AN INTELLIGIBILITY ASSESSMENT OF TODDLERS WITH CLEFT LIP AND PALATE WHO RECEIVED AND DID NOT RECEIVE PRESURGICAL INFANT ORTHOPEDIC TREATMENT

EMMY M. KONST, HANNY WEERSINK-BRAKS,
TONI RIETVELD, and HERMAN PETERS
Department of Orthodontics and Oral Biology, University of Nijmegen, Nijmegen, The Netherlands

A randomized, prospective, clinical study was performed investigating the effects of presurgical infant orthopedic treatment (PIO) in children with unilateral cleft lip and palate (UCLP). The influence of PIO on speech intelligibility was evaluated with two groups, each consisting of 10 children with UCLP. One group used PIO during the first year of life, whereas the other group did not use the device. Eight children without cleft served as a second control group. Intelligibility was assessed by lay listeners using two methods: transcription and listener rating. The ratings proved to be reliable and to have sufficient validity, but they did not completely reflect intelligibility defined as the proportion of words understood by the listener. Children in the treatment group were rated as exhibiting greater intelligibility than those in the nontreatment group. However, data obtained by means of transcriptions indicated that, in fact, there were no group differences in actual intelligibility. Only in comparison with their noncleft peers were the children with cleft lip and palate significantly less well understood.   © 2000 by Elsevier Science Inc.

Educational Objectives: Readers will learn about (1) the reliability and validity of a rating paradigm used for assessing intelligibility of children with UCLP and (2) the relationship between the use of PIO and intelligibility in toddlers with UCLP.

KEY WORDS: Intelligibility; Cleft lip and palate; Presurgical infant orthopedics

INTRODUCTION

Children born with cleft lip and palate are likely to have associated speech problems to some extent. Frequently mentioned potential problems associated with

cleft palate are: nasalization of vowels, audible nasal emission, frequent use of glottal stops and pharyngeal fricatives, palatization and retraction of consonants, and weakened fricatives, plosives, and affricates (Stengelhofen, 1989). These characteristics affect speech intelligibility and, consequently, the efficacy of oral communication.

Worldwide, cleft lip and palate patients are treated with presurgical infant orthopedics (PIO). One form of PIO treatment is the Zurich approach (Hotz, 1979, 1983), in which a maxillary appliance is fitted as soon as possible after birth. The appliance is adjusted periodically and is worn until the soft palate is closed surgically. It is intended to guide the growth and position of the maxillary segments and to improve feeding and tongue posture. Proponents of PIO treatment claim that the appliance leads to better speech (Gnoinski, 1990; Gruber, 1990; Hotz, Gnoinski, Perko, Nussbaumer, Hof, & Haubensak, 1986; Stuffins, 1981). Others say that speech may be negatively influenced by this protocol, because of delayed hard palate closure, inherent to PIO (Harding & Grunwell, 1993; Winters & Hurwitz, 1995; Witzel, Salyer, & Ross, 1984). Because the relevance of PIO treatment has not been experimentally documented, it remains controversial.

In 1993, a prospective clinical trial on the effects of infant orthopedic treatment according to the Zurich approach was started. The study investigates speech language development, surgical and orthodontic aspects, and more general aspects such as feeding and parental satisfaction (Kuijpers-Jagtman, Konst, Prahl, & Severens, 1998). Two groups are being observed in the trial: a group of children treated with PIO (the PIO group) and a group who did not receive this therapy (the NPIO group). The children entered the trial within 2 weeks after birth and were assigned to one of the treatment groups by means of computerized balancing for alveolar cleft width and birth weight. In the PIO group, a palatal appliance was inserted as soon as possible after birth. This appliance was made of soft and hard acrylic and covered the whole cleft, including the alveolar ridge and the soft palate. Apart from treatment with PIO, all interventions were the same in both cleft groups.

A report on the assessment of prelexical development of the children involved in this clinical trial showed that PIO facilitated the production of alveolar consonants, at least in the short term (Konst, Weersink-Braks, Rietveld, & Peters, 1999). In the present article, the effects of treatment with presurgical infant orthopedics on competence in oral communication reflected by intelligibility at age 2.5 years is described. The results for children with cleft lip and palate are compared with those of a control group of noncleft peers.

The literature provides information about speech characteristics and development in children with clefts, regardless of whether they are compared with noncleft peers, but the parameter of speech intelligibility has given rise to controversy. Although the topic of speech intelligibility has been of great interest and concern to professionals and researchers concerned with several other impairment, such as motor speech disorders and hearing impairment, it has been studied less in children with cleft palate. Authors who omitted speech intelligibility from their framework did so because of reliability and validity issues (Sell, Har-

ding, & Grunwell, 1994; Wyatt, Sell, Russell, Harding, Harland, & Albery, 1996). An additional reason for omitting this variable is that it is not only influenced by cleft-related variables, but also by intonation, accent, stress, and rate (Wyatt et al., 1996). Research into intelligibility of individuals who are hearing impaired or have motor speech disorders has shown that suprasegmental factors, such as phonatory control, timing, and speech rate, affect intelligibility (Olson Ramig, 1992; Weismer & Martin, 1992). Moreover, linguistic aspects such as context, redundancy, syntactic complexity of the utterance, and the use of ungrammatical structures may also influence speech intelligibility (Garcia & Dagenais, 1998; Yorkston, Dowden, & Breukelman, 1992). In addition, the relationship between intelligibility and articulation errors in cleft palate speech is not altogether clear (Subtelny, Van Hattum, & Myers, 1972). These authors compared articulation ratings to intelligibility ratings and found that only 4% of the samples were judged to be unintelligible, whereas 34% were rated as severely defective in articulation. They concluded that cleft palate speech characterized by many articulation errors was not inevitably rated as unintelligible.

Although there is controversy about the relevance and reliability of the parameter speech intelligibility, other authors proposed to include intelligibility in the minimal standards for reporting treatment results (Dalston, Marsh, Vig, Witzel, & Bumsted, 1988). We regard intelligibility as a basic aspect of speech performance because it provides an estimate of the viability of communication. Especially in young children, good intelligibility is not only indispensable for successful interaction with the environment, but also indirectly important for speech and language development that, at that age, is influenced by interaction with other language users.

Intelligibility can be measured using a transcription procedure whereby listeners are asked to write down what they understand of the speech. We agree with Samar and Metz (1988) that "the write-down paradigm, arguably, has clear face validity with respect to our current intuitions of the necessary properties of a suitable speech intelligibility construct." (p. 307) The main disadvantage of a write-down method, however, is that it is time consuming. Therefore, in clinical practice and in research experiments, rating scales are commonly used to assess intelligibility. This is a convenient method that does not require much time to obtain an impression of speech performance, but its use can only be justified if reliability and validity can be demonstrated.

The validity of clinical intelligibility ratings in patients with cleft palate ranging in age from 6 to 43 years was investigated by Subtelny et al. (1972). The correlation between the ratings and the write-down data in that study was statistically significant, but not very high ($r = 0.70$; $n = 104$). Van Erp (1991) reported a similar correlation coefficient of 0.74 in an experiment with four trained listeners ($n = 10$).

In another study, Subtelny (1977) described the validity of intelligibility ratings in a different type of speaker group, namely hearing-impaired speakers. This investigation yielded a much higher correlation of 0.87 ($n = 156$) between

the ratings and the write-down findings. Curvilinearity was present in the correlational data, suggesting that the association was not consistent over the entire intelligibility range. Samar and Metz (1988) also investigated the validity and reliability of this rating scale procedure in a group of hearing-impaired speakers. They determined the criterion validity by calculating the correlation between the write-down scores and the ratings. Although they found a high criterion validity coefficient of 0.94 for the rating scale measurements, their results suggested that rating scale scores could not distinguish adequately between those whose transcription scores ranged between 20% and 80%. They concluded that the write-down procedure was superior to the scaling method and found little justification for the use of rating scales, although they restricted this conclusion to intelligibility assessment of hearing-impaired persons.

In our study, intelligibility measurements were performed on toddlers with unilateral cleft lip and palate (UCLP) in the framework of a prospective clinical trial into the effects of presurgical infant orthopedic treatment. Intelligibility was assessed by two methods: a write-down method (with clear face validity, but time consuming) and a rating scale procedure (requiring less time, but with unproven validity).

METHOD

Patients

Twenty-eight toddlers aged 2.5 years participated in this study. Twenty were born with complete UCLP, whereas eight had no history of palatal clefting. All children with clefts participated in the Dutch intercenter prospective clinical investigation of the effects of PIO. In the study reported here, a subgroup of 10 children (eight male, two female) with cleft palate was treated with the PIO; a second group of 10 children (nine male, one female) with cleft palate did not receive this treatment. The eight children without cleft (two male, six female) served as a second control group.

All children with cleft palate entered the trial as soon as possible after birth. Apart from treatment with PIO, all interventions were the same in both cleft groups. Surgical intervention took place at 15 weeks of age (primary lip repair according to the Millard technique) and at 12 months (velum closure according to a modified Von Langenbeck procedure). Hard palate surgery was delayed until approximately 8 years of age, as is usual in the PIO treatment regimen (Hotz et al., 1986). All children were of Caucasian origin with both parents fluent in Dutch. None of the children had been diagnosed as having cognitive or neurologic impairment, or other congenital malformations. All participants, except for one child in the NPIO group, had normal hearing at the time of the recording. Eleven children (six PIO and five NPIO) in the cleft palate group and two control group children went through middle-ear infections with hearing loss in the past. The other children, nine children with cleft palate (four PIO and five

NPIO) and six children in the noncleft control group never had ear infections or hearing problems. The children's receptive language skills measured with a standardized test (Reynell Developmental Language Scales—Dutch Version) were within the normal range for six children in the PIO group, for eight children in the NPIO group, and for all control group children. Four children in the PIO group could not be tested because they were lacking concentration. In the NPIO group, the receptive language skills of two children were below normal.

The treatment protocol required the children in the PIO group to wear the appliance day and night until velum closure at 12 months of age. However, two children stopped wearing the plate at an earlier stage because they rejected the appliance.

Data Collection

A sample of spontaneous speech was recorded from all the children in the child's home environment by the same team of investigators. Speech was recorded using high-quality audio equipment (Sony TCD-D7 DAT Walkman with a Sennheiser MD421U-4 dynamic microphone) and a Panasonic NV-M40E video camera. The child and one of the researchers were engaged in semistructured play with a fixed set of toys. The toys were selected to represent a range of phonemes in words that are seen in the expressive vocabulary of a normally developing 2.5-year-old child (see Appendix). Most words contained singleton consonants in a one- or two-syllable structure. It was ensured during recording that each recorded conversation sample contained a minimum of 100 well-recorded utterances and that it was representative for the child's speech according to the parent.

A unique subset of 10 utterances per child was selected from the recorded sample to be used as stimulus material in the experiment. All these utterances were spontaneous with a clarified exact meaning for all the words. Clarification of meaning was carried out by one of the investigators while listening to the rephrasing in the conversation and watching the video recording showing the context of the conversation (toys that the child was playing with). Mean length of utterance was also accounted for while selecting the utterances: the mean length of the 10 selected utterances had to equal the mean length of utterance of the entire conversation sample. Furthermore, to control for the role of contextual cues, semantically related utterances were not used.

Procedure

Speech intelligibility was assessed by means of two methods: a write-down method and a rating procedure. Sixteen inexperienced, lay listeners (seven men and nine women) of between 20 and 40 years of age participated in the experiments. The listeners were either university students or graduates. They all had normal hearing and were not familiar with cleft palate speech or with the objective of this study. A high-quality digital stimulus tape was constructed for use in the listening experiment. The tape included 36 speech samples of 10 spontane-

ous utterances each, all recorded with a sampling frequency of 40 kHz. Twenty of the samples were taken from the children with cleft lip and palate (10 PIO and 10 NPIO); there were eight samples from the normally developing children. Six samples were duplicated for calculation of intrarater reliability, and two additional samples of cleft palate speech were included to give the listeners an opportunity to practice the task. The tape was administered by means of headphones to the listeners; the video recordings were not used in the experiment. The 10 utterances of each child were presented to the listeners in sequence; the order in which these samples were presented was randomized. Each utterance was played twice with a subsequent pause, which enabled the listener to indicate in normal spelling what he or she had understood of the utterance. Immediately after the transcription of the last utterance of a sample, the listeners rated the intelligibility of the entire speech sample on a 10-point rating scale marked by the contrasting labels "unintelligible" (rating 1) to "intelligible" (rating 10). After the experiment, the written responses were compared with the exact meaning of the utterances, and the number of correct words per sample was counted. The write-down intelligibility score of a sample was defined as the proportion of correctly perceived words of the total number of words.

Statistics

To determine whether the intelligibility ratings and the write-down intelligibility scores were reliable, Cronbach's α was used to compute interrater reliability. Intrarater reliability was defined by the Pearson's product moment correlation between test and retest scores. Criterion validity of the intelligibility ratings was determined by the correlation (Pearson's r) between the write-down scores and the ratings. The main objective of this study was to evaluate a difference in intelligibility between the two cleft groups. Student t test was used for this purpose because of its robustness. An additional one-way analysis of variance and a post hoc test (Tukey's Honest Significant Difference (HSD) test) were used to compare the cleft groups with the control group.

RESULTS

Interrater Reliability

Write-down task. A high interrater reliability (16 raters) of 0.99 (Cronbach's α) was obtained for the write-down procedure. Tukey's test for additivity was not significant ($p > 0.05$). Therefore, interaction between rater and objects ($n = 28$) was not present in the data.

Intelligibility ratings. The coefficient obtained for reliability of the listeners' judgments of intelligibility was also very high. Cronbach's α yielded 0.97 for the rating scale scores (see Table 1). Again, there was no interaction between rater and objects ($p > 0.05$).

Intrarater Reliability

Write-down task. Almost perfect correlation was found between the test and the retest samples ($n = 6$) in the transcription task (Pearson's $r = 0.98$; $p < 0.01$). A t test for paired observations showed that the mean write-down scores of the test and the retest samples did not differ significantly in magnitude ($t_5 = -0.10$; $p > 0.05$).

Intelligibility ratings. Intrarater reliability for the intelligibility ratings was also very high when tested with six test–retest samples. The correlation between the ratings and the repeated ratings was 0.95 ($p < 0.01$). Again, there was no significant difference between the ratings and the repeated ratings ($t_5 = -0.25$; $p > 0.05$).

Validity of Intelligibility Ratings

The validity of the ratings was evaluated by calculating the correspondence between the ratings and the write-down scores. For this purpose, the percentage of correctly perceived words was transformed into arcsin values. The (Pearson) correlation between the arcsin values and the ratings yielded 0.88. This coefficient may be corrected for attenuation resulting from the slightly imperfect reliability of the write-down scores (Pedhazur, 1982) by dividing the observed correlation by the square roots of the two reliability coefficients. The resulting validity coefficient was 0.90 $(0.88 / [\sqrt{0.99} \times \sqrt{0.97}] = 0.90)$. A scattergram of these values shows that the correlation for ratings of 4 and higher was good (Figure 1), whereas for ratings of less than 4, the correspondence was poorer. The correlation coefficients for the corresponding subsets, which are listed in Table 2, confirm this finding: they were 0.89 for ratings of 4 and higher ($n = 14$) and only 0.26 for ratings less than 4 ($n = 14$). A test for curvilinearity (Table 2) showed that the relationship was linear when measured over the whole range of scale values ($p < 0.01$) and also linear for ratings of 4 and higher ($p < 0.01$); however, this was not valid for ratings of less than 4 ($p > 0.05$).

The disappointing correlation between the two intelligibility methods at the lower end of the intelligibility range was further investigated by calculating the standard deviations of the scores from both methods. Figures 2 and 3 show the results of this analysis. Standard deviations of the scores from the write-down

Table 1. Interrater (16 raters) and Intrarater (Six Retest Samples) Reliability of Two Intelligibility Methods

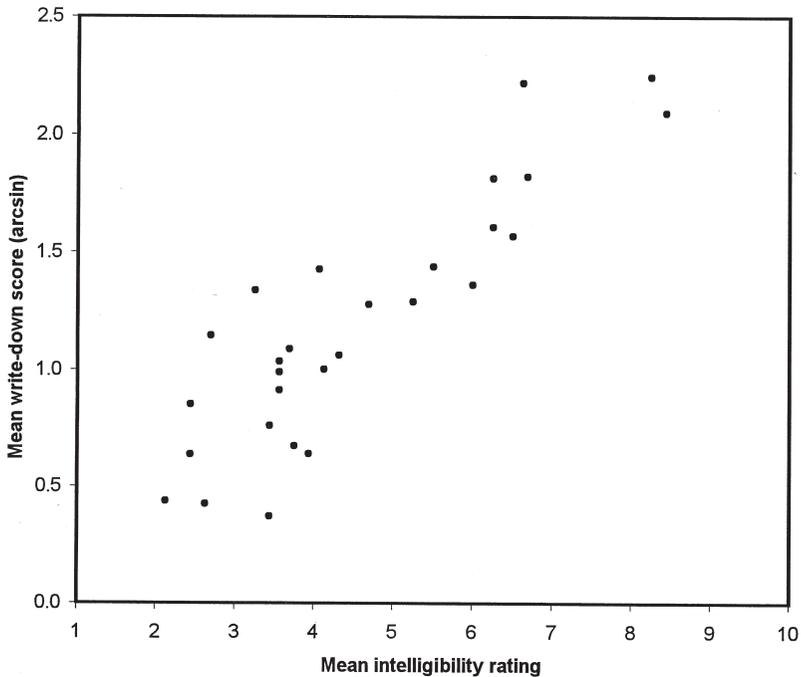|	Interrater Reliability (Cronbach's α)	Intrarater Reliability (Pearson's r)
Write down method	0.99	0.98
Rating procedure	0.97	0.95

Figure 1. The correlation between the mean intelligibility rating and the mean write-down score (arcsin).

method were relatively high at the lower end of the intelligibility range, where few words were correctly recognized and where they decreased proportionately to the number of words correctly understood. This relationship was not present in the standard deviations from the ratings. Relatively high standard deviations were only found for ratings of between 3 and 4 and between 6 and 7.

To determine the stability of the rating scale measure, the true range of write-down scores corresponding with a given scale value was assessed. We calcu-

Table 2. Pearson's Product Moment Correlation Between Arcsin Transformation of the Percentage of Words Correct and the Listeners' Judgements

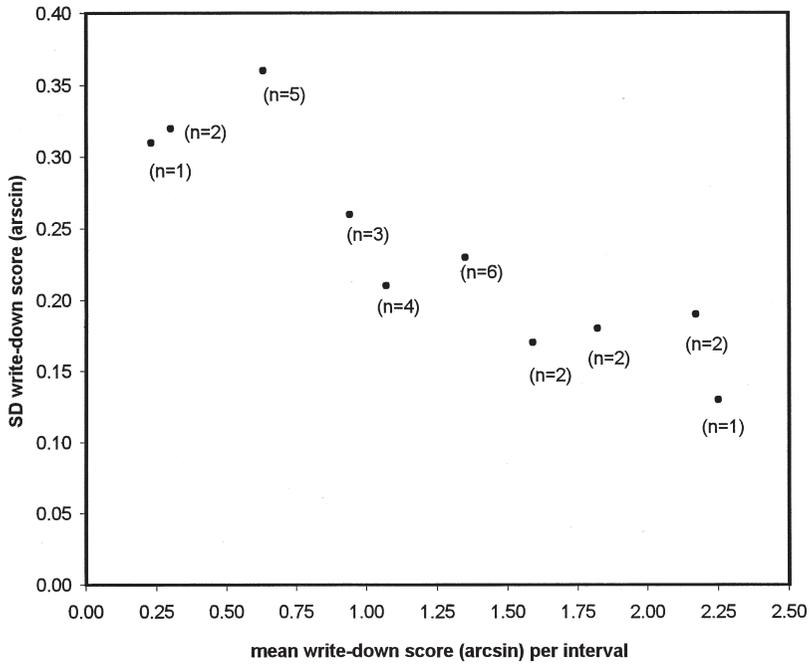|	Listeners' Judgements	Linearity of Relationship
Arcsin of % words correct over entire scale range ($n = 28$)	0.88	Linear, $p = 0.000$
Arcsin of % words correct over scale values of 4 or higher ($n = 14$)	0.89	Linear, $p = 0.000$
Arcsin of % words correct over scale values <4 ($n = 14$)	0.26	Nonlinear, $p = 0.374$

Figure 2. Standard deviation of the write-down score plotted per write-down interval. n = the number of observations in each interval.

lated the mean write-down scores and 95% confidence intervals for all ratings in a given scale interval. For example, the mean write-down score and 95% confidence interval was computed for all ratings of between scale values 2 and 3 ($n = 5$), between scale values 3 and 4 ($n = 9$), and so on. Figure 4 shows the mean write-down score in percentages with the 95% confidence intervals plotted against the scale values. All 95% confidence intervals were smaller than 0.09. This means that over the entire range of intelligibility ratings, we can state with 95% confidence that the true write-down score never differed by more than 4.5 percentage points from the observed mean. No relationship was found between the rated value and size of the confidence interval: confidence intervals were of similar size over the entire range of intelligibility ratings. These results indicate that the rating paradigm used in this experiment was a stable measure.

Speech Intelligibility

Effect of presurgical infant orthopedic treatment on speech intelligibility. The criterion measure for intelligibility (i.e., the write-down task) did not show a statistically significant difference between the two cleft groups. There
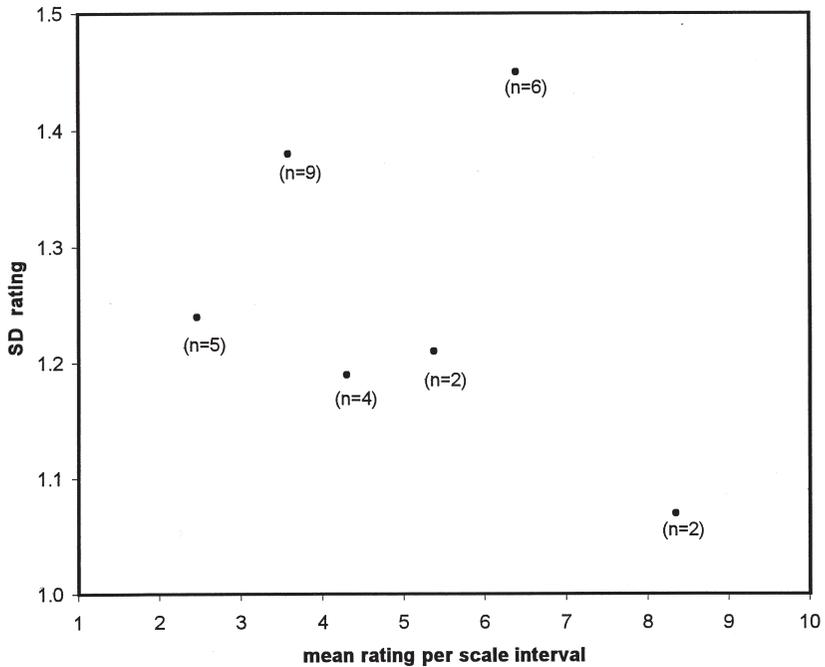
Figure 3. Standard deviation of the intelligibility ratings plotted per scale interval. n = the number of observations in each interval.

was a mean percentage of 31.2 correctly perceived words in the PIO group ($SD =$ 0.24) versus 18.1% in the NPIO group ($SD = 0.08$; see Table 3). Although the speech of the children treated with PIO was—on the whole—better understood than that of the children without PIO, the variation within the groups was such that a t test did not reach significance ($t_{18} = 1.30$; $p > 0.05$). The power of the test was low (0.26), which may be explained by the small number of objects.

In contrast with the criterion measure, the rating experiment showed a significant difference between the two cleft groups ($t_{18} = 2.44$; $p < 0.05$). Children treated with PIO obtained higher ratings for intelligibility (mean rating = 4.51) than children who did not receive PIO treatment (mean rating = 3.29).

Comparison with noncleft peers. The results of the two cleft groups were compared with those of their noncleft peers by means of a one-way analysis of variance followed by post hoc comparisons (Tukey's HSD test). The analysis of variance showed a significant effect ($p < 0.01$) both for the arcsin of the write-down variable and for the ratings. Post hoc comparisons (see Table 4) revealed that the noncleft children were significantly more intelligible than the two cleft groups ($p < 0.05$). The percentage of correctly perceived words in noncleft children's speech was 56.5, versus 31.2% (PIO) and 18.1% (NPIO).
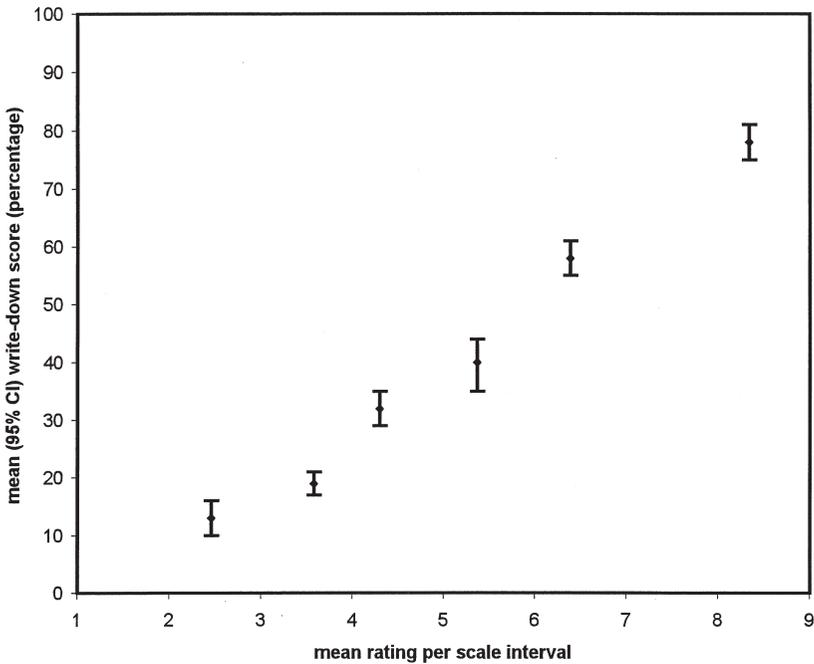
Figure 4. Mean write-down score with 95% confidence interval plotted against the mean intelligibility rating.

Tukey's HSD test also discriminated the intelligibility ratings for children with cleft lip and palate from the normally developing children, whose speech intelligibility was rated higher. The difference in intelligibility rating between the two cleft groups obtained with the t tests was not significant when tested against the control group by means of the Tukey procedure.

Table 3. Results of The t-Tests: Mean Values and Standard Deviations for Both Cleft Groups, t Value and Significance of Difference Between Groups

	Presurgical Infant Orthopedic Treatment ($n = 10$)		No Presurgical Infant Orthopedic Treatment ($n = 10$)			
	Mean	SD	Mean	SD	t (df $= 18$)	p
Percentage words correct	31.2	(23.8)	18.1	(8.12)		
Arcsin of % words correct	1.09	(0.62)	0.81	(0.26)	1.30	0.22
Listeners judgement	4.51	(1.41)	3.29	(0.70)	−2.44	0.03

df = degrees of freedom.

Table 4. Results of Post Hoc Comparisons (Tukey's HSD Test): Mean Values and Standard Deviations for Three Groups and the Significance of the Difference Between the Groups

	Presurgical Infant Orthopedic Treatment (PIO; $n = 10$)		No Presurgical Infant Orthopedic Treatment (NPIO; $n = 10$)		No Cleft Palate (NCP; $n = 8$)		Difference Between		
	Mean	SD	Mean	SD	Mean	SD	PIO–NPIO	PIO–NCP	NPIO–NCP
Percentage words correct	31.2	23.8	18.1	8.12	56.5	16.3		*	*
Arcsin of % words correct	1.09	0.62	0.81	0.26	1.71	0.35		*	*
Rating untrained listeners	4.51	1.41	3.29	0.70	6.18	1.81		*	*

*difference significant at $p < 0.05$.

DISCUSSION

Speech intelligibility can be defined as the match between the intention of the speaker and the perception of the listener. The write-down paradigm, in which words intended by the speaker are compared with words understood by the listener, meets this definition very well and was therefore chosen as the criterion measure in our study. In clinical practice and experimental research, intelligibility is often assessed using rating scales. A rating scale procedure is only justified if its reliability and validity are proven. The reliability of the rating procedure in our study appeared to be very high: the lay listeners who rated the speech were able to do so reliably. Nevertheless, it is not certain whether the ratings are valid and whether they reflect the viability of oral communication defined as intelligibility. As Samar and Metz (1991) stated: "No matter what the raters think they are listening for and rating, there is no a priori guarantee that the most salient perceptual features of complexly disordered speech will correspond to the target construct. It is one thing to tell raters what you want them to rate. It is quite another thing to demonstrate that they are actually capable of rating it." (p. 701).

The high correlation between the write-down task and the rating procedure in our study indicated that both methods measured the same concept to a large extent. A closer examination of the relationship between the two methods at issue revealed that the correlation at the lower end of the intelligibility range (ratings of less than 4) was low and not significant. Apparently, the ratings were a good reflection of the write-down scores for more easily intelligible speech, but not for poorer speech. The poor correspondence between the two intelligibility methods at the lower end of the intelligibility range may have resulted from the wide variation in the lower write-down scores. The standard deviations for the write-down task were larger—on the whole—when only a few words were correctly recognized. Apparently, when speech was almost unintelligible, listeners varied more in their ability to understand this speech. Obviously, guessing what the speaker intended is easier in fairly intelligible speech because of greater redundancy of the message. It seems likely that guessing produces the same results for most listeners when the speech is not too difficult to understand; therefore, variation in the write-down scores decreased with better intelligibility. Although none of the raters were experienced in listening to children's utterances, some of them were fairly competent in reconstructing speech with low intelligibility.

The above discussion shows that the two intelligibility methods did not measure exactly the same concept. When both methods were used to assess the intelligibility of two groups of differently treated children with cleft lip and palate, the rating paradigm showed a significant difference between the groups, whereas the write-down procedure did not. An explanation for this result could lie in the fact that the write-down task was relatively insensitive to aspects of speech quality other than intelligibility, whereas these aspects may have influ-

enced ratings. Ratings may, for example, be negatively affected by a characteristic such as nasality, which may be distracting but not necessarily degrading to the intelligibility. Although the listeners in our experiment were instructed to rate intelligibility on the basis of the percentage of words they understood, it is not certain that they actually did so. When the ratings (range, 1–10) were compared with the write-down scores, it appeared that the listeners systematically overrated the intelligibility of the speakers. This notion was confirmed by the regression equation (% rating $= -0.2008 + 11.8 \times$ proportion write-down).

A note should also be made about the requirement in the write-down paradigm that the intention of all utterances must be known by the investigator. The use of a fixed set of toys and the video recording always provided the investigators with enough contextual cues to select and understand the 10 utterances per child. However, the speech of some children with cleft lip and palate was very difficult to understand, even when contextual cues were present. In children who produced many highly unintelligible utterances, the representativeness of the sample was degraded by the requirement of the write-down paradigm that only completely comprehended utterances can be used. In these cases, the intelligibility measurements may therefore have overestimated the children's actual communication competence.

Another remark concerns the speech material used in this study. Picture naming or having the child repeat a list of words would have sufficed to assess intelligibility. However, the same material was also used to investigate other speech and language aspects, so a conversation sample of spontaneous speech was necessary. Typically spontaneous speech varies widely in content and utterance complexity. It is known from the literature (Garcia & Dagenais, 1998; Kent, Miolo, & Bloedel, 1994; Osberger, 1992) that linguistic context, semantic predictiveness, and length of utterance play an important role in how well speech is understood. In this study, we attempted to overcome these difficulties by using a fixed set of toys in the conversation and by accounting for mean length of utterance when selecting utterances for the experiments. Care was also taken not to include utterances that were semantically related to each other.

A final remark should be made about the absence of visual cues in the stimulus material. Obviously, contextual support provided by visual information is an important aspect in communication. Visual information might have improved the listeners' understanding of the short sentences of these toddlers. It should therefore be emphasized that the intelligibility assessment in these experiments only provided information about the competence of the speaker to transmit a message by means of spoken language, and did not reflect the child's competence as an interactive communication partner.

CONCLUSIONS

The main purpose of this study was to investigate the effect of treatment with presurgical infant orthopedics on the speech intelligibility of young children

(2.5 years) with cleft lip and palate. Lay listeners used two assessment methods to judge speech intelligibility: a write-down paradigm and a rating scale procedure. The reliability of both methods was high. When the ratings were correlated with the criterion measure (a write-down paradigm), good validity was observed for ratings of 4 and higher. Correlation between the two methods at the lower end of the intelligibility range was poorer. The results show that children in the treatment group were rated as exhibiting greater intelligibility than those in the nontreatment group, however, data obtained by means of transcriptions indicated that, in fact, there were no group differences in actual intelligibility. Apparently, the prosthesis facilitated speech parameters other than intelligibility that influenced the listeners to give higher ratings. Although the use of PIO may enhance the desirability of the perceived speech, it should be made clear to clinicians and parents that intelligibility should not be expected to improve as a result of the use of the prosthesis.

REFERENCES

Dalston, R.M., Marsh, J.L., Vig, K.W., Witzel, M.A., & Bumsted, R.M. (1988). Minimal standards for reporting the results of surgery on patients with cleft lip, cleft palate, or both: A proposal. *Cleft Palate Journal, 25*, 3–7.

Garcia, J.M., & Dagenais, P.A. (1998). Dysarthric sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language, and Hearing Research, 41*, 1282–1293.

Gnoinski, W.M. (1990). Infant orthopedics and later orthodontic monitoring for unilateral cleft lip and palate patients in Zurich. In J. Bardach & H.L. Morris (Eds.), *Multidisciplinary management of cleft lip and palate* (pp. 578–585). Philadelphia: W.B. Saunders.

Gruber, H. (1990). Presurgical maxillary orthopedics. In J. Bardach & H.L. Morris (Eds.), *Multidisciplinary management of cleft lip and palate* (pp. 592–600). Philadelphia: W.B. Saunders.

Harding, A., & Grunwell, P. (1993). Relationship between speech and timing of hard palate repair. In P. Grunwell (Ed.), *Analysing cleft palate speech* (pp. 142–160). London: Whurr Publishers.

Hotz, M. (1979). Multidisziplinäre Betreuung von Patienten mit Lippen-Kiefer-Gaumen-Spalten in Zürich. *Stomatology, 29*, 944–954.

Hotz, M. (1983). Orofacial development under adverse conditions. *European Journal of Orthodontics, 5*, 91–103.

Hotz, M., Gnoinski, W., Perko, M., Nussbaumer, H., Hof, E., & Haubensak,

R. (1986). The Zurich approach, 1964 to 1984. In M. Hotz, W. Gnoinski, M. Perko, H. Nussbaumer, E. Hof, & R. Haubensak (Eds.), *Early treatment of cleft lip and palate* (pp. 42–48). Toronto: Hans Huber Publishers.

Kent, R.D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech Language Pathology, 3*, 81–95.

Konst, E.M., Weersink-Braks, H., Rietveld, T., & Peters, H.F.M. (1999). Pre-lexical development of unilateral cleft lip and palate babies with reference to presurgical infant orthopaedics: A randomized prospective clinical trial. *Clinical Linguistics and Phonetics, 13*, 395–407.

Kuijpers-Jagtman, A.M., Konst, E.M., Prahl, C., & Severens, J.L. (1998). A multicentre prospective clinical trial into the effects of presurgical infant orthopaedics in children with complete UCLP. Nijmegen: University of Nijmegen, second report.

Olson Ramig, L. (1992). The role of phonation in speech intelligibility. In R.D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 119–155). Amsterdam/Philadelphia: John Benjamins.

Osberger, M.J. (1992). Speech intelligibility in the hearing impaired: research and clinical implications. In R.D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 233–264). Amsterdam/Philadelphia: John Benjamins.

Pedhazur, E.J. (1982). *Multiple regression in behavioral research.* New York: Holt, Rinehart and Winston.

Samar, V.J., & Metz, D.E. (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research, 31*, 307–316.

Samar, V.J., & Metz. D.E. (1991). Scaling and transcription measures of intelligibility for populations with disordered speech: where's the beef? *Journal of Speech and Hearing Research, 34*, 699–704.

Sell, D., Harding, A., & Grunwell, P. (1994). A screening assessment of cleft palate speech (Great Ormond Street Speech Assessment). *European Journal of Disorders of Communication, 29*(1), 1–16.

Stengelhofen, J. (1989). *Cleft Palate: The nature and remediation of communication problems.* Edinburgh: Churchill Livingstone.

Stuffins, G.M. (1981). Speech and mental attitudes in the older presurgical child. In: B. Kehrer, T. Slingo, B. Graf, & M. Bettex (Eds.), *Long term treatment in cleft lip and palate.* Proceedings of the first international symposium (pp. 199–206). Bern: H. Huber.

Subtelny, J.D., Van Hattum, R.J., & Myers, B.B. (1972). Ratings and measures of cleft palate speech. *Cleft Palate Journal, 9*, 18–27.

Subtelny, J.D. (1977). Assessment of speech with implications for training. In F. Bess (Ed.), *Childhood deafness* (pp. 183–194). New York: Grune & Stratton.

van Erp, A.J.M. (1991). *The phonetic basis of personality ratings, with special reference to cleft palate speech* [doctoral dissertation, University of Nijmegen]. Leidschendam: Royal PTT Nederland.

Weismer, G., & Martin, R.E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R.D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 67–118). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Winters, J.C., & Hurwitz, D.J. (1995). Presurgical orthopedics in the surgical management of unilateral cleft lip and palate. *Plastic and Reconstructive Surgery, 95*(4), 755–764.

Witzel, M.A., Salyer, K.E., & Ross, R.B. (1984). Delayed hard palate closure: The philosophy revisited. *Cleft Palate Journal, 21*, 263–269.

Wyatt, R., Sell, D., Russell, J., Harding, A., Harland, K., & Albery, E. (1996). Cleft palate speech dissected: A review of current knowledge and analysis. *British Journal of Plastic Surgery, 49*, 143–149.

Yorkston, K.M., Dowden, P.A., Breukelman, D.R. (1992). Intelligibility measurement as a tool in the clinical management of dysarthric speakers. In R.D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 265–286). Amsterdam/Philadelphia: John Benjamins.

CONTINUING EDUCATION

An Intelligibility Assessment of Toddlers with Cleft Lip and Palate Who Received and Did Not Receive Presurgical Infant Orthopedic Treatment

QUESTIONS

1. Why is speech intelligibility often omitted from the research framework in the cleft palate field?
   a. Assessment procedures are too time consuming
   b. Difficulties regarding reliability and validity are considered to be too great and intelligibility is influenced by many variables other than cleft-related articulation errors
   c. Speech intelligibility is not considered to be important because it is not affected by cleft-related articulation errors.

    d. It is not clear how speech intelligibility should be assessed in this population

    e. Intelligibility of cleft palate speech cannot be measured reliably

2. Presurgical infant orthopedics, according to the Zurich, approach is:

    a. Not a common treatment in cleft care

    b. Only used as a feeding plate

    c. Combined with two-stage palatal surgery

    d. Applied as soon as possible after surgical lip closure

    e. Applied as soon as possible after surgical velum closure

3. From this study it can be concluded that:

    a. The reliability of intelligibility ratings given by lay listeners was not sufficient

    b. The reliability of the write-down procedure used by lay listeners was not sufficient

    c. The correlation between the write-down scores and the ratings was high for ratings of 4 and higher

    d. The correlation between the write-down scores and the ratings was poor for ratings of 4 and higher

    e. The correlation between the write-down scores and the ratings was high over the entire intelligibility range

4. The authors concluded that:

    a. Intelligibility ratings should not be used by lay listeners for intelligibility assessment of the speech of children with cleft lip and palate

    b. The write-down method was superior to the rating scale method in intelligibility assessment of children with cleft lip and palate

    c. Intelligibility ratings did not reflect perfectly the percentage of words correctly understood by the listener

    d. The rating scale method and the write-down paradigm were equally time consuming

    e. The rating scale method is a better method of assessing speech intelligibility in cleft lip and palate children than the write-down method

5. The results of this study show that:

    a. Children treated with presurgical infant orthopedics received higher intelligibility ratings from lay listeners than children treated without presurgical infant orthopedics

    b. Children treated with presurgical infant orthopedics received lower intelligibility ratings from lay listeners than children treated without presurgical infant orthopedics

    c. Children treated with presurgical infant orthopedics received higher scores in the write-down experiment than children treated without presurgical infant orthopedics

    d. Children treated with presurgical infant orthopedics received lower scores in the write-down experiment than children treated without presurgical infant orthopedics

e. No differences in intelligibility were found between children treated
   with or without presurgical infant orthopedics

Appendix. Phonemes That Were Targeted in Toys

Initial Consonants	Target Word(s)
/p/	poes (pussy); paard (horse); papa (daddy)
/t/	tas (bag); telefoon (telephone)
/k/	kam (comb); koe (cow); koek (cooky)
/b/	beer (bear); baby (baby); bal (ball)
/d/	dag (bye); doos (box); daar (there)
/m/	mama (mommy); mooi (nice); mes (knife)
/n/	nee (no); neus (nose)
/w/	weg (gone); wassen (to wash)
/j/	jas (coat); ja (yes)
/l/	lekker (good); lamp (lamp)
/f/	fiets (bike)
/s/	sok (anklet)
/v/	vis (fish); varken (pig)

Medial Consonants	Target Word(s)
/p/	appel (apple); open (open); papa (daddy)
/t/	auto (car)
/k/	lekker (good); pakken (to get)
/b/	baby (baby)
/m/	mama (mommy)
/n/	banaan (banana)
/j/	aaien (to stroke)
/l/	hallo (hello); ballon (balloon)
/s/	wassen (to wash)

Final Consonants	Target Word(s)
/p/	aap (monkey); op (finished); pop (doll)
/t/	bad (bath); boot (boat)
/k/	sok (anklet); koek (cooky); boek (book)
/m/	boom (tree); kam (comb); bloem (flower)
/n/	schoen (shoe); banaan (banana)
/l/	bal (ball); appel (apple)
/f/	woef (woof)
/s/	huis (house); vis (fish); neus (nose)
/x/	weg (gone); dag (bye)