# Effects of Examiner Expectancy on Speech Ratings of Individuals with Cleft Lip and/or Palate

**LORRAINE A. RAMIG, Ph.D.**
*Boulder, Colorado*

The label 'cleft palate,' any facial disfigurement associated with cleft lip and/or palate and the description 'hypernasal' were studied relative to their effects on ratings of speech. Sixty-four listeners rated hypernasality and articulation characteristics of ten males with cleft lip and/or palate. Results indicated that the speech ratings assigned to this group of individuals did not change significantly with this additional information. Severity of speech defectiveness, mode of stimulus presentation and listener reliability were identified as important variables to be considered in the interpretation of these data.

The influence of an examiner on the validity of behavioral evaluations has been defined as examiner effect (Rosenthal, 1966). One type of effect identified by Rosenthal is examiner expectancy. This occurs when an examiner's assessment of behavior is detectably biased in the direction of his expectancies.

The effect of examiner expectancy on the validity of behavioral evaluations has been well documented (Rosenthal, 1966; Schwartz and Flanigan, 1971; Clifford, 1973). Facial disfigurement, diagnostic labels and case history information have been identified as factors which may bias an examiner's assessment of behavior (Wright, 1960; Monohan, 1971; Hersh, 1971; Clifford, 1973; Machowsky, 1973). Although these factors are usually present during evaluations of speech, only a few investigators, such as Beasley and Harlton, 1973; Meitus et al., 1973; and Podol and Salvia, 1976, have studied their effect on the validity of speech evaluations. Furthermore, the results of these studies are in conflict. Beasley and Harlton (1973) and Meitus et al. (1973) concluded that speech evaluations are not influenced by biasing information. On the other hand, Podol and Salvia (1976) reported that biasing information selectively effected evaluations of patients with moderate speech problems.

Because potentially biasing information is usually present during speech evaluations of patients with cleft lip and/or palate, these evaluations are particularly relevant to the study of examiner expectancy. Typically, patients with cleft lip and/or palate have a comprehensive case history available prior to evaluation. This provides the speech examiner with the 'cleft palate' label as well as the results of previous evaluations of articulation and resonance characteristics. In addition, the facial disfigurement (such as lip scarring, facial disproportion, nasal grimaces and irregular dentition) frequently presented by these clients is visible to the examiner throughout the speech evaluation. Since such factors have been shown to bias many other forms of behavioral assessment, it is likely that they may also effect speech assessment. Because the results of speech evaluations of individuals with cleft lip and/or palate may influence recommendations for surgical treatment of the velopharyngeal mechanism, the validity of these results is particularly critical.

This study was designed to investigate the effects of the label 'cleft palate,' any facial disfigurement associated with cleft lip and/or palate, and the description 'hypernasal' on speech ratings.

Address editorial correspondence to: Lorraine A. Ramig, Ph.D., Assistant Professor, Department of Communication Disorders and Speech Science, Campus Box 409, University of Colorado, Boulder, Colorado 80309.

## Method

### SUBJECTS

*Listeners.* Sixty-four listeners were sampled from groups of undergraduate and graduate students and faculty in speech pathology. They represented five university populations in Wisconsin and Illinois. Within the context of a general information survey, listeners provided information about the extent of their academic study and clinical experience in the area of cleft lip and palate. This information was used to balance the experimental groups according to listener experience (Hays, 1973).

*Speakers.* Ten males, with cleft lip and/or palate, ranging in age from 8 to 12 years, provided the speech samples for this study. They were selected from children receiving clinical services at cleft palate clinics in Wisconsin.

Rank order data on the severity of hypernasality of these speakers are presented in Table 1. Twenty additional listeners, with clinical and academic experience comparable to the sixty-four listeners described above, participated in a paired comparison procedure (Guilford, 1954) from which these independently collected data were obtained. These data were necessary in order to interpret any relationship between severity of hypernasality and biasing factors.

### PROCEDURES

High quality video recordings were obtained while each speaker read a phonetically balanced passage ("Arthur the Young Rat") (Edmonston, 1963). From each speaker's recording, the identical 30 second section was extracted to form a ten-sample stimulus tape. Two randomly ordered versions of this ten-sample tape were prepared.

The sixty-four listeners were divided into four equal size groups. Each listener individually rated one of the two ordered versions of the tape. *Group I* listeners evaluated only the audio track of the video recording, no other information was provided. *Group II* listeners also evaluated the audio track, however, they were informed that they were rating speakers with 'cleft palate.' Listeners in *Group III* were presented both the audio and visual tracks (video) of the same recordings and were informed that they were rating 'cleft palate' speakers. Listeners in *Group IV* also evaluated both the audio and visual tracks (video) and were informed that they were rating 'cleft palate' speakers who had been described as 'hypernasal' by a cleft palate team. The stimuli were presented free field at a constant, comfortable intensity level and a constant listener-to-monitor distance (4 feet). All listeners rated severity of hypernasality and severity of articulation defectiveness for each of the ten speech samples.

The following instructions, designed so that the specific objectives of the experiment were not revealed, were read to each listener:

> We are currently developing training tapes for student clinicians. For this project we need to accumulate many ratings of speech samples. Your task will be to rate 10 speech samples today. The following tape contains 10 thirty second speech samples. Your job is to score each sample on the following speech parameters: severity of hypernasality and defectiveness of articulation. Score each of these parameters individually on a scale of 0 to 7; 0 not being clinically defective, 1 being slightly clinically defective, through 7, being extremely clinically defective. Circle your answers. Please direct your attention to the entire speech sample before marking your score sheet. The first sample you will hear is a warm-up to get you accustomed to the task.

## Results

### RELIABILITY

Intrajudge reliability for ratings of hypernasality and articulation was assessed by calculating Pearson product moment correlation coefficients (Bruning and Kintz, 1968) between original and repeated samplings of ratings assigned each speaker. Coefficients were based on a sample representing approximately one-third of the original listener sample ($n = 24$; 6 listeners from each of the 4 groups). These coefficients ranged from $r = .55$ to .99 for hypernasality ratings and $r = .60$ to .98 for articulation ratings. All coefficients were statistically significant ($p < .05$). Intrajudge reliability was considered acceptable for the purpose of this investigation.

Interjudge reliability was assessed by calculating Kendall's Coefficient of Concordance (W) for the initial hypernasality and articulation ratings assigned each speaker by

TABLE 1. Hypernasality Rank Order Data and Group Hypernasality Ratings Assigned Ten Speakers with Cleft Lip and/or Palate

| *Speaker Number* | *Hypernasality Rank** | *Mean Ratings of Hypernasality* | | | |
|---|---|---|---|---|---|
| | | Group I | Group II | Group III | Group IV |
| 1 | 4 | 4.00 | 4.63 | 4.56 | 4.31 |
| 2 | 2 | 4.69 | 4.50 | 4.50 | 3.81 |
| 3 | 7 | 1.88 | 2.75 | 1.94 | 3.56 |
| 4 | 1 | 6.38 | 6.38 | 6.25 | 6.13 |
| 5 | 9 | 0.06 | 0.75 | 0.25 | 0.75 |
| 6 | 3 | 3.00 | 4.69 | 3.44 | 4.13 |
| 7 | 5 | 1.25 | 3.13 | 2.00 | 2.94 |
| 8 | 6 | 1.06 | 2.31 | 0.51 | 1.38 |
| 9 | 10 | 0.25 | 0.38 | 0.43 | 0.25 |
| 10 | 8 | 1.75 | 2.31 | 2.63 | 2.88 |

* Rank of "1" indicates the most severe hypernasality, as determined from an independently derived rank ordering.

listeners in all groups. For individual speakers, these coefficients ranged from W = .27 to .85 for hypernasality ratings and from W = .06 to .88 for articulation ratings. Coefficients reflecting between listener agreement were statistically significant ($p < .05$) for less than half of the speakers. When interjudge reliability was assessed for group data, across all speakers, coefficients ranged from W = .65 to .73 for hypernasality ratings and W = .57 to .66 for articulation ratings. All coefficients calculated on group data were statistically significant ($p < .001$). Therefore, it was concluded that for individual speaker data, between listener agreement was poor. Only when group mean data were analyzed was between listener agreement acceptable. The practical significance, however, of calculating reliability coefficients on group mean data must be questioned.

## Group Effects

*Hypernasality.* Listeners' ratings of severity of hypernasality were analyzed by a three factor analysis of variance (Group (4) × Order (2) × Speaker (10)) with repeated measures on the third factor (speaker) (Winer, 1971). Results of this analysis indicated that the factor of speaker was significant ($p < .05$). Speaker × Group ($p < .05$) and Speaker × Order ($p < .05$) interactions also were significant.

Therefore, a two factor (Group × Order) analysis of variance (Winer, 1971) was carried out on the hypernasality ratings assigned each speaker. Results indicated that the factor of group was significant in the analyses of three of the ten speakers (speakers 3, 7 and 8). Inspection of the hypernasality rank order data (Table 1) indicated that these three speakers were ranked in the mid-ranks of severity within this sample of speakers.

Post hoc testing of the group mean data for these three speakers was performed using Tukey's HSD (Kirk, 1968). Results indicated that there were significant differences between hypernasality ratings assigned by listeners in Group I (auditory presentation only) and Group II (auditory presentation and 'cleft palate' label) as well as between Group III (auditory and visual (video) presentation and 'cleft palate' label) and Group IV (auditory and visual (video) presentation, 'cleft palate' label and description 'hypernasal'). Hypernasality was rated significantly more severe by listeners who were given more information about these three speakers. However, hypernasality was rated significantly less severe by listeners in Group III (auditory and visual (video) presentation and 'cleft palate' label) than by listeners in Group II (auditory presentation and 'cleft palate' label).

*Articulation.* Results of a three factor analysis of variance (Group (4 × Order (2) × Speaker (10)) with repeated measures on the third factor (speaker) (Winer, 1971) indicated that the factors of order and speaker were significant for listener ratings of articulation defectiveness. Listeners rated articulation defectiveness significantly more severe when speakers were presented in Order 2.

To probe the significant speaker effect, a two factor (Group × Order) analysis of variance (Winer, 1971) was completed on articu-

lation ratings assigned each speaker. Results indicated that the factor of group was significant in analysis of ratings assigned to two speakers (speakers number 2 and 4; Table 1). Post hoc testing of the group mean data for these speakers indicated that ratings of articulation defectiveness assigned by listeners in Group IV (auditory and visual (video) presentation, 'cleft palate' label and description 'hypernasal') were significantly greater than those assigned by listeners in Group III (auditory and visual (video) presentation and 'cleft palate' label). But, ratings assigned by listeners in Group III were significantly lower than those assigned by listeners in Group II (auditory presentation and 'cleft palate' label).

**Discussion**

The results of this study indicate that the label 'cleft palate,' any facial disfigurement associated with cleft lip and/or palate and the description 'hypernasal' did not significantly affect speech ratings assigned to this group of speakers. Further inspection of the data may offer explanations for these findings.

Podol and Salvia (1976) reported that speech adequacy is an important consideration in the interpretation of expectancy data. In their study, expectancy effects were measured only for ratings assigned to speakers exhibiting mild to moderate nasality. The results of the present study are consistent with those findings. Within this sample of ten speakers, only the hypernasality ratings assigned the three speakers falling in the mid-ranks of hypernasality were effected by the factors 'cleft palate' label and 'hypernasality' description. The hypernasality of these three speakers was rated significantly more severe in the presence of these factors. It could be hypothesized that examiner expectancy is more likely to influence hypernasality ratings in cases with mild to moderate hypernasality than with severe hypernasality. Future research on examiner expectancy should assess this possibility.

The mode of stimulus presentation may be another important issue in the interpretation of these data. A major assumption underlying the design of this study was that only one factor would differ between each group condition. For example, listeners in Groups I and II received auditory presentation of the speech samples, and the 'cleft palate' label was the only factor added to Group II's information. Similarly, listeners in Groups III and IV received auditory and visual (video) presentation of the speech samples, and the description 'hypernasal' was the only additional factor added to Group IV's information. Under these circumstances, it was assumed that the relationship between an added factor and any rating change could be interpreted. In both of these comparisons (Group I to Group II and Group III to Group IV) mean severity ratings increased with the addition of a biasing factor.

However, in the design of this study, consideration was not given to the fact that when the factor 'facial disfigurement' was added to the information Group III listeners received, the mode of stimulus presentation was also changed. Listeners in Group II received auditory presentation, while listeners in Group III received auditory and visual (video) presentation and the additional factor 'facial disfigurement.' Thus, interpretation of the independent effect of the factor 'facial disfigurement' is difficult. In fact, in contrast to the other between group comparisons, severity ratings decreased between Groups II and III with the addition of the biasing factor 'facial disfigurement.' It appears then that mode of presentation must be considered for its potential confounding effect on these ratings. Furthermore, the assumption that auditory and auditory and visual (video) modes of presentation would produce comparable ratings must be questioned and considered in the design of future research.

A final consideration in the interpretation of these data is the reliability of the ratings. It has been reported that the reliability of hypernasality ratings is questionable (Bradford, Brooks and Shelton, 1964; Counihan and Cullinan, 1970). Only when reliability is expressed for group ratings and calculated from mean scale values has it been considered adequate (Lintz and Sherman, 1961; Counihan and Cullinan, 1970). The reliability data in this study are no exception. The possible effect of large within-group listener variance on the lack of between-group differences must be considered in the interpretation of these results.

## Conclusion

The results of this study indicate that the factors of 'cleft palate' label, any facial disfigurement associated with cleft lip and/or palate and the 'hypernasality' description did not significantly effect speech ratings assigned to a group of individuals with cleft lip and/or palate. Severity of speech defectiveness, mode of stimulus presentation and listener reliability were identified as important variables to be considered both in the interpretation of these data as well as in the design of future research.

Although this study did not demonstrate that examiner expectancy influenced speech ratings, the assumption that speech evaluations are therefore free from expectancy effects is premature. The important methodological issues raised in this study emphasize the need for future research before any conclusions are reached. The fact that examiner expectancy has been well established in other forms of behavioral assessment suggests that it must continue to be a consideration in the evaluation of speech behavior.

## References


BEASLEY, D., and HARLTON, K. Biasing effects of visual cues upon language evaluations, *J. Mich Speech Hear. Assoc., 9:* 36–43, 1973.

BRADFORD, L., BROOKS, A., and SHELTON, R. Clinical judgement of hypernasality in cleft palate children, *Cleft Palate J., 1:* 329–335, 1964.

BRUNING, J., and KINTZ, B. Computational Handbook of Statistics, Glenview, Illinois: Scott Foresman, 1968.

CLIFFORD, E. Psychosocial aspects of orofacial anomalies: speculations in search of data, *A.S.H.A. Report*, No. 8, 1973.

COUNIHAN, D., and CULLINAN, W. Reliability and dispersion of nasality ratings *Cleft Palate J., 7:* 261, 1970.

EDMONSTON, W. *Laradon Articulation Scale*, Los Angeles, California: Western Psychological Services, 1963.

GUILFORD, J. *Psychometric Methods*, New York: McGraw Hill, 1954.

HAYS, W. Statistics for the Social Sciences, New York: Holt, Rinehart and Winston, 1973.

HERSH, J. Effects of referral information on testers, *J. Consult Clin Psych., 37:* 116–122, 1971.

KIRK, R. *Experimental design: procedures for the behavioral sciences*, Belmont, California: Wadsworth Publishing, 1968.

LINTZ, L. and SHERMAN, D. Phonetic elements and perception of nasality, *J. Speech Hear. Res, 4:* 381–396, 1961.

MACHOWSKY, H. The effects of teachers' locus of control, teachers' knowledge of intellectual potential, and informational source status on teachers' judgements of childrens' expected academic ability, unpublished doctoral dissertation, Temple University, 1973.

MEITUS, I., RINGEL, R., HOUSE, A., and HOTCHKISS, J. Clinician bias in evaluating speech proficiency, *Br J. Comm. Dis., 8:* 146–151, 1973.

MONOHAN, L. Norms for physical disability: an analysis of expectations, unpublished doctoral dissertation, University of Oregon, 1971.

PODOL, J. and SALVIA, J. Effects of visibility of a prepalatal cleft on the evaluation of speech. *Cleft Palate J., 13:* 361–366, 1976.

ROSENTHAL, R. *Experimenter Effects in Behavioral Research*, New York: Appleton-Century-Crofts, 1966.

SCHWARTZ, R., and FLANIGAN, D. Evaluation of examiner bias in intelligence testing, *Am. J. Ment. Def., 76:* 262–265, 1971.

WINER, B. J., *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

WRIGHT, B. *Physical Disability—a psychological approach*, New York: Harper and Row, 1960.