

Ratings and Measures of Cleft Palate Speech

JOANNE D. SUBTELNY, Ph.D.
ROLLAND J. VAN HATTUM, Ph.D.
BRENDA B. MYERS, M.A.

Rochester, N.Y., 14603

Rating procedures are commonly employed in the diagnosis of communication efficiency of speakers with cleft palates to make assessments of speech intelligibility and articulation. Justification for the continued use of these convenient procedures should depend upon the reliability and validity of the ratings obtained.

Considerable evidence to support the use of ratings of articulation is available. Good intra- and inter-judgment agreement for ratings of articulation in the speech of individuals with cleft palates has been reported (6, 7). Positive and significant relationships between ratings of overall articulatory performance and number of errors as identified by conventional articulation testing have also been reported (4, 6, 9).

In general, less evidence is available to support the continued use of overall ratings of intelligibility. In simple terms, measures of intelligibility may be obtained by comparing the words comprehended by listeners with words intended by the speaker. Such measures are practical references for describing competence in communication and for expressing the communicative significance of disordered speech. Unfortunately, however, the procedures necessary for this type of measurement are extremely costly in time since they require carefully controlled tape recordings, playback to multiple listeners and rather tedious procedures in analysis. This latter factor is largely responsible for the common practice of rating overall intelligibility rather than measuring intelligibility as previously described.

This study was undertaken: 1) to evaluate differences between clinical ratings of overall intelligibility and measures of intelligibility, and 2) to describe a clinical format for the organization of multiple speech ratings and measures. Intelligibility and articulation are known to be significantly related; however, the two attributes are not the same. For this

Dr. Joanne D. Subtelny and Mrs. Brenda B. Myers are affiliated with the Department of Orthodontics, Eastman Dental Center, Rochester, New York. Dr. Van Hattum is Professor and Chairman, Speech Pathology and Audiology, State University College at Buffalo, New York.

This study was supported by PHS Research Grant DE-01837, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland.

reason, the present study involved: 1) overall ratings of articulation; 2) overall ratings of intelligibility; 3) evaluation of articulation using conventional testing procedures; and, 4) evaluation of intelligibility using testing procedures. The results were then subjected to statistical analyses to evaluate the reliability and validity of respective ratings.

Procedure

Subjects-Speech Materials. The study sample included 52 cleft palate patients ranging in age from six to 43 years. Identical speech material was recorded twice (before and after pharyngeal flap surgery) for each subject, thus yielding a total of 104 recordings.

All recordings were made in a sound treated room with constant mouth to microphone distance and high fidelity recording equipment. To obtain articulatory evaluation, words from the Templin-Darley Articulation Test were recorded at five seconds intervals and phonetically transcribed. Transcriptions were then analyzed to determine articulation error for 91 consonants appearing as single items. Customary procedures were employed in analyzing errors; that is, sounds which were omitted, substituted or distorted were counted as errors. The total number of errors, expressed in percentage, was used as the articulation measure.

The Lehiste-Peterson 50 Item Word Lists (2) and a picture intelligibility test were recorded for intelligibility testing. The picture test, developed by Van Hattum (8) for use with pre-literates, consisted of 50 words phonemically balanced and matched with the other lists. Subjects were

ARTICULATION ANALYSES

Test _____

Total Number of Consonants Tested _____ Plosives _____

Fricatives _____ Nasals _____ Glides _____ Affricates _____

E R R O R * T Y P E **

Test	Total	Plosive	Fricative	Nasal	Glide	Affricate	Omission	Substitution	Distortion
	# %	# %	# %	# %	# %	# %	# %	# %	# %
1st	— —	— —	— —	— —	— —	— —	— —	— —	— —
2nd	— —	— —	— —	— —	— —	— —	— —	— —	— —
3rd	— —	— —	— —	— —	— —	— —	— —	— —	— —
4th	— —	— —	— —	— —	— —	— —	— —	— —	— —
5th	— —	— —	— —	— —	— —	— —	— —	— —	— —

* Percentage of error determined by dividing total number of errors by number of consonants tested; percentages for each classification similarly determined. i.e., $\frac{\text{plosive errors}}{\text{\# plosives tested}}$

** Percentage for error type determined by dividing total number of errors by number of errors per type. Glottal stops are counted substitutions; pharyngeal fricatives are counted distortions.

Glottal stops: absent _____ inconsistent _____ consistent _____ Date _____

Assessment of Quality: (1. mild; 2. moderate; 3. severe)

Hoarse _____ Breathily _____ Harsh-Strident _____

FIGURE 1.

first familiarized with the material. Each word was then presented at five second intervals for recording. Tapes were played back for auditing by ten adult listeners with normal hearing and analyzed by Shupe (5). Words were scored as unintelligible when written responses were not phonetically equivalent with the intended utterance. For purposes of this study, the intelligibility measure is defined as the total percentage of words accurately identified by listeners.

A continuous speech passage consisting of 20 short sentences was recorded for separate scaled ratings of articulation and intelligibility. The two judges (senior authors with extensive experience in the cleft palate area) independently audited the entire passage as many times as desired to rate each parameter. A total of 416 ratings were obtained. Respective ratings were recorded by circling the appropriate scale number and definition. The articulation scale was defined as follows:

- (1) Superior articulation.
- (2) Normal for age and sex.
- (3) Substandard—few errors, i.e., slighting of final consonants, blends and clusters—therapy not recommended.
- (4) Mild articulation problem—few errors—therapy recommended.
- (5) Moderate problem—consistent errors noticeable to layman—therapy required.
- (6) Severe problem—many articulatory errors.

Intelligibility was rated:

- (1) Superior.
- (2) Normal for age and sex.
- (3) Mild difficulty in understanding—repetition not required.
- (4) Moderate difficulty—repetition required infrequently.
- (5) Marked difficulty—repetition required frequently.
- (6) Unintelligible with repetition.

By both scales, ratings of 1 and 2 indicated normal or better than normal performance.

Findings

Judgment Agreement. The degree of agreement between judges on ratings of articulation and intelligibility is reported in Table 1. About half of the ratings for both parameters were identical, with over 90% falling within one rating increment. Allowing for borderline cases and the difference in scale gradations of the two judges (Table 2) a good consistency pattern is evident. The correlation coefficients between judges for the articulation and intelligibility ratings were .74 and .80 respectively.

The gradation tendencies of each judge based on regression analysis are shown in Table 2. Although there are differences in rating tendencies between judges, it is clear that these differences are relatively small. In overview, the scale gradations used are reasonable and distinguishable with a high degree of consistency.

SPEECH RATING PROFILE

Circle appropriate scale rating below and enter at left.

Name	Date							
1	2	3	4	5	<u>Intell.</u>	1. Normal for age and sex. 2. Mild difficulty in understanding-repetition not required. 3. Moderate difficulty-repetition required infrequently. 4. Marked difficulty-repetition required frequently. 5. Unintelligible, with repetition.		
1	2	3	4	5	<u>Artic.</u>	1. Normal for age and sex. 2. Substandard, few errors, i.e., slighting of final consonants, blends and clusters; no therapy recommended. 3. Mild problem-few errors-therapy recommended. 4. Moderate problem-consistent errors noticeable to layman-therapy required. 5. Severe problem-many errors.		
1	2	3	4	5	6	7	<u>Nasality</u>	1. Hyponasal-consistent absence of nasal resonance-"cold in the nose" quality; perceptible to layman. 2. Hyponasal-slight; perceptible to trained ear. 3. Normal quality. 4. Nasal-slight; perceptible to trained ear. 5. Nasal-moderate; perceptible to layman. 6. Hypernasal; slight reduction in intraoral pressure; slight nasal airflow; few consonants grossly distorted. 7. Hypernasal-excessive; marked reduction in intraoral pressure; prominent nasal airflow; most consonants grossly distorted.
1	2	3	4	5	6	7	<u>Nasal Emission</u>	1. None. 2. Slight-perceptible to trained ear. 3. Moderate-perceptible to layman on a few consonants. 4. Marked-perceptible on most pressure sounds. 5. Excessive-all pressure consonants grossly distorted.
1	2	3	4	5	6	7	<u>Grinace</u>	1. None. 2. Slight-restricted to nares. 3. Marked-restricted to nares. 4. Slight, not restricted to nares. 5. Marked, not restricted to nares.
1	2	3	4	5			<u>Rate</u>	1. Slow, labored, intelligibility reduced. 2. Slow, but monitored for clarity. 3. Normal. 4. Too rapid for articulatory skill, slight reduction of intelligibility. 5. Much too rapid for articulatory skill; moderate reduction of intelligibility.
1	2	3	4	5			<u>Intensity</u>	1. Inadequate intensity; marked reduction of intelligibility. 2. Weak, slight reduction of intelligibility. 3. Normal. 4. Slightly above normal intensity level. 5. Loud, unpleasant.
1	2	3	4	5			<u>Pitch</u>	1. Much below optimum pitch level. 2. Slightly below optimum. 3. Appropriate for age and sex. 4. Slightly above optimum. 5. Much above optimum.

NORMAL

FIGURE 2.

TABLE 1. Interjudge agreement on ratings of articulation and intelligibility.

<i>level of agreement</i>	<i>articulation</i>			<i>intelligibility</i>		
	<i>No.</i>	<i>%</i>	<i>Cum. %</i>	<i>No.</i>	<i>%</i>	<i>Cum. %</i>
perfect agreement.....	49	47	47	51	49	49
within one scale value.....	45	43	90	46	44	93
within two scale values.....	9	9	99	7	7	100
within three scale values...	1	1	100	0	0	100

Articulation. Tabulation of articulation ratings as noted in Table 3 revealed 12% of the subjects were rated 1 or 2. For this group, the mean articulation error was 3%. Nine percent of the subjects were rated 3 (mean 4% articulation error); 23% of the subjects were rated 4 (mean 15% error); 33% were rated 5 (mean 20% error); and 23% were rated 6 (mean 34% error). These figures indicate a reasonably good distribution in ratings was attained. Progressively higher or poorer ratings were associated with higher percentages of error.

The correlation coefficients between ratings and measures of articulation are reported in Table 4. The correlations, ranging from $r .65$ to $.71$, are significant beyond the .01 level of confidence. In all cases, the correlations between ratings and measures are very high, particularly when viewed in the context of correlating a continuous variable, the measurement, with a saw-toothed rating scale, making a 1.00 correlation coefficient impossible to obtain.

Judge A had a slightly higher correlation coefficient than Judge B with

TABLE 2. Scale gradation tendencies of Judge A and Judge B based on regression analysis.

	<i>measurement range</i>	
	<i>judge A</i>	<i>judge B</i>
<i>articulation rating</i>		
1	—	—
2	0.0-5.8	0.0-3.5
3	5.9-11.7	3.6-11.6
4	11.8-18.6	11.7-19.7
5	18.7-25.5	19.8-27.8
6	25.6-Over	27.9-Over
<i>intelligibility rating</i>		
1	89.5-100.0	85.9-100.0
2	78.7-89.4	74.7-85.8
3	67.9-78.6	63.5-74.6
4	57.1-67.8	52.3-63.4
5	46.3-57.0	41.1-52.2
6	0.0-46.2	0.0-41.0

TABLE 3. Judgment ratings and associated measures of articulation error and intelligibility. (n = total number of ratings for each category for two judges expressed in percentage)

<i>articulation error</i>				<i>intelligibility</i>			
<i>rating</i>	<i>n</i>	<i>mean</i>	<i>SD</i>	<i>rating</i>	<i>n</i>	<i>mean</i>	<i>SD</i>
1	1%	.55%	.61	1	1%	89.50%	6.36
2	11%	4.14%	3.61	2	24%	81.76%	10.30
3	9%	4.40%	3.66	3	30%	72.31%	11.32
4	23%	15.01%	7.56	4	21%	62.19%	18.01
5	33%	20.41%	10.52	5	20%	48.70%	18.66
6	23%	33.98%	14.82	6	4%	36.60%	19.02

both measures; however, neither difference was statistically significant. Particularly interesting is the fact that only marginal improvement over Judge A ratings alone could be obtained by adding the impact of Judge B ratings, either as a simple average or the optimum weighting of the two ratings obtained from a multiple linear regression model. Specifically, the improvements in the coefficients were only .04 and .01 for articulation and intelligibility, respectively.

Intelligibility. A tabulation of respective ratings of intelligibility (Table 3) showed: 25% of the subjects were rated 1 or 2. For this group, the mean intelligibility score was 84%. Thirty percent of the subjects were rated 3 (mean 72% intelligibility); 21% were rated 4 (mean 62% intelligibility); 20% were rated 5 (mean 49%); and 4% were rated 6 (mean 37%). Again, a reasonably good distribution in ratings was achieved. Progressively poorer ratings were associated with progressively lower intelligibility scores, defined by testing.

Correlations between ratings and measures of intelligibility for each judge independently, for averaged ratings, and for the optimum weighting of the two ratings are included in Table 4. All coefficients ranging from .63 to .70 are significant beyond the .01 level.

TABLE 4. Correlation coefficients between ratings and measures of speech articulation and intelligibility.

<i>rating</i>	<i>measurement</i>	
	<i>articulation</i>	<i>intelligibility</i>
judge A67†	— .69†
judge B65†	— .63†
averaged rating (A and B)71†	— .69†
optimum weighting of A and B*71†	— .70†

* Obtained by multiple linear regression.

† Significant at .01 level.

Discussion

Although a relatively good distribution in ratings for articulation and intelligibility was achieved, some basic differences are evident. Proportionately, a much higher percentage of speech samples (55%) were rated normal or mildly deviate in intelligibility than was indicated by analogous articulation ratings (21%). As a corollary, only 4% of speech samples were rated unintelligible whereas 34% were rated severely defective in articulation. It appears, therefore, that speech characterized by many articulation errors is not inevitably unintelligible. In overview, ratings tended to skew to the left or the good end of the intelligibility continuum, and skew to the right or poor end of the articulation scale. In this regard, present findings indicate a difference between parameters of articulation and intelligibility in cleft palate speech.

Since speech samples were recorded before and after pharyngeal flap surgery, material for judgment varied grossly in regard to nasal resonance. Such variation may partially explain or contribute to the differences observed. Nasalization alone may lower articulation performance proportionately more than intelligibility. Most clinicians have observed nasalized cleft palate speech which is quite intelligible.

Speech which is nasalized and distorted by perceptible nasal emission would rightfully be rated as defective in articulation. By measurement procedure, distortion is one specific type of articulation error. By intelligibility testing, distortions are not errors unless phoneme identity is lost. When this basic difference in criteria for error is recognized, some differences between articulation and intelligibility measures should be expected.

Recent articulation and intelligibility test data for a young aglossic patient (10) provides further evidence of a significant difference between parameters. The speaker is reported to have consistently distorted sibilants, affricates and velar stops on the Templin-Darley Test but intelligibility averaged 86%. In this case, the difference between parameters cannot be explained on the basis of nasalization or by the nature of speech disordered by palatopharyngeal incompetence. In sum, a difference which makes a difference clinically is indicated.

How well a patient makes himself understood has very broad implications. Relative adequacy in communication may exert a profound effect upon personal adjustment of the speaker and upon clinical decisions as related to therapeutic, educational and vocational planning. The number and type of articulation errors determined by testing provides essential information of a different type which does not replicate but rather complements and supplements the intelligibility assessment.

Critical articulation testing assists in identifying errors related to: delayed maturation, inadequate palatopharyngeal valving, deviate tongue function and oral malformations. Repeated testing on a longitudinal basis is also needed to evaluate the efficacy of therapy and/or treatment. In

borderline cases of palatopharyngeal incompetence, a period of intensive therapy is frequently recommended before decisions are made relative to surgical or prosthetic intervention. If standardized test material and conventional procedures are rigorously followed in giving the test and in analyzing the results, articulation test data as summarized in the speech profile can provide a very effective basis for arriving at clinical decisions.

The impact of deviate articulation and/or excessive nasalization as related to communication may best be expressed by intelligibility assessment. For this reason, such assessment can and should be part of clinical evaluation. Since intelligibility testing may be impractical in most settings, ratings rather than measures are justified.

The present scale for intelligibility rating has proved useful. The results of the analysis indicate it is well gradated. Correlations between the two judges were slightly higher for intelligibility than for articulation ($r .80$ vs $r .74$). Correlations between ratings and measures of intelligibility closely approximated analogous correlations between ratings and measures of articulation. Compositely, these findings permit the conclusion that intelligibility rating is just as effective as articulation rating in clinical speech situations.¹ In such situations, it is recommended that the rating of intelligibility be based upon conversational speech material which has the advantage of reducing anticipatory cues for the listener. With elementary school age youngsters, names of siblings or descriptions of events, are preferable to rote material.

The six point scales for rating of intelligibility and for rating articulation were adequate. In retrospect, however, both scales could effectively be reduced to five points simply by combining ratings of 1 and 2, which designated better than normal speech performance. Since only 1% of the speech samples were rated better than normal, the reduction in range of both scales was definitely indicated. The revised scales are incorporated as part of the speech profile currently used to catalog longitudinal cleft palate speech data.

One hundred and forty recorded samples of cleft palate speech were rated by the same two judges to evaluate the revised scales. Interjudgment agreement for both articulation and intelligibility ratings was found to be higher than when the six point scale was used. Sixty-five percent of the ratings of articulation were in perfect agreement with 92% falling within one scale value. Seventy-one percent of the intelligibility ratings were in perfect agreement with 96% falling within one scale value.

The seven point nasality scale, ranging from hypo- to hypernasality, included in the profile has been very useful. Interjudgment agreement for

¹The correlation between averaged articulation ratings and measures ($r .71$) is in basic agreement with similar data reported for speakers with functional disorders of articulation (1). Jordan's correlation between ratings (nine point scale) and measures (number of single defective sounds) was $r .78$. In contrast, Van Demark's (6) correlation coefficient between overall judgements of articulation defectiveness and total number of articulatory errors in cleft palate speech was $r .88$ or considerably higher than indicated in this analysis.

SPEECH PROFILE
PALATOPHARYNGEAL IMPEDENCE

Patient _____ Birthdate ____ / ____ / ____

Address _____

Father _____ Phone _____

Cleft Classification: _____

Lip:

Bilateral _____ Comp. _____ Incomp. _____

Unilateral _____ Comp. _____ Incomp. _____

Palate:

Bilateral _____ Comp. _____ Incomp. _____

Unilateral _____ Comp. _____ Incomp. _____

Posterior _____ Submucous _____ CPI _____

Post Adenoid. _____ Neurologic _____

Other _____

Date of Evaluation

1. _____ by _____

2. _____ by _____

3. _____ by _____

4. _____ by _____

5. _____ by _____

Status of Physical Management - Palate:

Primary Secondary Pharyn. Pharyn. Prosthesis Palatal

Unoperated () Repair () Repair () Flap () Implant () Phar. Bulb () Lift ()

Date: _____

Fistula: None () Sublabial () Anterior to Vault () Vault () Posterior ()

Estimate of Size _____ Obturated: Prosthesis: Yes _____ No _____ Date _____

Surgery: Yes _____ No _____ Date _____

Palatopharyngeal Gap During Phonation - Oropharyngeal Examination - Date _____

None () Questionable () Small () Moderate () Large ()

Cephalometric Examination: /u/ _____ mm /s/ _____ mm - Date _____

Audiometric Assessment Date: _____ Findings: _____

Management: Pediatrician/Physician _____ Surgeon _____ Otologist _____

Speech Pathologist _____ Orthodontist _____ Prosthodontist _____

FIGURE 3.

nasality ratings is slightly lower than for the articulation and intelligibility ratings; however, the correlation between judges for nasality rating ($r .73$) approximates the analogous correlations for articulation rating ($r .74$) and for intelligibility rating ($r .80$).

The rating profile was designed to provide a simple graphic report of repeated observations. Cleft palate speech is composed of varied attributes. Articulation defectiveness, intelligibility and nasality are known to be related, however, intercorrelations reported in the literature vary greatly. Reasons for this variation have been carefully identified and discussed by Moll (3). The pertinent fact is that the conflicting data does exist. For this reason, predictions of one parameter from measures derived for another are not justified. It appears most prudent to separately assess varied aspects of cleft palate speech, then synthesize information, as in the speech profile. The specific format used to organize pertinent speech observations is largely a matter of personal choice. The speech profile has been reproduced here simply because it has been very effective over the long term in our particular clinical setting.

Summary

Tape recorded samples of continuous speech produced by individuals with cleft palate were separately rated for articulatory defectiveness and intelligibility by two speech pathologists. Conventional testing procedures were also employed to measure both parameters, thus permitting statisti-

cal analysis to evaluate the reliability and validity of respective ratings. Correlation coefficients between the two judges were slightly higher for intelligibility rating than for articulation. Correlations between ratings and measure of intelligibility closely approximated analogous correlations between ratings and measures of articulation.

It is concluded: a) intelligibility rating is just as effective as articulation rating in clinical speech situations; b) articulation and intelligibility should be separately assessed simply because differences between parameters do exist; c) the differences observed are important and do not permit clinical predictions of speech intelligibility from articulation data or visa versa. It is recommended that varied aspects of cleft palate speech (nasality, articulation and intelligibility) be separately assessed with information then organized and synthesized as discussed to facilitate clinical planning.

Acknowledgments: The authors gratefully acknowledge the cooperation of the Center for Craniofacial Anomalies of the University of Illinois, the research assistance of Charles R. Elliott, Ph.D. and the statistical consultation and analysis rendered by James R. Mills.

reprints: Joanne D. Subtelny, Ph.D.
Eastman Dental Center
800 Main Street East
Rochester, New York 14603

References

1. JORDAN, E. P., Validation of articulation test measures. *J. Speech Hearing Res.*, 3, 303-320, 1960.
2. LEHISTE, I., and PETERSON, G. E., Linguistic considerations in the study of speech intelligibility. *J. Acous. Soc. Amer.*, 131, 250-286, 1959.
3. MOLL, K. L., Speech characteristics of individuals with cleft lip and palate. In *Cleft Palate and Communication*, D. C. Priestersbach and D. Sherman, Eds., New York, Academic Press, 1968.
4. MORRIS, H. L., Communication skills of children with cleft lips and palates. *J. Speech Hearing Res.*, 5, 79-90, 1962.
5. SHUPE, L. K., Speech intelligibility measures of cleft palate speakers before and after pharyngeal flap surgery. Ph.D. Dissertation, State University of New York at Buffalo, 1968.
6. VAN DEMARK, D. R., Misarticulations and listener judgements of the speech of individuals with cleft palates. *Cleft Palate J.*, 1, 232-245, 1964.
7. VAN DEMARK, D. R., and VAN DEMARK, A. A., Misarticulations of cleft palate palate children achieving velopharyngeal closure and children with functional speech problems. *Cleft Palate J.*, 4, 31-37, 1967.
8. VAN HATTUM, R. J., An intelligibility word list for nonreaders. Unpublished study, Eastman Dental Center, Rochester, New York, October, 1964.
9. VAN HATTUM, R. J., Articulation and nasality in cleft palate speakers. *J. Speech Hearing Res.*, 1, 383-387, 1958.
10. WEINBERG, B., and PARAS, N., Speech intelligibility of a seven-year-old girl with severe congenital hypoplasia of the tongue. *Cleft Palate J.*, 7, 436-442, 1970.