NIH Public Access
Author Manuscript
*Cleft Palate Craniofac J*. Author manuscript; available in PMC 2012 September 1.

Published in final edited form as:
*Cleft Palate Craniofac J*. 2011 September ; 48(5): 538–549. doi:10.1597/09-166.

# Reliability and Validity of a Computer Mediated Single-Word Intelligibility Test: Preliminary Findings for Children with Repaired Cleft Lip and Palate

**David J. Zajac, PhD[Associate Professor and Director]**,
Speech-Language Pathology, Craniofacial Center, Department of Dental Ecology, School of Dentistry, University of North Carolina at Chapel Hill.

**Caitrin Plante, BA[graduate student researcher]**,
Craniofacial Center, University of North Carolina at Chapel Hill. She completed portions of this research as part of an undergraduate Honors Thesis.

**Amanda Lloyd, BA**, and
currently a graduate student at the University of Virginia. She completed portions of this research as part of an undergraduate Honors Thesis at UNC-CH.

**Katarina L. Haley, PhD[Associate Professor]**
Division of Speech and Hearing Sciences, Department of Allied Health Sciences, School of Medicine, University of North Carolina at Chapel Hill.

## Abstract

**Objective—**To determine the reliability and validity of a computer-mediated, 50 word intelligibility test designed to be a global measure of severity of speech disability in children with repaired cleft lip and palate (CLP).

**Design—**A prospective between group design was used with convenience sampling of patients from a university craniofacial center.

**Participants—**Thirty-eight children between the ages of 4 and 9 years. Twenty-two had repaired CLP while 16 had no clefts. Twenty adults served as listeners.

**Main Outcome Measure(s)—**Speech intelligibility scores were calculated for repeated administrations of a single-word test based upon the number of correct orthographically transcribed words by 4 groups of 5 listeners per child. Measures of parallel forms, inter-listener, and intra-listener reliability were estimated; measures of construct validity were also determined.

**Results—**All measures of reliability were adequate. Parallel forms reliability of the test based upon mean scores from 5 listeners per child was high (r=.97). Thirty-seven of 38 children had differences between forms of 11 percentage points or less. Construct validity of the test was shown by a) significantly lower speech intelligibility scores for children with CLP than controls, and b) a moderately high correlation (r=.79) between intelligibility scores and percent consonants correct for all children.

Address Editorial Correspondence to: David J. Zajac, PhD, CB# 7450, Craniofacial Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, Phone: 919-966-1362, Fax: 919-966-6185, david_zajac@dentistry.unc.edu.

Portions of this research were presented at the Annual Convention of the American Speech-Language-Hearing Association, Chicago, IL, November 2008

**Conclusions—**A computerized, single-word intelligibility test was described which appears to be a reliable and valid measure of global speech deficits in children with CLP. Additional development of the test may further facilitate standardized assessment of children with CLP.

**Keywords**

speech intelligibility; articulation; cleft lip/palate; computerized testing

Speech intelligibility is the degree to which a speaker can be understood by a listener. Yorkston and Beukelman (1978) stated that intelligibility is the most important measure of a speech disorder and increasing intelligibility is the primary goal of therapy interventions. Children with structural defects such as cleft lip and/or palate (CL/P) may have significantly reduced speech intelligibility due to a number of factors. Typical speech characteristics of children with cleft palate may include hypernasality, audible nasal air emission, weak oral pressure consonants, and compensatory articulations such as glottal stops, pharyngeal stops, pharyngeal fricatives, and mid-dorsum palatal stops. In addition, alveolar sounds may be affected if clefts involve the lip and alveolus. Whitehill and Chau (2004) noted that even after surgical repair of the palate, a number of factors may affect speech production and impact intelligibility, including the presence of oral-nasal fistulae, coexisting hearing impairments, problems with dentition, and velopharyngeal inadequacy. In addition, established patterns of misarticulation may persist and affect intelligibility.

Kent et al. (1989) emphasized that speech intelligibility is a relative measure that “can vary with a host of variables, many of which pertain to the listener and the communicative environment as well as to the speaker.” Kent et al. cited Flanagan (1972) who indicated that non-speaker variables such as test material, personnel, training, and test procedures all may influence intelligibility. Konst et al. (2000) identified speaker factors such as “intonation, accent, stress, and rate” in addition to linguistic factors such as context and redundancy that can influence speech intelligibility. Prosodic and linguistic factors may be especially influential when sentence-level and/or conversational speech samples are used to estimate intelligibility.

Kent et al. (1989) noted that estimates of intelligibility in speech pathology have been obtained primarily by two methods – scaling procedures such as equal-appearing interval scales and item identification. McWilliams et al. (1990) observed that early studies that investigated intelligibility in children with repaired cleft palate employed item identification via transcription or write-down techniques (e.g., Prins and Bloomer, 1965, 1968; Fletcher, 1978). In general, these studies indicated a) clear differences in intelligibility between individuals with and without cleft palate, b) a tendency for articulation and intelligibility scores to be correlated (especially errors involving stop consonants), and c) a tendency for intelligibility and velopharyngeal (VP) function to be related. Although transcription-based studies generated important information, Subtelny et al. (1972) advocated the use of rating scales to evaluate intelligibility in speakers with cleft palate given that write-down techniques were too time-consuming.

Kent et al. (1989) further noted that “intelligibility assessment tends to be one of the most variable components in assessment protocols.” Indeed, Whitehill (2002) substantiated this statement relative to individuals with cleft palate. Based upon a review of 57 published articles between 1960 and 1998, Whitehill reported that intelligibility was assessed by various methods including global judgments (15.8%), rating scales (47.4%), articulation test scores (8.8%), transcription tasks (14.0%), other means (5.3%), and unspecified (8.8%). In addition, Whitehill expressed concern that the majority of studies employed interval rating scales which research has shown to be less valid measures of speech intelligibility.

Schiavetti (1992), for example, indicated that intelligibility is a prothetic construct that is not reliably represented by equal-appearing interval scales. Whitehill recommended that future studies employ both reliable and valid measures of intelligibility such as transcription, multiple-choice, and magnitude estimation tasks.

At least one study has directly compared equal-appearing interval scales and orthographic write-down techniques to evaluate intelligibility in children with CL/P. As part of a study to investigate the influence of presurgical infant orthopedics on later speech intelligibility, Konst et al. (2000) had 16 listeners orthographically transcribe and rate the spontaneous utterances of toddlers with CL/P and controls. The rating scale consisted of equal-appearing intervals ranging from 1 (unintelligible) to 10 (intelligible). While the investigators reported intelligibility differences among infants with CL/P based upon the rating scale, there were no differences based upon write-down techniques. Konst et al. questioned the validity of equal-appearing interval scaling of intelligibility in children with cleft palate and suggested that factors such as nasality and linguistic content may have influenced the listeners.

Recently, two studies have described speech intelligibility tests using single-word identification procedures for speakers with cleft palate. Whitehill and Chau (2004) investigated speech intelligibility in Cantonese speakers with repaired cleft palate using a multiple-choice word test based upon phonetic contrasts as described by Kent et al. (1989). Both children and adults were studied using 13 phonetic contrasts that were identified as being problematic for speakers with cleft palate. The contrasts that most contributed to reduced intelligibility scores were a) place of articulation for stops and nasals, b) stop versus fricative, and c) stop versus affricate. Whitehill and Chau reported that single-word intelligibility could be predicted with 87% accuracy using 2 contrasts, stop versus fricative and initial consonant versus null. The investigators noted, however, that “some speakers who had similar intelligibility scores had very different phonetic contrast error profiles,” which may suggest different underlying causes of unintelligibility. As further noted by Whitehill and Chau, a limitation of their study involved the relatively small number of speakers evaluated, especially young children under 12 years of age.

Hodge and Gotzke (2007) evaluated the construct and concurrent validity of a computer-mediated single-word intelligibility test called the Speech Intelligibility Probe for Children with Cleft Palate (SIP-CCLP). Fifteen English-speaking children between the ages of 3 years, 5 months and 6 years, 7 months were evaluated. Only 5 of the children, however, had repaired cleft palate. Hodge and Gotzke obtained three indexes of speech intelligibility: 1) a global index based upon 100 consecutively transcribed utterances from spontaneous speech, 2) a single-word intelligibility score using a multiple-choice (or closed set) approach, and c) a single-word intelligibility score using a write-down (or open set) approach. The single-word intelligibility scores were obtained using the computer program which was designed to facilitate elicitation and recording of target words and listener responses. The closed and open-set tests included 115 to 124 words per child that reflected underlying phonetic contrasts known to be problematic for children with CL/P based upon a comprehensive review of the literature. The targeted phonetic contrasts were: manner, place, sibilants, voicing, and consonant clusters. Hodge and Gotzke reported a high correlation between spontaneous speech intelligibility scores and scores obtained from the open-set task (r=.88). On average, the spontaneous speech index of intelligibility was 10 percentage points higher than the open-set word score. This difference most likely reflected the additional linguistic information available from connected speech. As expected, intelligibility scores from the closed-set (multiple-choice) task were the highest. Relative to phonetic contrasts, Hodge and Gotzke reported that for children with cleft palate, the majority of errors involved place (49%) followed by manner (35%).

Intelligibility tests that use single-word identification procedures as described above have multiple advantages over other techniques. First, as indicated by Schiavetti (1992), word-identification tasks in general are more valid measures of speech intelligibility than equal-appearing interval scales. Second, confounding factors related to conversational speech such as prosody and linguistic structure are largely eliminated in tests that use single-words. As suggested by Konst et al. (2000), these factors might bias listeners when rating scales are used. The elimination of linguistic structure may be a disadvantage, however, if one wishes to focus on understanding a child's spoken language in interactions with others. Third, if tests are constructed using minimal pairs based upon phonetic contrasts, then underlying causes of speech unintelligibility may be identified and targeted during remediation. Finally, and perhaps most important, single-word intelligibility tests that use standardized procedures including a common corpus of words may facilitate assessment of speakers across centers. Such tests, if found to be reliable and valid, may encourage collaborative research and/or clinical trials to evaluate behavioral and/or surgical treatment outcomes in children with cleft palate.

Given the advantages of single-word intelligibility tests, it is surprising that only two recent studies have used this approach with speakers with cleft palate. Of those studies, only one has used a computer-mediated approach, which may reduce test administration time. Although the findings of Hodge and Gotzke (2007) appear promising, only 5 children with cleft palate between the ages of 3 and 6 years were evaluated. The purpose of the present study was to determine the reliability and validity of a preliminary version of a computer-mediated, 50 word intelligibility test designed to be a global measure of severity of speech disability in children with repaired cleft lip and palate. Unlike the SIP-CCLP, the present test was modeled upon previous tests that use randomly selected words from a larger pool for each speaker (Morris et al., 1995; Yorkston and Beukelman, 1980). An open-set orthographic transcription task was used by listeners. The construction of the test, speaker recording procedures, and listener transcription procedures are described.

## Method

### Participants

**Speakers—**Speakers consisted of 38 children ranging in age from 4 to 9 years. Twenty-two children had repaired cleft lip and palate (CLP) without known syndromes (mean age=87 months, SD=17). Fourteen of these children had unilateral CLP; 8 had bilateral CLP. There were 15 girls and 7 boys. Sixteen of the children did not have CLP and served as controls (mean age=74 months, SD=18). There were 11 girls and 5 boys. All of the children learned English as a first language. Children in the control group were recruited to be in the age range of interest (4 to 9 years) but they were not age matched to the children with CLP. Because of an age difference between groups, we used age as a covariate in the statistical analyses as described later. All children underwent hearing screening, articulation screening, and pressure-flow assessment of velopharyngeal (VP) function as described in following sections.

**Listeners—**Twenty undergraduate students participated as listeners. The listeners ranged in age from 19 to 23 years and consisted of 12 females and 8 males. They passed the same pure tone hearing screening described below. All listeners reported English as their primary language and no history of speech or hearing anomalies.

All study procedures were approved by an Institutional Review Board that oversees research. Signed informed consent was obtained from the parents of all children and from the adult listeners. All participants were compensated monetarily.

## Hearing Screening

Hearing screening consisted of responding to pure tone frequencies of 500, 1000, 2000, and 4000 Hz presented at 25 dB HL in at least one ear. Children in the control group were required to pass hearing screening. Some children with CLP failed hearing screening as noted in Table 1.

## Articulation Screening

The articulation of all children was screened by the first author, an experienced speech-language pathologist, using the Preschool Screening form of the Analysis of Phonological Processes–Revised (APP-R) (Hodson, 1986). This form samples 28 consonants and three /r/ colored vowels in 12 words. Children in the control group were required to achieve a percent consonants correct (PCC) score of at least 80%. Using this criterion, some of the younger children in the control group exhibited developmental articulation errors (e.g., w/r substitutions and/or derhotacized /r/). None of the controls, however, reported receiving articulation therapy.

Thirty-five of the children were also recorded during articulation screening to calculate PCC scores as a measure of construct validity. Three children (2 controls and 1 child with CLP) were screened but audio recordings were not obtained. The children were recorded using an AKG head-mounted condenser microphone and a digital recorder (Marantz, model PMD670, Mahwah, NJ). The first author and a graduate student researcher with phonetic transcription training and cleft palate experience independently transcribed the recordings using procedures described by Shriberg and Kent (2003). Substitutions, omissions, common clinical distortions (e.g., derhotacized /r/, lateralized /s/), and audible/turbulent nasal air emission were considered as consonant errors. We included audible/turbulent nasal air emission as an error due to its potential impact on single-word identification. Point-by-point transcription agreement for the target consonants between the two transcribers was 94%. All disagreements were reviewed by both transcribers and resolved by consensus agreement. PCC scores of the children with CLP are listed in Table 1; PCC scores of the children in the control group are listed in Table 2.

## Screening of VP Function

All children were assessed using pressure-flow procedures described by Zajac (2000). All children in the control group were required to have adequate VP function. This was defined as having estimated VP closure during /p/ of “hamper” of 3 $mm^2$ or less (Zajac, 2000). Table 1 notes the VP status of the children with CLP.

## Intelligibility Test Construction

A 50-word speech intelligibility test was constructed modeled upon tests described by Morris et al. (1995) and Yorkston and Beukelman (1980). A corpus of 510 monosyllabic words was created and divided into 50 sets of phonetically similar words. Approximately 85% of the words were chosen from a corpus of words that are frequently used by children aged 4 to 5 years (Hall et al., 1984). The remaining words were judged to be age appropriate by the researchers and found to be familiar by adult listeners using the Neighborhood Database (Washington University, St. Louis). These words were scored a six or seven on a familiarity scale of one (low) through seven (high). All words were divided into 50 sets based on the initial consonant or consonant cluster that represented most of the English consonants. Each set contained between six and fourteen unique monosyllabic words (see Appendix).

We need to note that the test was not designed to target specific phonetic contrasts involving consonants known to be problematic to children with cleft palate. Rather, the test included most of the obstruent consonants which, in general, are problematic for children with cleft palate (Peterson-Falzone et al., 2001). We excluded the voiced fricative /z/ due to the limited number of age-appropriate words that begin with this sound. We also excluded the nasal consonants given that children with VP dysfunction typically do not have articulation difficulty with these sounds. Some children, however, may exhibit hyponasality on these sounds due to structural problems such as enlarged tonsils and/or reduced nasal area which in severe cases might reduce intelligibility. As described in later sections, the computer program is flexible and nasal consonants can be included depending upon specific assessment needs. Finally, some word sets were also designed to target the same consonant or consonant cluster in a low and high vowel environment. This was done given that nasalization has been shown to alter spectral properties of vowels which might affect identification (Philips and Kent, 1984).

## Speech Recording Procedures

Each child was recorded producing two unique 50-word tests in a sound-attenuated booth using Speech Measures (Haley, 2008). Speech Measures is a computer software program that was developed for quantitative assessment of speech intelligibility, acoustic speaking rate (i.e., syllables per second), and segmental duration. The program has modules that a) identify speakers and listeners, b) construct perceptual tests, c) elicit stimuli and record speakers, d) review and edit recordings, and e) run perceptual tests (e.g., orthographic transcription). Although the original target population was stroke survivors with aphasia and/or apraxia of speech, the program is designed to be flexible for use with other populations and to accommodate programming expansion and customization.

Elicitation procedures of Speech Measures include reading orthographically presented words, repeating words played by prerecorded audio files, or repeating words presented by a live speaker. We chose to use live elicitation for the present study. This was done because a) some of the younger participants could not read, and b) we anticipated that a live speaker would be better able to engage and maintain the attention of the younger children versus audio recordings. All children were instructed that they were to simply repeat words spoken by the investigator (i.e., to play a game of "copy cat"). As part of practice, the investigator would say a word and then point to the child as a prompt to repeat the word. Once a child demonstrated understanding of the procedures, the program randomly selected a word from each of the 50 sets and displayed the word on the screen of a laptop computer to the investigator. One of two investigators (the second and third authors) said the word and then prompted the child to repeat the word. Both investigators had what can be considered a general American dialect. Together, they reviewed all 510 words prior to the study and agreed upon standard pronunciations. A head-mounted miniature condenser microphone (AKG, model C-420, Northridge, CA) was used to record the words directly to the laptop computer. To ensure high-quality audio recordings, an external audio capture device (Edirol, model UA-25, Los Angeles, CA) was used to digitize the words at a sampling rate of 22.05 kHz with 24-bit resolution. The investigator controlled the recordings to the laptop via a mouse. Immediately following the elicitation and recording of the first 50-word test, a second test was randomly generated and recorded to the laptop by the same investigator. Following the completion of both tests, the investigator used the editing functions of Speech Measures to review all recorded words and cut inadvertent recordings of the investigator.

## Perceptual Testing Procedures

Four groups of 5 different listeners orthographically transcribed the recorded words of 9 to 10 children using the perceptual testing module of Speech Measures. Each listener transcribed a total of 900 to 1000 words (i.e., 9 to 10 children × 2 tests × 50 words). Each group of children included four controls and five to six children with CLP. We attempted to create groups of children who had fairly consistent ranges of PCC scores. This was done to limit the influence of variability on the strength of correlations within groups and is discussed later. We used multiple groups of children and listeners to a) limit the total transcription time of listeners to a reasonable period, and b) reduce the possibility of learning effects by the listeners. The order of presentation of the 50-word tests was conditionally randomized for each listener so that tests from the same child did not occur consecutively. All listeners heard a single presentation of the words in a sound-attenuated booth via headphones. If a listener was unable to understand a word and had no idea as to what the intended word was, he/she was instructed to type the word “nothing” into the computer. A typical transcription session lasted approximately 90 minutes. Listeners took short breaks following the completion of every 5 tests (i.e., 250 words). Three listeners from group 1 repeated the transcription of the same children approximately 3 weeks later.

Following transcription by each listener, the investigator used Speech Measures to manually check for homophones in the transcribed error responses (e.g., “bee” transcribed as “be”). A percent intelligibility score based upon the total number of correct responses was then automatically calculated by the program. These procedures resulted in each child receiving mean intelligibility scores for the 2 tests based upon the transcribed responses from 5 different listeners.

### Reliability Analyses

The reliability of the intelligibility testing procedures was evaluated in the following ways. First, parallel (or alternate) forms reliability was estimated for the 2 tests based upon both mean scores from 5 listeners per child and scores from individual listeners. Pearson Product Moment Correlation coefficients and the differences between tests were computed. Second, inter-listener reliability of the intelligibility scores was estimated by calculating intraclass correlation coefficients for each group of listeners. Finally, intra-listener reliability was estimated by computing Pearson Product Moment Correlation coefficients for the 3 listeners who repeated the same intelligibility tests 3 weeks following the initial procedures. We need to note that we selected only 3 listeners to repeat the intelligibility tests given that all listeners judged two parallel tests by each speaker. Thus, the estimates of parallel form reliability also included intra-listener reliability to some extent.

### Validity Analyses

Construct validity of the 50-word test was determined by 1) comparing the mean intelligibility scores between the children with CLP and controls, and 2) establishing the relationship between the mean intelligibility scores and PCC. Relative to the former, because age differences existed between the groups, we used an analysis of covariance (ANCOVA) with age as a covariate. Relative to the latter, we computed a Person Product Moment Correlation coefficient between the mean intelligibility and PCC scores for the 35 children who were audio recorded during articulation screening.

## Results

### Child Groups Relative to PCC

As noted previously, we attempted to create groups of children who had fairly consistent ranges of PCC scores. One child with CLP, however, had an extremely low PCC score compared to all other children, which made balancing groups problematic. Table 3 presents the means, standard deviations, low, and high PCC scores for the 4 groups of children. The table includes only the 35 children whose PCC scores were calculated from audio recordings. A one-way analysis of variance indicated that there was equal variance and no significant group differences for mean PCC scores ($F[1,3]=0.29$, $p=0.889$). Because the PCC scores were not distributed normally, we also conducted a Kruskal-Wallis one-way analysis of variance on ranks. This analysis also indicated no significant group differences (H[1,3]=0.0868, p=0.993). Although these tests suggest that the groups were relatively balanced, the increased range of PCC scores for children in group 2 may have influenced the strength of some correlations as noted below.

## Parallel Forms Reliability

Tables 4–7 list the children by groups, mean intelligibility scores for the 2 tests based upon all 5 listeners, and the absolute difference in mean intelligibility scores between the 2 tests. Across all children, the correlation coefficient between the mean scores of the 2 tests was r=.97 (p<.001). The mean difference between the tests was 1 percentage point (test 1 mean=75% [SD=16], test 2 mean=74% [SD=18]). Differences between scores of the 2 tests for individual children ranged from a low of 0 to a high of 18 percentage points. The single high difference of 18 percentage points for subject CLP8 (Table 7, group 4) appears to be an outlier and is discussed in following sections. Overall, 37 of 38 children (97%) had test score differences of 11 percentage points or less and 32 children (84%) had test score differences of 5 percentage points or less.

Tables 4–7 also list the intelligibility scores from the 5 individual listeners for each child for both tests. For the 10 children in group 1, the correlation coefficients between tests 1 and 2 ranged from .85 to .96 across the five listeners. The mean differences in scores between tests 1 and 2 for each listener (L) were as follows: L1 was 2 percentage points (range from 0 to 16), L2 was 5 percentage points (range from 0 to 16), L3 was 2 percentage points (range from 0 to 6), L4 was 1 percentage point (range from 0 to 12), and L5 was 2 percentage points (range from 0 to 20) [see Table 4].

For the 10 children in group 2, the correlation coefficients between tests 1 and 2 ranged from .94 to .99 across the five listeners. The mean differences in scores between tests 1 and 2 for each listener were as follows: L6 was 4 percentage points (range from 0 to 12), L7 was 1 percentage point (range from 0 to 16), L8 was 3 percentage points (range from 0 to 6), L9 was 4 percentage points (range from 0 to 10), and L10 was 2 percentage points (range from 0 to 16) [see Table 5].

For the 9 children in group 3, the correlation coefficients between tests 1 and 2 ranged from .78 to .92 across the five listeners. The mean differences in scores between tests 1 and 2 for each listener were as follows: L11 was 2 percentage points (range from 1 to 12), L12 was 1 percentage point (range from 0 to 10), L13 was 1 percentage point (range from 0 to 11), L14 was 1 percentage point (range from 0 to 11), and L15 was 2 percentage points (range from 4 to 10) [see Table 6].

For the 9 children in group 4, the correlation coefficients between tests 1 and 2 ranged from .82 to .94 across the five listeners. The mean differences in scores between tests 1 and 2 for each listener were as follows: L16 was 1 percentage point (range from 0 to 13), L17 was 2 percentage points (range from 0 to 16), L18 was 4 percentage points (range from 0 to 18), L19 was 1 percentage point (range from 0 to 18), and L20 was 1 percentage point (range from 0 to 24) [see Table 7]. The relatively large range of differences between tests for listeners in group 4 appears to be influenced by subject CLP8 and is discussed in a following section.

As noted previously, the increased range of PCC scores of children in group 2 may have influenced the strength of correlations for this group. Overall, the children in group 2 had the highest range of correlations across the 5 listeners. Even so, the differences in scores between tests 1 and 2 for the listeners of group 2 appeared to be similar to the listeners in the other groups. For example, the mean difference between tests for listeners in group 2 was 2.8 percentage points while it was 2.4 percentage points for listeners in group 1.

### Inter-Listener Reliability

Intraclass correlation coefficients (ICCs) were calculated to estimate inter-listener reliability for each group of 5 listeners as a function of tests 1 and 2. For listeners in group 1, ICCs were .97 for test 1 and .98 for test 2. For listeners in group, 2, ICCs were .99 for test 1 and .99 for test 2. For listeners in group 3, ICCs were .96 for test 1 and .94 for test 2. For listeners in group 4, ICCs were .98 for test 1 and .98 for test 2.

### Intra-Listener Reliability

Pearson correlation coefficients for L2, L3, and L4 of group 1 who repeated orthographic transcription of 10 children (20 intelligibility tests) were .95, .94, and .92 ($p<.001$), respectively. The mean differences between the repeated intelligibility scores were as follows: 5 percentage points for L2 (range from 0 to 18), 2 percentage points for L3 (range from 0 to 16), and 2 percentage points for L4 (range from 0 to 12). Across the three listeners, the majority of tests (55 of 60) differed by 10 percentage points or less.

### Construct Validity

Construct validity of the 50-word intelligibility test was determined by comparing children with CLP to the controls. For this analysis, we used the mean of the 2 tests for each child. The mean intelligibility score for the control children was 81% (SD=13, range from 59 to 95%); mean intelligibility for the children with CLP was 70% (SD=18, range from 15 to 90%). Because the children with CLP were a year older on average than the controls, we analyzed the groups using age as a covariate. The ANCOVA revealed a significant difference between groups with an adjusted mean of 85% for the controls and 67% for the children with CLP ($F[1, 35]=14.249$, $p<.01$).

Construct validity was also estimated by determining the relationship between the mean intelligibility scores and PCC for the 35 children who were audio recorded during articulation screening. There was a moderately high correlation between the measures for all 35 children ($r=.79$, $p<.001$). A similar correlation was obtained for the 21 children with CLP who were audio recorded ($r=.78$, $p<.001$).

### VP Function and Intelligibility Scores

Fifteen of the 22 children with CLP had adequate VP closure as determined by pressure-flow testing. As an additional analysis, we compared the children with CLP who had adequate VP closure to those who did not. Mean speech intelligibility for the children with adequate VP closure was 71% (SD=14) and for the children with incomplete VP closure it was 66% (SD=27). This difference was not significant ($t[20]=0.578$, $p>.05$).

### Number of Listeners and Intelligibility Scores

We also performed an analysis to determine if similar intelligibility scores would be obtained using a reduced number of listeners. We recalculated the speech intelligibility scores for the children in group 2 using all 10 possible combinations of 3 listeners. We selected group 2 because it contained 10 children and included the child (CLP7) with an extremely low PCC score. Table 8 presents the means and standard deviations of the intelligibility scores for the group based upon all 5 listeners and the 10 combinations of 3 listeners. A one-way analysis of variance indicated that there were no significant differences among any of the mean scores ($F[1,10]=0.009$, $p=1.0$).

## Discussion

The purpose of this study was to determine the reliability and validity of a preliminary version of a computer mediated, single-word speech intelligibility test administered to children with cleft lip and palate (CLP). The test was designed to provide a global measure of speech disability by sampling most of the English consonants. Although intelligibility may be considered the most important measure of a speech disorder and/or management outcome, there has been relatively few studies reporting data obtained using word identification approaches with children with CLP that use standardized procedures with known reliability and validity.

The results of this study suggest that the 50-word speech intelligibility test is reliable. Reliability of parallel forms of the test was high (r=.97) when intelligibility scores were derived from group responses of listeners. All but one of the children (37 of 38) received mean intelligibility scores based upon 5 listeners that differed by 11 percentage points or less on parallel forms. One child with CLP (subject CLP8) exhibited a mean difference between tests of 18 percentage points. Review of her recorded test words indicated that she tended to over emphasize and prolong each phoneme of words during test 1 but not during test 2. This child was currently receiving speech therapy. Although children were instructed to say the words normally, we believe that she started test 1 by using strategies learned in therapy which improved intelligibility but did not continue using these strategies during test 2. This child highlights the “relative” nature of speech intelligibility as noted by Kent et al. (1989). That is, even within a brief period of time, speaker characteristics may change, which ultimately affects intelligibility.

Reliability of parallel forms of the test was moderate to high (r=.78 to .99) when intelligibility scores were derived from individual listeners. Excluding listeners in group 4 who transcribed subject CLP8, the largest difference in test scores for any one listener was 20 percentage points (L5 in group 1). Across the 15 listeners in groups 1–3, the majority of test scores (130 of 145) differed by 11 percentage points or less. These findings are similar to results reported by Morris et al. (1995) for a 50-word intelligibility test that targeted children without cleft palate. In that study, the largest difference between parallel forms of the test was 16 percentage points for one listener and 14 percentage points for a second listener.

Construct validity of the 50-word speech intelligibility test was demonstrated by poorer scores for children with CLP than controls and a moderately high correlation with articulation ability as reflected by PCC scores (r=.79). Adjusting for age, children with CLP scored approximately 20 percentage points lower than controls. This occurred even though the majority of children with CLP exhibited adequate velopharyngeal (VP) closure as discussed in a following section.

There have been few studies of English-speaking children with cleft palate that have used computer-mediated assessment of single-word intelligibility. Hodge and Gotzke (2007) reported a mean single-word intelligibility score of 54.6% for five children with cleft palate and a mean of 79.1% for 10 children without cleft palate using an open-set transcription task. While comparison to the present study is problematic due to differences in test design, procedures, and number of subjects, the mean intelligibility scores are roughly similar in that our adjusted mean for children with CLP was 67% and the adjusted mean for controls was 85%.

Previous studies have also reported that VP function is related to intelligibility. In the present study, 15 of the 22 children with CLP had adequate VP closure as determined by pressure-flow testing. Mean speech intelligibility for the children with adequate VP closure

(71%) was not significantly different from the children with incomplete VP closure (66%). We must note, however, that there were few children with incomplete VP closure and we used a VP area cutoff criterion (3 $mm^2$) that was based solely on normative data for children, not perceptual symptoms. Regardless, these findings suggest that the current test was sensitive to speech characteristics other than obligatory symptoms associated with VP dysfunction. Given that all of the children with cleft palate also had cleft lip and alveolus, we believe that structural defects of the primary palate such as maxillary collapse, anterior dental crowding, missing teeth, malocclusion, and/or dental cross bite may have affected intelligibility. Additional research using single-word tests designed specifically to target alveolar-palatal phonetic contrasts is needed to investigate this possibility.

As stated by Whitehill (2002), “there remains a need for global measures of speech performance such as speech intelligibility.” We believe that the use of single-word speech intelligibility measures have the potential to fill this need. Intelligibility tests that use standardized elicitation and recording procedures along with a common corpus of words might greatly facilitate outcome comparisons across centers, at least when a common language is involved. The use of single words and computer-assisted recording and perceptual testing procedures can also reduce the “time consuming” nature of write-down procedures. A single speaker, for example, can be recorded saying 50 words in less than 10 minutes. Likewise, a single listener can transcribe 50 words in a similar amount of time. Obviously, the use of multiple listeners will necessitate greater testing time. Based upon the additional results reported, it appears that as few as 3 listeners may provide reliable estimates of intelligibility. Indeed, Morris et al. (1995) suggested that a single clinician could make reliable estimates of intelligibility, even of the same speaker over time, using randomly generated tests of 50 words. We need to note, however, that the test described by Morris et al. used a larger corpus of words (600), an equal number of words per set (12), and closed set responses. Ultimately, the use of automatic speech recognition systems (e.g., Schuster et al., 2006) may eliminate the need for human listeners entirely.

### Limitations and Future Test Development

The current version of our 50-word intelligibility test used fewer total words than similar tests described in previous studies. This occurred because we attempted to select words that were either in the vocabulary of and/or familiar to younger children. Given that the reliability of parallel forms depends upon random selection of items from a larger pool, reliability might be reduced for some generated tests when sets contain an unequal number of words. This possibility, however, can be reduced by increasing the number of total words in the test and having equal numbers per word set.

Because we used two different investigators to elicit words from the children, modeling inconsistencies may have occurred, which might also reduce reliability. Although both investigators “calibrated” their production of the words prior to the study, we cannot rule out that some inadvertent inconsistencies occurred. Speech Measures includes the option of using prerecorded audio files and/or pictures for elicitation. Use of these options in future tests would further standardize the elicitation procedures. Hodge and Gotzke (2007), for example, reported using picture stimuli in combination with live modeling during elicitation procedures for the SIP-CCLP.

We also used a limited number of listeners to repeat the entire perceptual testing procedures to estimate intra-listener reliability. Although intra-listening reliability was high, a larger number of listeners should be employed to evaluate future versions of tests.

Finally, the current study did not include measures of concurrent validity such as determining intelligibility during spontaneous speech. Future tests will need to show such validity measures along with test sensitivity and specificity characteristics.

## Summary

We describe a preliminary version of a single-word speech intelligibility test that may facilitate evaluation of children with repaired cleft lip and palate. The test is computer-based which facilitates both recording of children and orthographic transcription by listeners. The test was shown to have adequate parallel forms, inter-listener, and intra-listener reliabilities. The test was also shown to have construct validity. We believe that further refinement and adoption of such tests by clinicians and researchers will lead to more standardized speech intelligibility testing in children with cleft lip and palate, which may ultimately facilitate our understanding of speech disorders and/or treatment options in this population. Finally, because of the flexibility of the computer program, tests can be created and/or easily modified for use with other populations such as children without cleft palate who exhibit speech disorders.

## Acknowledgments

Supported by Grant Numbers R56 DE018004 from the National Institute of Dental & Craniofacial Research and R03 DC006163 from the National Institute on Deafness and Other Communication Disorders

## References

Flanagan, JL. Speech analysis, synthesis, and perception. New York: Springer-Verlag; 1972.

Fletcher, SG. Diagnosing speech disorders from cleft palate. New York: Grune and Stratton; 1978.

Haley, KL. Computer-Mediated Assessment of Speech Intelligibility in Aphasia and AOS; Presented at the Annual Convention of the American Speech Language Hearing Association; November 2008; Chicago, IL.

Hall, WS.; Nagy, WE.; Linn, R. Spoken words: Effects of situation and social group on oral word usage and frequency. Hillsdale: Lawrence Erlbaum Associates, Inc.; 1984.

Hodge M, Gotzke CL. Preliminary results of an intelligibility measure for English-speaking children with cleft palate. Cleft Palate Craniofac J. 2007; 44:163–174. [PubMed: 17328639]

Hodson, B. The assessment of phonological processes – revised. Danville, IL: Interstate; 1986.

Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. J Speech Hear Dis. 1989; 54:482–499.

Konst E, Weersink-Braks H, Rietveld T, Peters H. An intelligibility assessment of toddlers with cleft lip and palate who received and did not receive presurgical infant orthopedic treatment. J Commun Disord. 2000; 33:483–501. [PubMed: 11141029]

McWilliams, BJ.; Morris, HL.; Shelton, RL. Cleft Palate Speech. 2nd Edition. Philadelphia: BC Decker; 1990.

Morris SR, Wilcox KA, Schooling TL. The preschool speech intelligibility measure. Am J Speech Lang Pathol. 1995; 4(4):22–27.

Neighborhood Database. St Louis: Washington University; Available at: http://128.252.27.56/neighborhood/Home.asp.

Peterson-Falzone, S.; Hardin-Jones, M.; Karnell, M. Cleft Palate Speech. 3rd Edition. St. Louis: Mosby; 2001.

Philips, BJ.; Kent, R. Acoustic-phonetic descriptions of speech production in speakers with cleft palate and other velopharyngeal disorders. In: Lass, N., editor. Speech and language: advances in basic research. New York: Academic Press; 1984.

Prins D, Bloomer H. A word intelligibility approach to the study of speech change in oral cleft patients. Cleft Palate J. 1965; 2:357–368. [PubMed: 5213338]

Prins D, Bloomer H. Consonant intelligibility: A procedure for evaluating speech in oral cleft subjects. J Speech Hear Res. 1968; 11:128–137. [PubMed: 5648527]

Schiavetti, N. Scaling procedures for the measurement of speech intelligibility. In: Kent, RD., editor. Intelligibility in Speech Disorders: Theory, Measurement and Management. Philadelphia: John Benjamins; 1992.

Schuster M, Maier A, Haderlein T, Nkenke E, Wohlleben U, Rosanowski F, Eysholdt U, Noth E. Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition. Int J Pediatr Otorhinolaryngol. 2006; 70:1741–1747. [PubMed: 16814875]

Shriberg, LD.; Kent, RD. Clinical Phonetics. 3rd Edition. Boston: Allyn & Bacon; 2003.

Subtelny JD, Van Hattum RJ, Myers BA. Ratings and measures of cleft palate speech. Cleft Palate J. 1972; 9:18–27. [PubMed: 4500292]

Whitehill TL. Assessing intelligibility in speakers with cleft palate: A critical review of the literature. Cleft Palate Craniofac J. 2002; 39:50–58. [PubMed: 11772170]

Whitehill TL, Chau C. Single-word intelligibility in speakers with repaired cleft palate. Clin Linguist Phon. 2004; 18:341–355. [PubMed: 15259575]

Yorkston KM, Beukelman DR. A comparison of techniques for measuring intelligibility of dysarthric speech. J Commun Disord. 1978; 11:499–512. [PubMed: 739065]

Yorkston KM, Beukelman DR. A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. J Commun Disord. 1980; 13:15–31. [PubMed: 7354139]

Zajac D. Pressure-flow characteristics of /m/ and /p/ production in speakers without cleft palate: Developmental findings. Cleft Palate Craniofac J. 2000; 37:468–477. [PubMed: 11034029]

**Table 1**

Age, gender (M=male, F=female) cleft type (UCLP=unilateral cleft lip and palate, BCLP=bilateral cleft lip and palate), hearing status (P=pass, F=fail), percent consonants correct (PCC), and velopharyngeal (VP) function (A=adequate, I=incomplete) of the children with cleft lip and palate (CLP).

| Subject | Age (mos.) | Gender | Cleft Type | Hearing | PCC | VP |
|---|---|---|---|---|---|---|
| CLP1 | 60 | F | UCLP | F | 77 | I |
| CLP2 | 64 | M | BCLP | P | 94 | A |
| CLP3 | 66 | F | UCLP | P | 94 | A |
| CLP4 | 66 | F | UCLP | F | 84 | A |
| CLP5 | 69 | M | UCLP | P | 90 | A |
| CLP6 | 69 | F | UCLP | P | 100 | A |
| CLP7 | 70 | M | BCLP | F | 32 | I |
| CLP8 | 80 | F | BCLP | P | 77 | I |
| CLP9 | 85 | F | UCLP | P | 94 | A |
| CLP10 | 85 | M | UCLP | P | 77 | A |
| CLP11 | 86 | F | UCLP | P | 90 | A |
| CLP12 | 86 | F | BCLP | P | 80 | I |
| CLP13 | 88 | F | UCLP | P | 87 | A |
| CLP14 | 93 | F | UCLP | P | 74 | A |
| CLP15 | 96 | M | BCLP | P | 100 | A |
| CLP16 | 97 | F | UCLP | P | 90 | A |
| CLP17 | 99 | F | BCLP | P | 71 | A |
| CLP18 | 106 | F | BCLP | P | 71 | A |
| CLP19 | 109 | M | UCLP | P | 77* | I |
| CLP20 | 109 | M | BCLP | P | 100 | A |
| CLP21 | 111 | F | UCLP | P | 97 | I |
| CLP22 | 113 | F | UCLP | P | 97 | I |

* not audio recorded – PCC determined during screening

## Table 2

Age, gender (M=male, F=female), and percent consonants correct (PCC) of the children without cleft lip and palate in the control (C) group.

| Subject | Age (mos.) | Gender | PCC |
|---|---|---|---|
| C1 | 49 | F | 90 |
| C2 | 50 | M | 81 |
| C3 | 59 | F | ≥80* |
| C4 | 61 | M | 97 |
| C5 | 63 | F | 100 |
| C6 | 64 | M | 100 |
| C7 | 65 | F | 87 |
| C8 | 73 | M | 100 |
| C9 | 74 | F | 100 |
| C10 | 76 | F | ≥80* |
| C11 | 77 | F | 100 |
| C12 | 80 | F | 100 |
| C13 | 86 | M | 100 |
| C14 | 93 | F | 100 |
| C15 | 96 | F | 97 |
| C16 | 115 | F | 100 |

*not audio recorded – PCC determined during screening

**Table 3**

Means, standard deviations (SD), and range (low and high) of percent consonants correct (PCC) scores of children in the four listening groups.

| | PCC Scores (%) | | | |
|---|---|---|---|---|
| **Group** | **Mean** | **SD** | **Low** | **High** |
| 1 (n=9) | 92 | 8 | 74 | 100 |
| 2 (n=10) | 87 | 21 | 32 | 100 |
| 3 (n=8) | 90 | 12 | 71 | 100 |
| 4 (n=8) | 90 | 10 | 77 | 100 |

## Table 4

Speech intelligibility scores of subjects in group 1 as judged by listeners L1–L5. ADM=absolute difference between means of Test 1 and 2.

| | Group 1 Intelligibility Scores (%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Test 1 | | | | | | Test 2 | | | | | | |
| **Subject** | **L1** | **L2** | **L3** | **L4** | **L5** | **Mean** | **L1** | **L2** | **L3** | **L4** | **L5** | **Mean** | **ADM** |
| CLP2 | 84 | 84 | 86 | 76 | 84 | 83 | 84 | 82 | 86 | 80 | 84 | 83 | 0 |
| CLP3 | 48 | 50 | 50 | 38 | 26 | 42 | 46 | 46 | 50 | 44 | 46 | 46 | 4 |
| CLP11 | 70 | 70 | 72 | 74 | 78 | 73 | 74 | 70 | 74 | 78 | 66 | 72 | 1 |
| CLP13 | 80 | 76 | 74 | 66 | 70 | 73 | 76 | 68 | 74 | 60 | 64 | 68 | 5 |
| CLP14 | 62 | 58 | 62 | 47 | 56 | 57 | 50 | 48 | 56 | 38 | 50 | 48 | 9 |
| CLP15 | 85 | 92 | 92 | 86 | 88 | 89 | 86 | 88 | 88 | 90 | 92 | 89 | 0 |
| C3 | 70 | 78 | 66 | 76 | 70 | 72 | 82 | 72 | 72 | 64 | 68 | 72 | 0 |
| C5 | 82 | 82 | 86 | 76 | 80 | 81 | 86 | 84 | 80 | 88 | 74 | 82 | 1 |
| C7 | 78 | 84 | 72 | 74 | 76 | 77 | 62 | 68 | 68 | 64 | 68 | 66 | 11 |
| C9 | 92 | 96 | 94 | 90 | 92 | 93 | 86 | 90 | 90 | 90 | 88 | 89 | 4 |

## Table 5

Speech intelligibility scores of subjects in group 2 as judged by listeners L6–L10. ADM=absolute difference between means of Test 1 and 2.

| | Group 2 Intelligibility Scores (%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Test 1 | | | | | | Test 2 | | | | | | |
| Subject | L6 | L7 | L8 | L9 | L10 | Mean | L6 | L7 | L8 | L9 | L10 | Mean | ADM |
| CLP5 | 58 | 52 | 58 | 66 | 58 | 58 | 62 | 60 | 58 | 66 | 56 | 60 | 2 |
| CLP7 | 14 | 8 | 16 | 20 | 16 | 15 | 14 | 20 | 16 | 10 | 10 | 14 | 1 |
| CLP10 | 54 | 60 | 48 | 60 | 66 | 58 | 66 | 68 | 54 | 60 | 54 | 60 | 2 |
| CLP12 | 80 | 82 | 76 | 82 | 72 | 78 | 82 | 82 | 80 | 80 | 82 | 81 | 3 |
| CLP19 | 82 | 88 | 84 | 88 | 86 | 86 | 88 | 90 | 90 | 80 | 80 | 86 | 0 |
| CLP22 | 80 | 92 | 82 | 84 | 76 | 83 | 86 | 76 | 80 | 78 | 78 | 80 | 3 |
| C1 | 56 | 64 | 68 | 60 | 62 | 62 | 66 | 50 | 66 | 52 | 46 | 56 | 6 |
| C6 | 56 | 66 | 58 | 64 | 56 | 60 | 46 | 59 | 52 | 54 | 48 | 52 | 8 |
| C12 | 92 | 94 | 92 | 92 | 90 | 92 | 96 | 98 | 98 | 94 | 96 | 96 | 4 |
| C16 | 94 | 98 | 94 | 90 | 88 | 93 | 96 | 92 | 100 | 96 | 96 | 96 | 3 |

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## Table 6

Speech intelligibility scores of subjects in group 3 as judged by listeners L11–L15. ADM=absolute difference between means of Test 1 and 2.

| | Group 3 Intelligibility Scores (%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Test 1 | | | | | | Test 2 | | | | | | |
| Subject | L11 | L12 | L13 | L14 | L15 | Mean | L11 | L12 | L13 | L14 | L15 | Mean | ADM |
| CLP4 | 48 | 66 | 74 | 58 | 60 | 61 | 60 | 64 | 65 | 69 | 67 | 65 | 4 |
| CLP17 | 70 | 72 | 82 | 70 | 82 | 75 | 76 | 73 | 74 | 69 | 76 | 74 | 1 |
| CLP18 | 64 | 57 | 71 | 60 | 68 | 64 | 60 | 64 | 82 | 56 | 58 | 64 | 0 |
| CLP20 | 90 | 86 | 94 | 86 | 88 | 89 | 82 | 90 | 94 | 92 | 94 | 90 | 1 |
| CLP21 | 84 | 90 | 88 | 90 | 82 | 87 | 82 | 80 | 92 | 86 | 90 | 86 | 1 |
| C4 | 73 | 69 | 82 | 82 | 78 | 77 | 82 | 72 | 80 | 82 | 88 | 81 | 4 |
| C11 | 94 | 92 | 94 | 86 | 90 | 91 | 86 | 88 | 94 | 82 | 96 | 89 | 3 |
| C13 | 76 | 88 | 94 | 86 | 86 | 86 | 86 | 90 | 88 | 86 | 90 | 88 | 2 |
| C15 | 75 | 88 | 88 | 79 | 83 | 83 | 76 | 88 | 86 | 80 | 76 | 81 | 2 |

NIH-PA Author Manuscript
NIH-PA Author Manuscript



## Table 7

Speech intelligibility scores of subjects in group 4 as judged by listeners L16–L20. ADM=absolute difference between means of Test 1 and 2.

| | Group 4 Intelligibility Scores (%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Test 1 | | | | | | Test 2 | | | | | | |
| **Subject** | **L16** | **L17** | **L18** | **L19** | **L20** | **Mean** | **L16** | **L17** | **L18** | **L19** | **L20** | **Mean** | **ADM** |
| CLP1 | 54 | 68 | 64 | 66 | 66 | 64 | 64 | 64 | 62 | 64 | 60 | 63 | 1 |
| CLP6 | 84 | 74 | 80 | 86 | 86 | 82 | 86 | 88 | 82 | 90 | 86 | 86 | 4 |
| CLP8 | 53 | 56 | 60 | 60 | 62 | 58 | 40 | 40 | 42 | 42 | 38 | 40 | 18 |
| CLP9 | 80 | 82 | 80 | 82 | 78 | 80 | 80 | 66 | 70 | 82 | 80 | 76 | 4 |
| CLP16 | 94 | 90 | 94 | 96 | 90 | 93 | 92 | 90 | 76 | 88 | 90 | 87 | 6 |
| C2 | 64 | 54 | 52 | 60 | 60 | 58 | 70 | 56 | 62 | 58 | 64 | 62 | 4 |
| C8 | 92 | 92 | 88 | 88 | 88 | 90 | 90 | 92 | 88 | 94 | 94 | 92 | 2 |
| C10 | 92 | 94 | 92 | 90 | 94 | 92 | 90 | 92 | 94 | 96 | 92 | 93 | 1 |
| C14 | 90 | 94 | 88 | 88 | 86 | 89 | 94 | 94 | 82 | 90 | 98 | 92 | 3 |

**Table 8**

Means and standard deviations (SD) of intelligibility scores for subjects in group 2 based upon 5 listeners and 10 combinations of 3 listeners.

| | **Intelligibility Scores (%)** | |
|---|---|---|
| **Listeners** | **Mean** | **SD** |
| All 5 | 68 | 24 |
| 1_2_3 | 68 | 25 |
| 1_2_4 | 69 | 24 |
| 1_2_5 | 68 | 24 |
| 1_3_4 | 68 | 23 |
| 1_3_5 | 67 | 23 |
| 2_ 3_4 | 70 | 24 |
| 2_4_5 | 69 | 23 |
| 3_4_5 | 68 | 22 |
| 4_5_1 | 68 | 22 |
| 5_2_3 | 68 | 24 |

## Appendix

Word Sets of the Intelligibility Test: Sets were determined by initial consonant and/or consonant cluster. Some sets also target **low** or **high** vowels.

| **Set 1** | **Set 2** | **Set 3** | **Set 4** | **Set 5** | **Set 6** | **Set 7** | **Set 8** | **Set 9** | **Set 10** |
|---|---|---|---|---|---|---|---|---|---|
| **/p/ low** | **/gr/** | **/sp/** | **/t/ low** | **/p/ high** | **/tr/** | **/ʃ/ low** | **/kr/** | **/k/ low** | **/gl/** |
| pack | grab | spat | tack | peace | trace | shall | crab | cab | glad |
| pad | grace | speak | tag | peach | track | share | crack | call | glare |
| pail | grade | spell | tail | peep | trade | shark | crash | can | glass |
| pan | grape | spit | tall | pick | train | shed | cream | cap | glaze |
| pass | grass | spoil | tan | pig | trap | shell | creed | cat | gleam |
| pat | grease | spook | tap | pill | trash | shock | creep | catch | glide |
| pet | great | spoon | taught | pit | trays | shop | cries | cause | glob |
| pot | green | sport | tock | pitch | treat | shore | cross | cob | globe |
| | grim | spot | top | pull | trees | shot | crumb | cop | gloom |
| | group | spout | toss | push | trick | | crunch | cot | gloss |
| | growl | spud | | put | trip | | crust | cough | glove |
| | grows | | | | | | | | glued |
| Set 11 | Set 12 | Set 13 | Set 14 | Set 15 | Set 16 | Set 17 | Set 18 | Set 19 | Set 20 |
| /r/ high | /d/ low | /j/ | /k/ high | /s/ high | /t/ high | /fr/ | /fl/ | /ʧ/ low | /w/ low |
| reach | dad | year | cook | seat | teach | fraud | flag | chair | walk |
| read | dam | yell | cool | seed | team | freak | flat | chalk | wall |
| real | dawn | yep | could | seem | teen | freeze | fleece | chance | want |
| rich | den | yes | keep | seen | teeth | fresh | flies | check | was |
| rip | dock | yet | keys | sees | tick | fret | flip | chess | watch |
| roof | dog | yoke | kick | sick | till | frog | float | chop | web |
| room | doll | young | kid | sill | tin | fruit | floor | chore | well |
| root | dot | your | kill | sin | tip | | flows | | went |
| | | yuck | kin | sing | took | | flush | | wet |
| | | yum | kiss | sip | tool | | flute | | whale |
| | | | kit | sit | toot | | | | what |
| | | | | soon | tube | | | | when |

| **Set 1** | **Set 2** | **Set 3** | **Set 4** | **Set 5** | **Set 6** | **Set 7** | **Set 8** | **Set 9** | **Set 10** |
|---|---|---|---|---|---|---|---|---|---|
| **/p/ low** | **/gr/** | **/sp/** | **/t/ low** | **/p/ high** | **/tr/** | **/ʃ/ low** | **/kr/** | **/k/ low** | **/gl/** |
| | | | | soup | | | | | where |
| | | | | suit | | | | | wore |
| Set 21 | Set 22 | Set 23 | Set 24 | Set 25 | Set 26 | Set 27 | Set 28 | Set 29 | Set 30 |
| /sk/ low | /ʃ/ high | /f/ high | /st/ | /b/ high | /d/ high | /r/ low | /sm/ | /br/ | /θ/ |
| scab | sheep | fear | stair | beach | deal | rack | smack | braid | thick |
| scar | sheet | feed | state | bead | deed | ran | small | brain | thin |
| scare | ship | feel | steal | beak | deep | rap | smash | brand | thing |
| scat | shoes | feet | stem | bean | deer | rat | smear | brave | third |
| score | shook | fill | step | beat | did | red | smell | bread | thought |
| scum | shoot | fin | stick | beef | dig | rob | smirk | break | thumb |
| skull | should | fish | still | been | dim | rock | smog | breathe | |
| | | food | stool | beep | dip | rod | smoke | brick | |
| | | fool | stoop | bib | | wreck | smudge | bridge | |
| | | foot | stop | big | | wrong | smug | bright | |
| | | full | stove | bit | | | smut | bring | |
| | | | | bull | | | | broke | |
| | | | | bush | | | | | |
| Set 31 | Set 32 | Set 33 | Set 34 | Set 35 | Set 36 | Set 37 | Set 38 | Set 39 | Set 40 |
| /bl/ | /l/ high | /h/ low | /pr/ | /f/ low | /w/ high | /sn/ | /kl/ | /l/ low | /tw/ |
| black | lead | had | praise | fail | week | snack | clam | lap | tweed |
| blast | leak | hair | press | fall | whip | snake | clap | laugh | tweet |
| bleed | lean | hall | price | fan | whiz | snap | class | leg | twice |
| bless | leap | ham | prick | far | wig | sneak | clean | less | twig |
| blind | leave | has | pride | fat | will | snip | clear | let | twin |
| blink | lick | hat | prime | fed | win | snitch | click | lob | twine |
| block | lid | have | probe | fell | wish | snob | climb | lock | twirl |
| blood | lip | head | prong | for | wit | snoop | clip | log | twist |
| blur | lit | hen | | | witch | snooze | clock | lot | twitch |

| **Set 1** | **Set 2** | **Set 3** | **Set 4** | **Set 5** | **Set 6** | **Set 7** | **Set 8** | **Set 9** | **Set 10** |
|---|---|---|---|---|---|---|---|---|---|
| **/p/ low** | **/gr/** | **/sp/** | **/t/ low** | **/p/ high** | **/tr/** | **/ʃ/ low** | **/kr/** | **/k/ low** | **/gl/** |
| | look | hog | | | with | snore | close | | |
| | loop | hop | | | wood | snort | cloud | | |
| | loose | hot | | | wool | snout | clown | | |
| | lose | | | | | snub | club | | |
| | | | | | | snug | | | |
| Set 41 | Set 42 | Set 43 | Set 44 | Set 45 | Set 46 | Set 47 | Set 48 | Set 49 | Set 50 |
| /g/ | /s/ low | /pl/ | /ʧ/ high | /sl/ | /sk/ high | /h/ high | /ʤ/ | /b/ low | /dr/ |
| gas | sack | place | cheat | slap | scoop | heal | gel | back | drain |
| gate | sad | plane | cheek | sleep | scoot | heat | jack | bad | drank |
| gave | said | plant | cheap | sleeve | skid | here | jail | badge | dread |
| goat | sat | plate | cheer | slurp | skin | hick | jam | ball | dream |
| god | sauce | please | cheese | slick | skip | hill | jar | bam | dress |
| goes | sawed | plop | chick | slide | skirt | him | jazz | ban | dried |
| good | sock | plug | chill | slime | skit | hip | jet | bang | drink |
| goose | | plus | chin | slip | | his | job | bar | drip |
| got | | plays | chip | slot | | hit | join | bat | drive |
| guess | | | choose | sleigh | | | joke | beg | drop |
| gum | | | | | | | jug | bet | drowse |
| | | | | | | | juice | boss | drug |
| | | | | | | | | | drum |