitated task and the other half as a spontaneous task. Although the aim of the current study was to compare the sensitivity of these tools and conditions, the results should be interpreted with caution as the testing conditions were not identical for each test.

Conclusion

The validity of the International Speech Screener (ISS, McLeod, 2013) as a tool for screening children's speech was examined for 267 English-speaking preschool children whose parents were concerned about their speech. The findings support using imitated elicitation and analysis of percentage of consonants correct in word-initial position as a time-efficient procedure when screening the speech of preschool children.

ACKNOWLEDGMENTS

This research was supported by the following sources: Australian Research Council Discovery Grant DP130102545 (awarded to Sharynne McLeod, Elise Baker, Jane McCormack, Yvonne Wren and Sue Roulstone) and the Charles Sturt University Faculty of Arts and Education Research Assistant Scheme. The authors thank Kate Crowe for data collection support and Charlotte Howland, Felicity McKellar, Grear McAdam, and Ninh Dang Vu for data entry.

REFERENCES

Armstrong, S. & Ainley, M. (1988). *South Tyneside Assessment of Phonology*. Northumberland, UK: Stass.

Australian Bureau of Statistics (2008). *An introduction to Socio-Economic Indexes for Areas (SEIFA) 2006*. Canberra, Australia: Australian Bureau of Statistics.

Baarda, D., de Boer-Jongsma, N. & Haasjes-Jongsma, W. (2001). *LOGO-art articulatieonderzoek* [Dutch]. Ternat/Axel, The Netherlands: Baert.

Beckman, J. N. (1998). *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*. New York, NY: Routledge.

Bernhardt, B. H., & Stemberger, J. P. (2000). *Workbook in nonlinear phonology for clinical application*. Austin, TX: ProEd.

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language and Communication Disorders, 39*(3), 303-324. doi:10.1080/13682820310001625589

Brumbaugh, K. M., & Smit, A. B. (2013). Treating children ages 3-6 who have speech sound disorder: A survey. *Language, Speech, and Hearing Services in Schools, 44*(3), 306-319. doi:10.1044/0161-1461(2013/12-0029)

Cilibrasi, L., Stojanovik, V., & Riddell, P. (2015). Word position and stress effects in consonant cluster perception and production. *Dyslexia, 21*(1), 50-59. doi:10.1002/dys.1488

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London, UK: Psychological Corporation.

Edwards, J., Fox, R. A., & Rogers, C. L. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. *Journal of*

Speech, Language, and Hearing Research, 45(2), 231-242. doi:1092-4388/02/4502-0231

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

Flipsen, J. P., & Ogiela, D. A. (2015). Psychometric characteristics of single-word tests of children's speech sound production. *Language, Speech, and Hearing Services in Schools, 46*(2), 166-178. doi:10.1044/2015_LSHSS-14-0055

Friberg, J. C. (2010). Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy, 26*(1), 77-92. doi:10.1177/0265659009349972

Glascoe, F. P. (2000). *Parents' Evaluation of Developmental Status: Authorized Australian Version.* Parkville, Australia: Centre for Community Child Health.

Goldstein, B., Fabiano, L., & Iglesias, A. (2004). Spontaneous and imitated productions in Spanish-speaking children with phonological disorders. *Language, Speech, and Hearing Services in Schools, 35*(1), 5-15. doi:10.1044/0161-1461(2004/002)

Guimarães, I. & Grilo, M. (1996). *Curso teórico-prático sobre articulação verbal* [A course in articulation]. Lisboa, Portugual: Fisiopraxis.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627–635.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. New York, NY: John Wiley & Sons.

IBM (2016). *IBM SPSS Statistics for Windows*, Version 24.0, Armok, NY: IBM Corp.

Joffe, V., & Pring, T. (2008). Children with phonological problems: A survey of clinical practice. *International Journal of Language and Communication Disorders, 43*(2), 154-

164. doi:10.1080/13682820701660259

Johnson, S., & Somers, H. (1978). Spontaneous and imitated responses in articulation testing. *International Journal of Language and Communication Disorders, 13*, 107–116. doi:10.3109/13682827809011332

Kresheck, J., & Socolofsky, G. (1972). Imitative and spontaneous articulation assessment of 4-year-old children. *Journal of Speech and Hearing Research, 15*, 729–733.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for speech and language delay: A systematic review of the literature. *Health Technology and Assessment, 2*(9), 1–183.

Leonard, L. B., Schwartz, R. G., Folger, M. K., & Wilcox, M. J. (1978). Some aspects of child phonology in imitative and spontaneous speech. *Journal of Child Language, 5*, 403-415. doi:10.1017/S0305000900002075

Long, S. H., Fey, M. E., & Channell, R. W. (2008). *Computerized profiling* (Version MS-DOS version 9.7). Milwaukee, WI: Marquette University.

Macrae, T. (2017). Stimulus characteristics of single-word tests of children's speech sound production. *Language, Speech, and Hearing Services in Schools, 48*, 219-233. doi:10.1044/2017_LSHSS-16-0050

Marshall, C. R., & Van der Lely, H. K. J. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, SLI and dyslexia. *Language, 85*, 39–57. doi:10.1353/lan.0.0081

Masso, S., McLeod, S., & Baker, E. (2018). Tutorial: Assessment and analysis of polysyllables in young children. *Language, Speech, and Hearing Services in Schools*, *49*(1), 42-58. doi: 10.1044/2017_LSHSS-16-0047.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*(1), 34-42. doi:10.1044/jshd.4901.34

McLeod, S. (2004). Speech pathologists' application of the ICF to children with speech impairment. *Advances in Speech-Language Pathology, 6*(1), 75-81. doi:10.1080/14417040410001669516

McLeod, S. (2012). Translation to practice: Creating sampling tools to assess multilingual children's speech. In S. McLeod & B. A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 144-153). Bristol, UK: Multilingual Matters.

McLeod, S. (2013). *International Speech Screen: Research Version*. Bathurst, Australia: Author.

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics and Phonetics, 28*(7-8), 508-531. doi:10.3109/02699206.2014.926994

McLeod, S., Baker, E., McCormack, J., Wren, Y., Roulstone, S., Crowe, K., . . . Howland, C. (2017). Cluster-randomized controlled trial evaluating the effectiveness of computer-assisted intervention delivered by educators for children with speech sound disorders. *Journal of Speech, Language, and Hearing Research, 60*(7), 1891-1910. doi:10.1044/2017_JSLHR-S-16-0385

McLeod, S., Crowe, K., Masso, S., Baker, E., McCormack, J., Wren, Y., . . . Howland, C. (2017). Profile of Australian preschool children with speech sound disorders at risk for literacy difficulties. *Australian Journal of Learning Difficulties, 22*(1), 15-33.

doi:10.1080/19404158.2017.1287105

McLeod, S., Harrison, L. J., McAllister, L., & McCormack, J. (2013). Speech sound disorders in a community study of preschool children. *American Journal of Speech-Language Pathology, 22*(3), 503-522. doi:10.1044/1058-0360(2012/11-0123)

McLeod, S., van Doorn, J., & Reed, V. A. (2001b). Consonant cluster development in two-year-olds: General trends and individual difference. *Journal of Speech, Language, Hearing Research, 44*, 1144-1171. doi:10.1044/1092-4388(2001/090)

McLeod, S., & Verdon, S. (2014). A review of 30 speech assessments in 19 languages other than English. *American Journal of Speech-Language Pathology, 23*(4), 708-723. doi:10.1044/2014_AJSLP-13-0066

Mullen, R., & Schooling, T. (2010). The National Outcomes Measurement System for pediatric speech-language pathology. *Language, Speech, and Hearing Services in Schools, 41*, 44-60. doi:10.1044/0161-1461(2009/08-0051)

Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the U.S. Preventive Services Task Force. *Pediatrics, 117*, e298–e319.

Oliveira, C., Lousada, M., & Jesus, L. M. (2015). The clinical practice of speech and language therapists with children with phonologically based speech sound disorders. *Child Language Teaching and Therapy, 31*(2), 173-194. doi:10.1177/0265659014550420

Peña, E. D., Spaulding, T. J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology, 15*, 247-254. doi:10.1044/1058-0360(2006/023)

Powell, T. W. (1997). Assessing consonant cluster production under imitative and more spontaneous conditions. *Perceptual and Motor Skills, 84*, 1134. doi:10.2466/pms.1997.84.3c.1134

Priester, G. H., Post, W. J., & Goorhuis-Brouwer, S. M. (2009). Problems in speech sound production in young children: An inventory study of the opinions of speech therapists. *International Journal of Pediatric Otorhinolaryngology, 73*(8), 1100-1104. doi:10.1016/j.ijporl.2009.04.014

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*, 271-277. doi: 10.1044/jshd.5203.271

Rvachew, S., & Andrews, E. (2002). The influence of syllable position on children's production of consonants. *Clinical Linguistics and Phonetics, 16*, 183-198. doi:10.1080/02699200110112222

Schisterman, E. F., Perkins, N. J., Liu, A., & Bondell, H. (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology, 16*, 73-81. doi:10.1097/01.ede.0000147512.81966.ba

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics, 5*(3), 225-279. doi: 10.3109/02699209108986113

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

Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research, 36*, 533-547. doi:10.1044/jshr.3603.533

Sturner, R. A., Layton, T. L., Evans, A. W., Heller, J. H., Funk, S. G., & Machon, M. W. (1994). Preschool speech and language screening: A review of currently available tests. *American Journal of Speech-Language Pathology, January*, 25–36.

Williams, A. L. (2002). Prologue: Perspectives in the assessment of children's speech. *American Journal of Speech-Language Pathology, 11*(3), 211–212.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2013). *Preschool Language Scale-5 Screening Test*. San Antonio, TX: Pearson.

Table 1

Correlation (r_s) Between the ISS (McLeod, 2013) and DEAP-Phonology (Dodd et al., 2002) for Percentage of Consonants Correct (PCC), Percentage of Vowels Correct (PVC) and Percentage of Phonemes Correct (PPC) (n = 267)

	ISS				DEAP-Phonology				Correlation	
Measure	Min	Max	Median	Mean (*SD*)	Min	Max	Median	Mean (*SD*)	r_s	p
PCC	29.20	98.50	78.10	77.12 (12.76)	29.10	98.60	74.50	72.57 (14.16)	.881	.000
PVC	71.60	100.00	96.20	94.55 (4.70)	56.00	100.00	93.70	92.75 (5.87)	.512	.000
PPC	40.74	98.60	84.98	83.76 (10.47)	47.30	98.60	79.50	78.76 (10.96)	.873	.000

ISS, International Speech Screener: Research Version (McLeod, 2013); DEAP-Phonology, Diagnostic Evaluation of Articulation and Phonology - Phonology (Dodd et al., 2002).

Table 2.

Sensitivity and Specificity between Percentage of Consonants Correct (PCC) Measures in Different Word Positions on the ISS (McLeod, 2013) Compared with PCC Standard Score on DEAP-Phonology (Dodd et al., 2002) (n = 267)

Measure	Total area	Standard error	Maximum Youden's Index	Threshold	Sensitivity	Specificity
PCC	.912	.019	.736	84.65%	.836	.900
PCC-initial	.896	.019	.734	81.65%	.850	.883
PCC-within word	.853	.025	.620	82.21%	.791	.829
PCC-final	.838	.027	.544	94.37%	.894	.650

Table 3

Crosstabulation Demonstrating True Positives, False Negatives, False Positives, and True Negatives for the ISS (McLeod, 2013) based on the DEAP-Phonology (Dodd et al., 2002) as the Reference (n = 267)

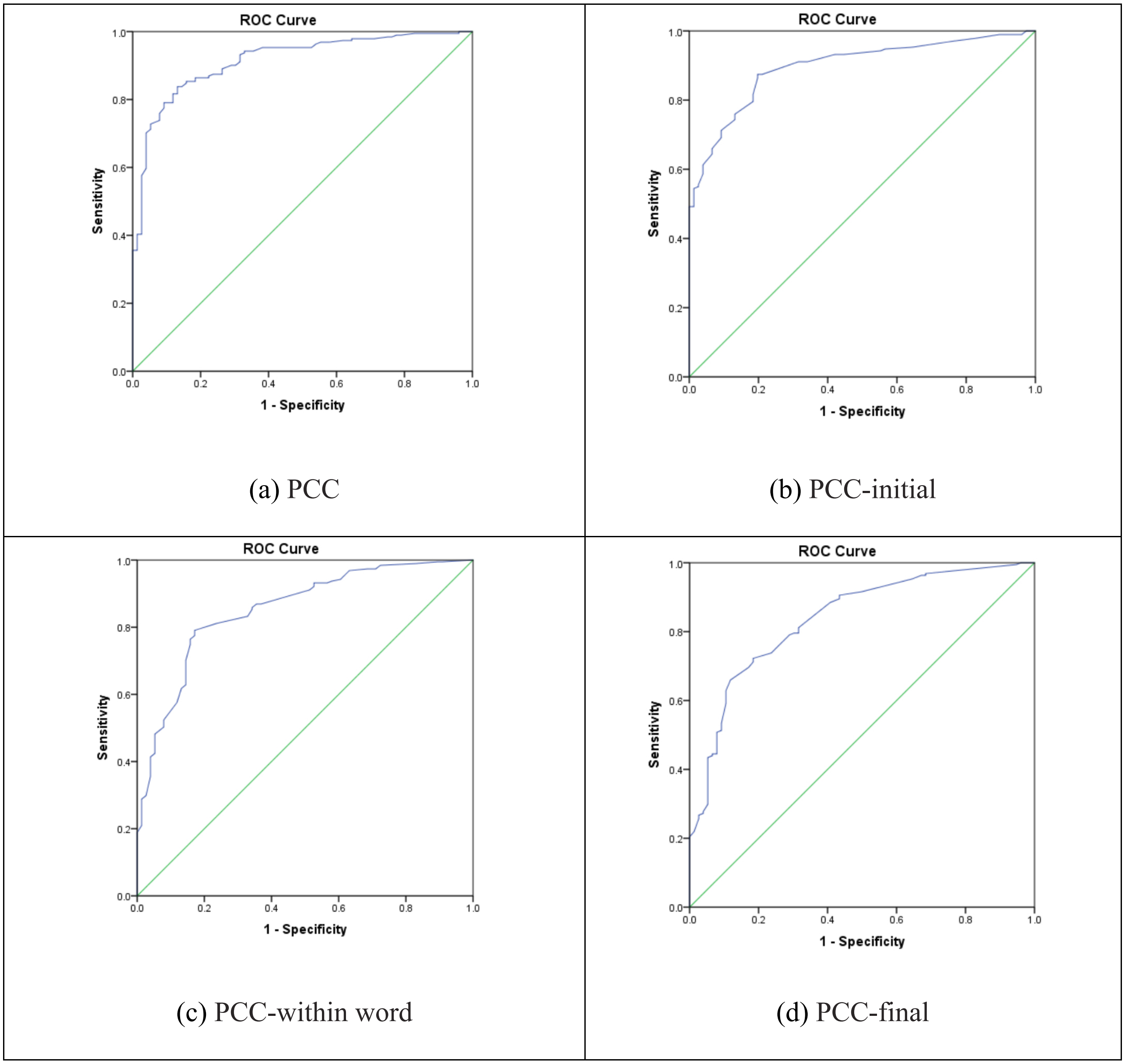
Measure	Classification	DEAP-Phonology (PCC)		
		Not-WNL	WNL	Total
ISS (PCC)	Not-WNL	165[a]	17[b]	182
	WNL	26[c]	59[d]	85
	Total	191	76	267
ISS (PCC-initial)	Not-WNL	167[a]	16[b]	183
	WNL	24[c]	60[d]	84
	Total	191	76	267
ISS (PCC-within word)	Not-WNL	151[a]	13[b]	164
	WNL	40[c]	63[d]	103
	Total	191	76	267
ISS (PCC-final)	Not-WNL	173[a]	33[b]	206
	WNL	18[c]	43[d]	76
	Total	191	76	267

[a] true positives, [b] false negatives, [c] false positives, [d] true negatives; WNL, within normal limits on the DEAP-Phonology; DEAP-Phonology, Diagnostic Evaluation of Articulation and Phonology - Phonology (Dodd et al., 2002); ISS, International Speech Screener: Research Version (McLeod, 2013).

Table 4.

Differences between Imitated and Spontaneous Productions of Single-Word Stimuli on the ISS (McLeod, 2013) by Word Position (n = 267)

						Mann-Whitney	
Measure	Group	*n*	*Mean*	*Median*	*SD*	*Z*	*p*
PCC	Imitated	127	79.02	80.30	11.68	-2.251	.024
	Spontaneous	140	75.39	76.80	13.49		
PCC-initial	Imitated	127	73.10	74.55	14.85	-.893	.372
	Spontaneous	140	71.76	74.55	14.47		
PCC-within word	Imitated	127	76.73	79.17	14.89	-2.071	.038
	Spontaneous	140	71.43	75.00	19.00		
PCC-final	Imitated	127	85.13	86.48	11.21	-2.342	.019
	Spontaneous	140	81.60	84.73	13.20		
PVC	Imitated	127	96.10	96.30	3.13	-5.094	.000
	Spontaneous	140	93.15	94.90	5.40		
PPC	Imitated	127	85.27	85.98	10.53	-2.768	.006
	Spontaneous	140	81.83	83.33	10.07		



(a) PCC

(b) PCC-initial

(c) PCC-within word

(d) PCC-final

Figure 1. Receiver operating characteristic (ROC) curve analysis to estimate the optimal sensitivity and specificity cut-off for PCC (PCC, PCC-initial, PCC-within word, PCC-final) calculated from the ISS (McLeod, 2013) when compared with PCC on the DEAP-Phonology (Dodd et al., 2002).