

6-1997

Gross Estimation: A Study of the Clinical Validity for Measuring Intelligibility

Heather Gail Clarke
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Communication Sciences and Disorders Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Clarke, Heather Gail, "Gross Estimation: A Study of the Clinical Validity for Measuring Intelligibility" (1997). *Dissertations and Theses*. Paper 5685.
<https://doi.org/10.15760/etd.7557>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

THESIS APPROVAL

The abstract and thesis of Heather Gail Clarke for the Master of Science in Speech Communication: Speech and Hearing Sciences were presented May 14, 1997, and accepted by the thesis committee and the department.

COMMITTEE APPROVALS:

[REDACTED]
Mary Gordon-Brannan, Chair

[REDACTED]
Doug Martin

[REDACTED]
Carol Mack
Representative of the Office of
Graduate Studies

DEPARTMENT APPROVAL:

[REDACTED]
Stephen A. Kosokoff, Chair
Department of Speech Communication

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by

[REDACTED] on

19 June 1997

ABSTRACT

An abstract of the thesis of Heather Gail Clarke for the Master of Science in Speech Communication: Speech and Hearing Sciences presented May 14, 1997.

Title: Gross Estimation: A Study of the Clinical Validity for Measuring Intelligibility

Intelligibility is the most fundamental factor for successful speech communication. Measurements of speech intelligibility carry important clinical consequences that relate to description of severity, need for intervention, intervention goals, service delivery options, and treatment efficacy. It is important, therefore, that speech-language pathologists use an approach that reflects an accurate and valid measure of intelligibility.

The purpose of this study was to investigate the relationship between the two seemingly most common procedures used by practicing speech-language pathologists for measuring speech intelligibility: the gross estimation of intelligibility procedure, and the orthographic transcription procedure. Twelve 100-word connected speech samples were analyzed by 4 trained listeners to determine percentage of speech intelligibility, first by gross estimation, and secondly, by orthographic transcription.

Both procedures resulted in a score that yields a percentage of intelligibility.

Results from this study indicate that mean Pearson-r correlation coefficients for interjudge reliability were strong at .87 and .88, for the gross estimation procedure and orthographic transcription procedure, respectively. Intrajudge reliability using a discrepancy model indicated that 3 of the 4 listeners were relatively consistent in their gross estimation ratings. A high correlation between the gross estimation and the orthographic transcription procedures in determining percent of intelligibility was indicated by the Pearson-r correlation coefficient of .85. Visual inspection of the data, however, indicated that the listeners were, in fact, not consistently in agreement in percentages assigned by the two procedures for individual speakers.

GROSS ESTIMATION: A STUDY OF THE
CLINICAL VALIDITY FOR MEASURING INTELLIGIBILITY

by

HEATHER GAIL CLARKE

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
in
SPEECH COMMUNICATION;
SPEECH AND HEARING SCIENCE

Portland State University
1997

ACKNOWLEDGEMENTS

This sincere word of thanks goes to **Mary Gordon-Brannan** for her patience, advice, and support throughout this project. She made me think even when I thought I just could not think anymore. I thank **Doug Martin** and **Carol Mack** for being members of my committee. Their questions and resulting discussion continue to keep the topic of this thesis an intriguing matter for me.

Without the friendship and support of **Kathy Belfiore** and **Barbara Putnam**, this project may have gone on for quite some time. Kathy was always there, whether it was for the whines or the good times, and always with such words of wisdom. Barbara prayed. A heartfelt thank you to both of you.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
CHAPTER	
I INTRODUCTION	1
Statement of Purpose	3
II REVIEW OF LITERATURE	
Definition of Intelligibility	5
Listener Variability.	6
Measures of Intelligibility	8
Open-Set Word Identification	
Closed-Set Word Identification	
Rating Scales	
Gross Estimation	
Considerations.	13
III METHOD	16
Subjects	16
Measures of Intelligibility	17
Procedures	17
Preliminary Procedures	
Speech Samples	
Speech Sample Presentation	
Scoring	
Reliability	
Data Analysis	21
IV RESULTS AND DISCUSSION	23
Results	23
Reliability	

Intelligibility Analysis

	Discussion	28
	Reliability Measures of Intelligibility	
V	SUMMARY AND IMPLICATIONS	36
	Summary	36
	Implications	37
	Clinical Research	
	REFERENCES	41
	APPENDIXES	
A	INFORMED CONSENT	46
B	LISTENING INSTRUCTIONS	48
C	ORTHOGRAPHIC TRANSCRIPTION INSTRUCTIONS	49
D	RAW DATA BY LISTENER - PERCENTAGE OF UNDERSTOOD IN CONNECTED SPEECH/ GROSS ESTIMATION PROCEDURE	50
E	RAW DATA BY LISTENER - PERCENTAGE OF UNDERSTOOD IN CONNECTED SPEECH/ ORTHOGRAPHIC TRANSCRIPTION PROCEDURE	51
F	LISTENER ESTIMATIONS FOR SAMPLES RATED TWICE	52
G	RAW DATA BY LISTENER - PERCENTAGE OF WORDS UNDERSTOOD IN CONNECTED SPEECH/COMPARISON OF GROSS ESTIMATION AND ORTHOGRAPHIC TRANSCRIPTION PROCEDURES	53
H	RAW DATA BY LISTENER IN RANDOM SPEAKER ORDER - PERCENTAGE OF WORDS UNDER- STOOD IN CONNECTED SPEECH/ ORTHOGRAPHIC TRANSCRIPTION PROCEDURE	54

LIST OF TABLES

TABLE		PAGE
1	Gross Estimation Correlation Matrix for Intelligibility Measures . . .	24
2	Orthographic Transcription Correlation Matrix for Intelligibility Measures	24

CHAPTER I

INTRODUCTION

The most fundamental factor for successful speech communication is intelligibility of a speaker by a listener (Bernthal & Bankson, 1988; Connolly, 1986; Kent, Miolo, & Bloedel, 1994). In fact, speech intelligibility has been designated as the single most practical measure of communication competency (Metz, Samar, Schiavetti, Sitler, & Whitehead, 1985). For this study, speech intelligibility is operationally defined as the percentage of words understood by a listener (Gordon-Brannan, 1993). Assessment of intelligibility is of prime importance when evaluating communication disorders in children. Judgments concerning the level of intelligibility in children with communication disorders carry important clinical consequences (Gordon-Brannan, 1993; Kent et al., 1994; Rafaat, Rvachew & Russell, 1995; Yorkston, Strand, & Kennedy, 1996). Speech intelligibility measures resulting from assessment directly relate to descriptions of severity, need for intervention, intervention goals, service delivery options, and treatment efficacy.

A review of the literature indicates there are many approaches to assessing speech intelligibility. There is little agreement, however, as to which procedure best addresses the measurement of speech intelligibility. The

two most common procedures seem to be orthographic transcription and gross estimation. Orthographic transcription, which Gordon-Brannan (1994) described as the standard for determining percentage of intelligibility, is a procedure used by speech-language pathologists (SLPs) in which the percentage of words understood in an utterance is obtained by writing down each intelligible word and noting unintelligible words (Connolly, 1986; Gordon-Brannan, 1994; Kent et al., 1994; Weston & Schriberg, 1992; Yorkston et al., 1996). A percentage of intelligibility is derived by dividing the number of intelligible transcribed words by the total number of words in the speech sample. This procedure is labor intensive, time consuming, and tedious (Bacon, 1995; Gordon-Brannan, 1993, 1994; Kent et al., 1994; Metz et al., 1985, Schiavetti, 1992).

SLPs more often use a gross estimation approach that is time efficient and requires no tracking or transcription of any kind (Bacon, 1995; Gordon-Brannan, 1994; Morris, Wilcox, & Schooling, 1995; Rafaat et al., 1995). SLPs use their trained ears to assign an estimated percentage of intelligibility purely from listening to a continuous speech sample. The question raised by the gross estimation approach is: Is the gross estimation procedure of measuring intelligibility in young children by trained listeners an accurate and valid measure of intelligibility that can be used in conjunction with a standardized measure

to qualify a child for services, or is an approach requiring orthographic transcription necessary to determine percentage of speech intelligibility? If it can be shown that there is a high correlation between the two procedures, then the less labor intensive and time efficient gross estimation procedure can be considered to be useful as a complementary measurement of intelligibility for determining whether or not speech and language services are needed, and/or as a measure to document gains as a result of treatment procedures.

STATEMENT OF PURPOSE

The purpose of this study was to investigate the relationship between the gross estimation of intelligibility procedure by trained listeners and the procedure of orthographic transcription by the same listeners when assessing intelligibility in young children with a wide range of phonological proficiency. The research question addressed was: What is the correlation between the gross estimation of percentage of intelligibility and the percentage of words understood derived from the orthographic transcription procedure? This research question is reflected in the null hypothesis for this study: There is not a significant correlation between the gross estimation of percentage of intelligibility and the percentage of words understood derived from the orthographic transcription procedure.

To address the research question, 12 continuous speech samples were analyzed by trained listeners to determine percentage of speech intelligibility, first by the gross estimation procedure, and secondly, by the orthographic transcription procedure. Both procedures resulted in a score that yields a percentage of intelligibility. The results were analyzed to determine the correlation between the two procedures.

CHAPTER II

REVIEW OF THE LITERATURE

The foci of this literature review are on the definition of intelligibility, listener variability, and common procedures for measuring intelligibility.

Definition of Intelligibility

According to Gordon-Brannan (1993), and for the purposes of this study, speech intelligibility is operationally defined as "the degree to which a person's speech is understood by a listener" (p. 7). The degree of intelligibility is often expressed as the percentage of words understood by a listener. The range of intelligibility extends along a continuum from a message being completely understandable to totally not understandable, with varying levels between the extremes (Bernthal & Bankson, 1988).

The success of speech communication, which encompasses intelligibility, may be influenced by many factors and their relationships, including articulation, phonology, semantics, pragmatics, context, and linguistic structure (Connolly, 1986; Gordon-Brannan, 1993; Hodson & Paden, 1991; Kent et al., 1994; Rice & Wilcox, 1995; Shriberg & Kwiatkowski, 1982; Weston & Shriberg, 1992). Etiological factors such as hearing impairment, cleft palate, laryngectomy, dysarthria, dyspraxia, and severe dysfluency

may impact speech intelligibility (Connolly, 1986).

Kent et al. (1994) added that intelligibility involves both a speaker and listener, hence successful outcomes will also be affected by listener familiarity with the speaker, contextual support (i.e., shared knowledge of time, place, topic, and purpose), message content, linguistic complexity (i.e., sentence structure and length), visual and acoustic signals of speech, and environmental factors. It is likely that a speaker has a "range of intelligibility potentials, depending on listener familiarity, nature of the linguistic message, physical setting, motivation, effort level, and so on" (Kent et al., 1994, p. 81-82).

Listener Variability

Because assessment of intelligibility is primarily accomplished through the auditory-perceptual modality, another important factor regarding judgements of intelligibility is listener variability. Listeners use different strategies when listening to different speakers, and factors, such as past experience with certain clinical populations, familiarity with a particular speaker, and linguistic experience, can affect results.

The research literature refers to experienced or trained listeners as those persons who have received specific and substantial instruction in speech disorders and are often certified speech-language pathologists, audiologists, or deaf educators. Researchers have shown

that experienced listeners generally make higher intelligibility judgements than untrained or inexperienced listeners when listening to speech of clinical populations (Boothroyd, 1985; Monsen, 1983). This may be explained by the ability of experienced listeners who learn to interpret and to adapt to atypical speech and language patterns.

Although not specifically trained, experienced listeners may also include those listeners familiar with a speaker, such as caregivers, siblings, and teachers. Researchers have commented that this important characteristic likely increases intelligibility judgements over those made by unfamiliar listeners (Gordon-Brannan, 1993; Kent, 1996). This may once again be explained by a familiar listener's ability to identify errors and adapt to atypical speech and language patterns.

Another listener characteristic is linguistic experience. Intelligibility is influenced by the linguistic components of speech and language that includes syntax, morphology, phonology, semantics, and pragmatics. Kent (1996) suggested that all listeners have a linguistic background that can have a strong affect on intelligibility judgments.

Intelligibility, therefore, is a complex element of speech. Assessment of speech intelligibility thus becomes a difficult task when these many factors influence the success of speech communication. Researchers have

suggested that assessment of intelligibility involves both measurement of overall intelligibility and analysis of these speech and language factors that underlie reduced unintelligibility (Gordon-Brannan, 1993; Kent et al., 1994; Metz et al., 1985).

Measures of Intelligibility

Review of the literature reveals that numerous approaches have been used by clinicians and researchers to measure speech intelligibility. The procedure chosen will depend on the purpose for judging intelligibility, which in turn will determine the stimulus materials and speech samples to be used (Boothroyd, 1985; Gordon-Brannan, 1994; Hodson & Paden, 1981; Kent et al., 1994; Monsen, 1981, 1983; Yorkston & Beukelman, 1978; Yorkston et al., 1996). A measure of intelligibility may be derived by transcribing and calculating a true percentage of words understood in a speech sample of single-syllable words, multi-syllable words, sentences, or connected speech. Intelligibility may also be measured by listener estimates of the percentage of words understood in a speech sample. Another technique is to use a multiple-choice measurement in which the listener chooses the word that has been spoken from a prepared list of similar sounding words.

Further, researchers group these many approaches under various categories. Gordon-Brannan (1994), for example, used the three general categories of open-set word

identification, closed-set word identification, and rating scale procedures. Kent, et al. (1994) employed five categories, namely: procedures that emphasize phonetic contrast analysis, procedures that emphasize phonological process analysis, procedures that are restricted to word identification without accompanying phonetic or phonological analysis, procedures that derive phonetic indices from continuous speech, and procedures that rely on a scaling method. Yorkston and Beukelman (1978), on the other hand, grouped eight techniques for measuring speech intelligibility into two categories: estimates of intelligibility and objective measures of intelligibility. For the purposes of this research, measurement techniques were reviewed under four classifications, namely, open-set word identification, closed-set word identification, rating scale procedures, and gross estimation. Following this review, comments are made with respect to the two clinically popular techniques that are the focus of this research: orthographic transcription and gross estimation.

Open-Set Word Identification

Orthographic transcription is the traditional measure used in the open-set word identification procedure. This procedure is considered the standard for determining percentage of intelligibility and has been used by many researchers (Connolly, 1986; Gordon-Brannan, 1994; Kent et al. 1994; Weston & Schriberg, 1992; Yorkston et al., 1996).

Speech intelligibility is assessed by the listener transcribing intelligible single words, sentences, or contextual speech samples. A percentage of intelligibility is derived by dividing the number of transcribed intelligible words by the total number of words in the speech sample.

The Weiss Intelligibility Test (Weiss, 1982) was designed to measure intelligibility through a subtest of isolated word identification and a contextual speech subtest. The isolated word identification subtest requires the listener to transcribe 25 words produced by the speaker. The second subtest requires the listener to transcribe a 200-word continuous speech sample. A percentage of intelligibility score is derived by averaging the total number of words understood from the two subtests. This score is compared with intelligibility norms and assigned a severity classification of normal, mild, moderate, severe, or profound.

Closed-Set Word Identification

Another procedure for quantifying intelligibility is identifying words spoken from a word list. The Preschool Speech Intelligibility Measure (P-SIM) (Morris et al., 1995) is a 50-item multiple choice intelligibility test designed for young children. This test requires the child to imitate an examiner's model of 50 randomly selected words. The child's responses are tape-recorded, and from

the tape, a listener identifies the 50 words spoken by the child. A percentage of correctly identified words is calculated and used as a measure of intelligibility.

The Children's Speech Intelligibility Test (CSIT) (Kent et al., 1994), consists of 14 subtests. The child imitates a model of target words and a listener identifies the child's responses from a multiple-choice list of words. Listener responses are analyzed to calculate scores of percentage of words correctly understood and phonological contrast.

Another closed-set word identification intelligibility test, the CID Word SPINE (Speech Intelligibility Evaluation), was developed by Monsen (1981) for use with persons with hearing impairments. This test consists of 10 sets of four consonant-vowel-consonant (CVC) words that are contrastive in a particular phonetic feature. The child reads aloud one of the four words from cards, and the listener identifies the child's response from the four possible choices. After all 10 sets of words have been presented, a percentage of words correctly understood is calculated.

Rating Scales

Scaling procedures require no tracking or transcription. The researcher basically asks "the listener to judge in some way how well his/her responses could match the speaker's list of intended words and yield a value on

some predetermined scale" (Schiavetti, 1992, p. 16). As with other procedures of measuring speech intelligibility, this procedure can be applied to words, sentences, and continuous speech samples. Intelligibility rating scales are typically either equal-appearing interval scales, or direct magnitude scales.

In equal-appearing interval scaling procedures, a number is assigned to a speech sample along a linear continuum (Gordon-Brannan, 1994; Kent et al., 1994; Schiavetti, 1992; Yorkston & Beukelman, 1978). Typically, a five-point, seven-point, or nine-point scale is used, with the various points along the scale corresponding to a descriptor. For instance, on a five-point rating scale, a rating of 1 could correspond to a speech sample being completely unintelligible, whereas a rating of 5 could correspond to a speech sample being completely intelligible. The National Technical Institute for the Deaf (NTID) rating scale is an example of a five-point equal-appearing interval scale (Johnson, 1975).

A direct magnitude scaling procedure involves the listener placing each speech sample along a continuous scale relative to a standard sample. Listeners judge each speech sample "with a number that is proportional to the perceived ratios of speech intelligibility among the speech samples" in contrast to "fixed maximum and minimum numbers at the extreme ends of the continuum" (Schiavetti, 1992, p.

20). The standard may be provided by the researcher, or the listener may assign a number if a standard is not provided.

Gross Estimation

Using a gross estimation procedure, a percentage of intelligibility (i.e., 50% intelligible, 20% intelligible, etc.) is derived from listening to a speech sample. No tracking or transcription of any kind is required. As Young (1969) has stated "... a measurement of a speech disorder is primarily a perceptual event, and the observer's response necessarily represents the 'final' validation for any measurements" (p. 135). Again, as with other procedures for measuring speech intelligibility, this approach can be applied to words, sentences, and continuous speech samples.

Considerations

From the numerous approaches available for assessing speech intelligibility, clinicians and researchers alike face a difficult task when choosing the appropriate approach from which speech intelligibility judgments will be made. Review of the literature indicated that the two most common clinical approaches in practice seem to be orthographic transcription and gross estimation.

Researchers have concluded from studies that the orthographic transcription procedure is an accurate and reliable measure of intelligibility, but have stated that

data collection and analysis are time consuming in comparison to other approaches, making this procedure difficult to use in clinical situations (Bacon, 1995; Kent, et al., 1994; Kwiatkowski & Shriberg, 1992; Samar & Metz, 1988; Schiavetti, 1992).

Research related specifically to a gross estimation procedure is lacking apart from acknowledgements by researchers that SLPs more often use a gross estimation approach that is time efficient and requires no tracking or transcription of any kind (Gordon-Brannan, 1994; Morris et al., 1995; Rafaat et al., 1995). For instance, Morris et al. (1995) stated that "the present measure of choice for assessing intelligibility in preschool children appears to be subjective estimation of performance such as assignment of percentage intelligibility" (p. 22). Gordon-Brannan (1994) claimed that "currently most practicing speech-language pathologists make gross estimates of percentage of intelligibility" (p. 22).

Both procedures result in a score that yields a percentage of intelligibility. Results expressed as percentages are easily understood by both professionals and nonprofessionals. A percentage measure provides descriptive information that can be used as an index of severity (Yorkston et al., 1996); to compare groups of related individuals, such as persons who are hearing impaired (Osberger, 1992) or dysarthric (Yorkston &

Beukelman, 1978); and to document gains as a result of treatment procedures (Gordon-Brannan, 1993; Schiavetti, 1992; Yorkston & Beukelman, 1978).

Throughout the literature, researchers have indicated that decisions regarding choice of procedure to measure intelligibility should include the criteria of validity and reliability. Metz et al. (1985) stated ease of use should be considered a third criterion. A procedure such as gross estimation is not only simple to use, it is a cost effective and a time efficient procedure of measuring speech intelligibility.

SLPs are often called upon to evaluate speech intelligibility in young children to qualify a child for services. In this study the two most commonly used procedures, orthographic transcription and gross estimation, were compared.

CHAPTER III

METHOD

The purpose of this study was to investigate the relationship between the gross estimation procedure and the procedure of orthographic transcription when trained listeners assess intelligibility in young children with a wide range of phonological proficiency. Speech intelligibility is defined as the percentage of words correctly understood by a listener (Gordon-Brannan, 1993). Twelve continuous speech samples were analyzed by trained listeners to determine percentage of speech intelligibility, first by the gross estimation procedure, and secondly, by the orthographic transcription procedure. Both procedures result in a score that yields a percentage of intelligibility.

Subjects

The listeners for this study were comprised of 4 speech-language pathologists employed by the Vancouver School District, Vancouver, Washington. Each listener held certification awarded by the American Speech-Language-Hearing Association, and had a minimum of 10 years of experience working with children with articulation/phonological disorders. More specifically, Listener 1 had the least amount of experience with 10 years, Listener 2 had 26 years of experience, Listener 3 had 18 years of

experience, and Listener 4 held the most experience with 28 years (range = 10 to 28 years; mean = 20.5 years). Additionally, the listeners had normal hearing, and were unfamiliar with the speakers of the continuous speech samples. The listeners signed an informed consent form (Appendix A).

Measures of Intelligibility

The first measure of intelligibility was made by the listeners using their trained ears to assign a percentage of intelligibility using the gross estimation procedure which is derived from listening to a continuous speech sample. The second measure of intelligibility was made by the same listeners using the orthographic transcription procedure which requires the listeners to write down each word they individually understood and noting which words they did not understand. A percentage of intelligibility was derived by dividing the number of intelligible transcribed words by the total number of words in the speech sample and multiplying by 100.

Procedures

Preliminary Procedures

The continuous speech samples used in this study were originally gathered by Gordon-Brannan (1993) in connection with an earlier study on intelligibility. The speakers were children from preschools and speech-language pathology caseloads located in the greater metropolitan area of

Portland. Informed consent forms were signed by the parents/caregivers prior to the commencement of any testing on their children.

The speakers were between the ages of 4:1 (years:months) and 5:6. The children had no known neurological, motor, or physical impairments that could affect speech production. No laryngeal or resonance deviancy existed at the time the speech samples were gathered. The speakers presented with a wide range in articulation/phonological ability, and accordingly for the Gordon-Brannan (1993) study, were placed into three groups of 16 based on intelligibility: (a) Group I, high intelligibility; (b) Group II, average intelligibility; and (c) Group III, low intelligibility.

The speech samples were elicited in an acoustically treated room at Portland State University. The speakers were both audio-taped using a Sharp SX D200 digital audiotape recorder, together with an AKG, Model C451, capacitor flat response microphone; and video-recorded using a Panasonic camcorder, VHS Recorder, AG-100. The 100-word speech samples were elicited using a book, The Relatives Came (Rylant & Gammell, 1985), with the children telling a story.

A scoring key of these samples was prepared by Gordon-Brannan (1993) through orthographic transcription. The content of the transcriptions was verified by the parents/

caregivers of the child who identified the words that were unintelligible or misunderstood in the prepared transcripts.

Speech Samples

The 12 continuous speech samples used in this study were selected by Mowe (1997) from the original continuous speech samples gathered by Gordon-Brannan (1993). The 12 samples were randomly selected such that at least 4 speakers were selected from each of the original three speaker groups. The 12 continuous speech samples were recorded in random speaker order from digital audiotapes onto regular audiocassette tapes.

Speech Sample Presentation

The stimulus material used to elicit the continuous speech samples was shown to the listeners to familiarize them with the story prior to listening to the audiotapes. Following this, the listeners listened to each tape on two separate occasions. For the first presentation, in a group setting, the listeners were instructed to use their trained ears to assign a percentage of intelligibility from listening to each sample, and to write down that percentage on a scoring sheet. The listeners were instructed not to track or to make notes of any kind. Written and verbal instructions were provided by the researcher (Appendix B). Unknown to the listeners and for the purpose of intrajudge reliability analysis, 3 of the 12 continuous speech samples

were presented twice. The repeated samples were placed at the end of the listener tape. For the first presentation, the listeners, then, listened to a total of 15 speech samples.

The second presentation of the speech samples occurred 3 weeks following the initial presentation. The listeners each listened to an audiocassette tape consisting of the 12 continuous speech samples, but each in a different random speaker order from the other. The listeners individually listened to the speech samples on their personal audiotape recorders in their homes, and orthographically transcribed each sample. The listeners were instructed to listen to each sample a maximum of three times. Completed transcriptions were returned to the researcher for scoring. Written instructions were provided by the researcher (Appendix C).

Scoring

The orthographic transcription for each 100-word continuous speech sample from this study were compared for accuracy with the scoring key prepared in the original study (Gordon-Brannan, 1993). Words transcribed that deviated from the scoring key, with the exception of morphological differences, were considered incorrect. A percentage of words understood for each 100-word speech sample was determined by counting the number of words correct.

Reliability

The Pearson product-moment correlation technique (Pearson- r) was used to determine interjudge reliability between each pair of listeners for the gross estimation procedure and orthographic transcription procedure for measuring speech intelligibility. Two correlation matrices were made to display the data.

To determine intrajudge reliability within each experienced SLP, the percentages assigned to the three repeated speech samples were examined through discrepancy scores. The discrepancy model involved determining the arithmetic differences in gross estimation ratings by each listener for the three repeated speech samples.

Data Analysis

All statistical analyses were performed using SPSS release 6.0 (1993, SPSS Inc., Chicago, Illinois). Statistical significance was set at the .05 level of confidence for all data analyses. The Pearson r was used to address the research question of the correlation between the gross estimation procedure for measuring speech intelligibility and the orthographic transcription procedure. The results indicated whether or not the gross estimation procedure significantly correlates with speech intelligibility as measured by the orthographic transcription procedure, and the extent of the correlation. Additionally, to further investigate the relationship

between the 2 measures, the mean and range of differences between the 2 measures for the speakers for each listener was examined by visual inspection.

CHAPTER IV

RESULTS AND DISCUSSION

Results

The purpose of this study was to investigate the relationship between the gross estimation of intelligibility procedure and the procedure of orthographic transcription when assessing intelligibility in young children with a wide range of phonological proficiency. The same listeners provided the data for both procedures. The research question addressed in this study was: What is the correlation between the gross estimation of percentage of intelligibility and the percentage of words understood derived from the orthographic transcription procedure? This was accomplished by comparing intelligibility scores (presented in percentages) obtained for 12 connected speech samples in a known context using the gross estimation procedure and the orthographic transcription procedure.

Reliability

Preliminary to comparing the two measures of intelligibility, interjudge and intrajudge reliability scores were obtained. The percentage data for each speaker sample by each listener appear in Appendixes D and E. Interjudge reliability for the listeners was assessed using the Pearson r . Tables 1 and 2 provide correlation coefficient data for interjudge reliability for the gross estimation and orthographic transcription methods,

Table 1

Gross Estimation Correlation Matrix for Intelligibility Measures

Listener	Listeners			
	1	2	3	4
1	1.00			
2	.87*	1.00*		
3	.95*	.87*	1.00	
4	.82*	.81*	.89*	1.00

Note: Each correlation represents 2 listeners. Marked correlations (*) are statistically significant at $p < .05$.

Table 2

Orthographic Transcription Correlation Matrix for Intelligibility Measures

Listener	Listeners			
	1	2	3	4
1	1.00			
2	.94*	1.00*		
3	.86*	.88*	1.00	
4	.81*	.86*	.94*	1.00

Note: Each correlation represents 2 listeners. Marked correlations (*) are statistically significant at $p < .05$.

respectively. Pearson r correlations for the percentage of words understood in connected speech ranged from .81 to .95 ($p < .05$) with a mean Pearson r of .87 for the gross estimation method, and a range from .81 to .94 ($p < .05$) with a mean Pearson r of .88 for the orthographic transcription method. These values indicate that, within each method, there was strong agreement within the listeners across samples in assessing speech intelligibility.

To examine intrajudge reliability, a discrepancy model was used (Appendix F). For one of the three speakers listened to twice, Listener 1 assigned the same estimated percentage of intelligibility; and for the remaining two speakers, estimates were within 10 percentage points of their first assigned percentage. The discrepancy scores of Listener 1 were 0, +10, and +10, with a discrepancy score mean of 6.66 percentage points. Listener 2 assigned the same estimated percentage on two presentations to one of the three speakers listened to twice. She assigned percentages within 3 and 40 percentage points of the first assigned percentage for the two remaining speakers. The discrepancy scores were +40, +3, and 0, with a discrepancy score mean of 14.33 percentage points. Listener 3 assigned estimated percentages of intelligibility to within 5 percentage points to two of the three speakers listened to twice. She assigned percentages within 15 percentage

points of the first assigned percentage for the remaining speaker. The discrepancy scores were +15, +5, and +5, with a discrepancy score mean of 8.33 percentage points. Listener 4 assigned the same estimated percentage on two presentations to one of the three speakers listened to twice. She assigned percentages within 10 and 15 percentage points of the first assigned percentages for the remaining speakers. The discrepancy scores were +10, 0, and +15, with a discrepancy score mean of 8.33 percentage points. Based on a discrepancy model of investigating intrajudge reliability and looking at the mean difference, 3 of the 4 listeners were relatively consistent in their assignments of gross estimation of intelligibility to the connected speech samples. The fourth listener was consistent on two of the repeated speech samples, but assigned percentages with a wide difference between the first and second presentation of one of the samples.

Intelligibility Analysis

The research question addressed was: What is the correlation between the gross estimation of percentage of intelligibility and the percentage of words understood derived from the orthographic transcription procedure? A Pearson r was used to determine the correlation between the two measures. Intelligibility percentage scores for each procedure for each of the 12 speaker samples is provided in Appendix G. The resultant Pearson r correlation was .85 (p

< .05). This result indicates that the gross estimation of percentage of intelligibility is highly correlated with the percentage of intelligibility derived from orthographic transcription. However, visual inspection of the data reflects that a difference in intelligibility scores exists between the two measures for some of the speakers indicating that the listeners were not consistently in agreement in intelligibility percentages assigned for the two procedures. The mean difference in percentage assigned between the 2 measures was 14% (range = 27 percentage points) for Listener 1, 18% (range = 36 percentage points) for Listener 2, 11% (range = 30 percentage points) for Listener 3, and 6% (range = 15 percentage points) for Listener 4.

In summary, the results from this study indicate that the mean Pearson r correlation coefficients for interjudge reliability were strong for both measures at .87 and .88. Intrajudge reliability using a discrepancy model indicated that 3 of the 4 listeners were relatively consistent in their rating. The Pearson r correlation coefficient of .85 indicates a high correlation between the two methods in determining percent of intelligibility; however, visual inspection of the data indicates that the listeners were not consistently in agreement between the percentages assigned for the two procedures for measuring intelligibility.

Discussion

Reliability

This study investigated the relationship of the seemingly two most common methods for clinically determining percentage of intelligibility in young children with a wide range of phonological proficiency, that is, gross estimation and orthographic transcription (Gordon-Brannan, 1994; Kent et al., 1994; Morris et al., 1995; Weston & Schriberg, 1992; Yorkston et al., 1996). The result of the Pearson- r for interjudge reliability for the gross estimation procedure was a mean of .87 and a mean of .88 for the orthographic transcription procedure. While these findings indicate a high linear association among the listeners, this does not necessarily mean that the listeners were in agreement in percentages assigned, but that the proportional differences between speaker samples was consistent. The percentage data for each speaker sample by each listener appear in Appendixes D and E.

Examples of high variability can be seen by examining the overall range in scores for individual speakers. The range in percentage of intelligibility scores for the gross estimation procedure (Appendix D) was from 10 to 50 percentage points; and for the orthographic transcription procedure (Appendix E), the range was from 11 to 24 percentage points. A wide range in variability can be seen by looking at specific speaker samples. For example,

intelligibility percentages for Sample #7 shows a range from 20% to 70% (50 percentage points) for gross estimation, and the orthographic transcription ranged from 58% to 81% (23 percentage points). In contrast, sample #1 had much smaller ranges with the gross estimation range of 90% to 100% (10 percentage points), and the orthographic transcription range from 83% to 97% (14 percentage points). Variability is also evidenced in that not one of the 12 speaker samples was assigned the same estimated percentage by all four listeners in either the gross estimation or orthographic transcription procedure.

It is true that the speakers were purposely selected in order to exhibit a wide range in articulation/phonological ability and thus a wide range of intelligibility. Further, it was expected that listeners would assign percentage of intelligibility scores for individual speakers within a narrow range. However, this latter prediction was not the case, and such extreme variability, as evidenced by the examples of range in scores for speaker sample #7 and #1, was not expected. This finding supports previous research with regard to listener variability (Boothroyd, 1985; Gordon-Brannan, 1993; Kent, 1996; Monsen, 1983). It is likely that the large range in scores results from the individuality of the listeners and their unique perceptual ability to assess

intelligibility through the auditory-perceptual modality, with some being better interpreters than others.

Variability in percentage of intelligibility scores was expected with the gross estimation procedure as it is a more subjective assessment. Each listener brings a variety of individual skills, as well as individual judgments, when assigning percentages. Thus, 10% intelligibility rating globally assigned by one listener may not mean 10% to another listener. It is not possible to define what 10% intelligibility represents to any given listener. Perhaps listeners were really estimating the comprehensibility of the speech samples rather than intelligibility, the actual percentage of words understood. Yorkston et al. (1996) defined comprehensibility as the extent to which a listener understands a speaker in a communication context. A communication context includes the shared knowledge by the speaker and listener as to time, place, topic, purpose, and other information surrounding the communication event.

Less variability between listeners was expected with the orthographic transcription procedure as it is a more objective measure. Each word is judged individually for understandability. Either the word is intelligible or it is not intelligible. Perhaps a contributing factor in a lower range of scores for orthographic transcription versus gross estimation is the number of times the listener hears the speech sample. When using the orthographic

transcription procedure, SLPs are able to, and often do, listen to the speech sample more than once in order to transcribe what has been heard. For this study, repeated review of each speaker sample was permitted a maximum of three times. This might result in a learning curve for correctly identifying the words uttered by a given speaker with an expected pattern of decreased variability.

However, visual inspection of the orthographic derived data shown in Appendix E reflects a relatively wide range of variability (11 to 24 percentage points), although the range is considerably less than the range for the gross estimation procedure (gross estimation mean range = 27; orthographic transcription mean range = 16). It is possible the decrease in variability resulted from repeated listening of each speech sample, but it is more likely this variability can be attributed to the more objective or precise nature of the orthographic transcription procedure (Gordon-Brannan, 1993; Kent et al., 1994; Shriberg & Kwiatkowski, 1982).

Measures of Intelligibility

It was expected that higher intelligibility scores would occur in each successive speech sample as the listener became more and more familiar with the speech sample content (Boothroyd, 1985; Monsen, 1983). This means that it was expected by the time a listener had listened to and transcribed from the first through the twelfth speech

sample, intelligibility scores would be close to 100%. The listener would have increasingly adapted to the atypical speech and language pattern of each speaker. However, this was not the case, and an order effect did not occur in either the gross estimation procedure, or the orthographic transcription procedure. This is evidenced by examination of the gross estimation derived data shown in Appendix D. For example, Listener 1 rated Sample #1 (Subject #36) as 95% intelligible, Sample #11 (Subject #44) as 70% intelligible, and Sample #12 (Subject #43) as 15% intelligible. Listener 2 rated Sample #1 (Subject #36) as 100% intelligible, Sample #11 (Subject #44) as 95% intelligible, and Sample #12 (Subject #43) as 2% intelligible. Listener 3 rated Sample #1 (Subject #36) as 95% intelligible, Sample #11 (Subject #44) as 75% intelligible, and Sample #12 (Subject #43) as 30% intelligible. Listener 4 rated Sample #1 (Subject 36) as 90% intelligible, Sample #11 (Subject #43) as 75% intelligible, and Sample #12 (Subject #44) as 30% intelligible.

Visual inspection of the orthographic derived data presented in Appendix H provides further evidence that an order effect did not occur. For example, Listener 1 rated Sample #1 (Subject #36) as 96% intelligible, Sample #11 (Subject #43) as 17% intelligible, and Sample #12 (Subject 33) as 73% intelligible. Listener 2 rated Sample #1

(Subject #13) as 47% intelligible, Sample #11 (Subject #9) as 56% intelligible, and Sample #12 (Subject #33) as 65% intelligible. Listener 3 rated Sample #1 (Subject #17) as 68% intelligible, Sample #11 (Subject #9) as 64% intelligible, and Sample #12 (Subject #16) as 36%.

Listener 4 is the only listener who appears to have an order effect as she rated Sample #1 (Subject #33) as 66% intelligible, Sample #11 (Subject #14) as 75% intelligible, and Sample #12 (Subject #36) as 87% intelligible. This could have resulted by random speaker order of least intelligible to most intelligible being presented on Listener 4's speech sample tape.

From the research literature, it was expected that a more experienced listener would have greater agreement in intelligibility scores between the two measures of intelligibility under investigation than a less experienced listener (Boothroyd, 1985; Gordon-Brannan, 1993; Kent, 1996; Monsen, 1983). It was expected that gross estimation intelligibility scores would closely agree with orthographic transcription intelligibility scores. All of the listeners for this study had at least 10 years of experience, but there was an 18-year range of experience. Visual inspection of the data (Appendix G) indicates that Listener 4, who has the most experience of the 4 listeners, did have the most consistent scores between the two procedures investigated. However, Listener 2, who has only

2 years less experience, has the least degree of consistency between the scores. This result was not expected, and again, it is likely that the difference in scores reflects the individuality of the listeners and their unique perceptual abilities.

Visual inspection of Appendix G reveals that 3 of the 4 listeners tended to underestimate intelligibility scores when a comparison is made of the gross estimation procedure to the orthographic transcription procedure. For example, Listener 1 underestimated 10 out of 12 speech samples with a range of 1 percentage point (Sample 1) to 33 percentage points (Sample 3). Listener 3 underestimated 8 out of 12 speech samples with a range of 1 percentage point (Sample 6) to 31 percentage points (Sample 7). Listener 4 underestimated 9 out of 12 speech samples with a range of 2 percentage points (Samples 5 and 11) to 17 percentage points (Sample 4). The fourth listener (Listener 2) underestimated 6 of the speech samples and overestimated 6 of the speech samples. This researcher speculates that this inconsistency results from listener definition of intelligibility. It is possible that when a gross estimation procedure of intelligibility is employed, the listener is actually judging comprehension, whereas, when the orthographic transcription procedure is used, the listener is judging word-for-word whether a word is intelligible or not intelligible. Perhaps in the

estimation procedure, the listeners are estimating the percent of content words understood. This could account for much of the discrepancy between the scores for each method.

This researcher speculates that assessment of percentage of intelligibility depends on listener variability as it relates to the perceptual skills unique to each listener, with some listeners being more skilled than others. Further, it could be construed that the orthographic transcription procedure is a more accurate and precise measure of speech intelligibility. This brings into question whether or not gross estimation is an appropriate procedure for measuring intelligibility as it has been operationally defined for this research project. Further research needs to be conducted using a gross estimation procedure to determine whether listeners are judging comprehensibility or intelligibility.

CHAPTER V

SUMMARY AND IMPLICATIONS

Summary

Intelligibility is the most fundamental factor for successful speech communication. Measurements of speech intelligibility carry important clinical consequences that relate to description of severity, need for intervention, intervention goals, service delivery options, and treatment efficacy. It is important, therefore, that speech-language pathologists use an approach that reflects an accurate and valid measure of intelligibility.

The purpose of this study was to investigate the relationship between the two seemingly most common procedures used by practicing speech-language pathologists for measuring speech intelligibility: the gross estimation of intelligibility procedure, and the orthographic transcription procedure. Twelve 100-word connected speech samples were analyzed by 4 trained listeners to determine percentage of speech intelligibility, first by gross estimation, and secondly, by orthographic transcription. Both procedures resulted in a score that yields a percentage of intelligibility.

Results from this study indicate that mean Pearson- r correlation coefficients for interjudge reliability were strong at .87 and .88, for the gross estimation procedure and orthographic transcription procedure, respectively.

Intrajudge reliability using a discrepancy model indicated that 3 of the 4 listeners were relatively consistent in their gross estimation ratings. A high correlation between the gross estimation and the orthographic transcription procedures in determining percent of intelligibility was indicated by the Pearson-r correlation coefficient of .85. Visual inspection of the data, however, indicated that the listeners were, in fact, not consistently in agreement in actual percentages assigned by the two procedures for individual speakers.

Implications

Clinical

Based upon the findings of this study, this researcher recommends that clinicians undertake their own individual study to determine how accurately their gross estimation percentage scores agree with their orthographic transcription scores, to find the procedure that best suits their unique perceptual ability to assess intelligibility. This could be accomplished by using both procedures for determining percentage of intelligibility for a number of connected speech samples and comparing the resulting scores. Should the scores agree within 10 percentage points of each other, this researcher believes the clinician could then use either method to assess intelligibility accurately. Should greater variability exist, the clinician would consider the purpose for which

intelligibility judgments are being made and choose the appropriate method accordingly.

The gross estimation procedure would be appropriate as a component of a referral screening when determining whether a complete evaluation is warranted for determining whether or not speech and language services are needed. Judgments and statements concerning intelligibility based on a gross estimation procedure would also be appropriate to augment objective evaluation protocols. Measures of intelligibility could be used to document gains in speech intelligibility over time. However, when accuracy and an objective measure is required, the orthographic transcription procedure would be the more appropriate clinical procedure for measuring speech intelligibility.

As this study has brought to question whether a gross estimation procedure does in fact measure intelligibility as operationally defined in this research project, clinicians should be mindful of what they are measuring when using this procedure: comprehensibility or intelligibility.

Research

Further research involving the gross estimation procedure is warranted. A sample size of 4 listeners as was used in this study makes it difficult to generalize to all listeners. It is suggested that this study be replicated with an increased sample size of at least 30

listeners. It is further suggested that listeners determine intelligibility based on percentage of content words understood, rather than all words uttered. Such an approach might provide useful information regarding the discrepancy between the scores for the two methods examined in this study. Perhaps, it will be shown that estimation and percent of content words understood will be less discrepant. It is further suggested that control of the number of times a listener hears the connected speech samples when using the orthographic procedure be implemented. In this study, while listeners were given verbal and written instructions for a maximum of three passes for each speech sample, and as this presentation took place in their homes, it was not possible to control for how many times each speech sample was, in fact, listened to. As stated previously, this may have influenced the results of this study.

It would also be of interest to conduct a study comparing the gross estimation procedure with a different procedure for determining speech intelligibility, such as the check-slash method. This is a less time-consuming procedure than the orthographic transcription procedure, and it would be interesting to determine if this method resulted in similar variability in percentages as did this study.

Another useful study would be an investigation of interjudge reliability when using a gross estimation procedure. A larger number of listeners and a larger number of speech samples could be used. This would provide further information relating to listener variability regarding judgments of intelligibility. A similar study could be undertaken investigating intrajudge reliability.

The results of this study and others demonstrate that measuring intelligibility accurately and reliably is a complex multi-faceted process. The results might also assist clinicians in choosing the procedure that most appropriately reflects that clinician's unique perceptual ability to assess speech intelligibility in children with communication disorders.

References

Bacon, V. J., (1995). Validity and efficiency of the check-slash transcription method for measuring intelligibility. Unpublished master's thesis, Portland State University, Portland, OR.

Bernthal, J. E., & Bankson, N. W. (1988). Articulation and phonological disorders (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Boothroyd, A. (1985). Evaluation of speech production of the hearing impaired: Some benefits of forced-choice testing. Journal of Speech and Hearing Research, 28, 185-196.

Connolly, J. H. (1986). Intelligibility: A linguistic view. British Journal of Communication Disorders, 21, 371-376.

Gordon-Brannan, M. (1993). Speech intelligibility assessment of young children with varying levels of phonological proficiency/deficiency. Unpublished doctoral dissertation, The Wichita State University, KS.

Gordon-Brannan, M. (1994). Assessing intelligibility: Children's expressive phonologies. Topics in Language Disorders, 14, 17-22.

Hodson, B. W., & Paden, E. P. (1981). Phonological processes which characterize unintelligible and

intelligible speech in early childhood. Journal of Speech and Hearing Disorders, 46, 369-373.

Hodson, B. W., & Paden, E. P. (1991). Targeting intelligible speech (2nd Ed.). Austin, TX: Pro-Ed.

Johnson, D. D. (1975). Communication characteristics of NTID students. Journal of the Academy of Rehabilitative Audiology, 8, 17-32.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. American Journal of Speech-Language Pathology, 5, 7-23.

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.

Kwiatkowski, J., & Shriberg, L. D. (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. Journal of Speech and Hearing Research, 35, 1095-1104.

Metz, D. E., Samar, V. J., Schiavetti, N., Sitler, R. W., & Whitehead, R. L. (1985). Acoustic dimensions of hearing-impaired speakers' intelligibility. Journal of Speech and Hearing Research, 28, 345-355.

Monsen, R. B. (1981). A usable test for the speech intelligibility of deaf talkers. American Annals of the Deaf, 126, 845-852.

Monsen, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. Journal of Speech and Hearing Disorders, 48, 286-296.

Mowe, K. (1997). Comparison of intelligibility estimation and orthographic transcription methods by preprofessional speech-language pathologists. Unpublished master's thesis, Portland State University, Portland, OR.

Morris, S. R., Wilcox, K. A., & Schooling, T. L. (1995). The preschool speech intelligibility measure. American Journal of Speech-Language Pathology, 4, 22-28.

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). Philadelphia, PA: John Benjamins Publishing.

Rafaat, S. K., Rvachew, S., & Russell, R. S. C. (1995). Reliability of clinician judgments of severity of phonological impairment. American Journal of Speech-Language Pathology, 4, 39-45.

Rice, M. L., & Wilcox, K. A. (1995). Building a language-focused curriculum for the preschool classroom, Volume I: A foundation for lifelong communication. Baltimore, MD: Brooks Publishing Company.

Rylant, C., & Gammell, S. (1985). The relatives came. New York: Bradbury Press.

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.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), Intelligibility in speech disorders (pp. 11-34). Philadelphia, PA: John Benjamins Publishing.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. Journal of Speech and Hearing Disorders, 47, 256-270.

Weiss, C. E. (1982). Weiss intelligibility test. Tigard, OR: CC Publications.

Weston, A. D., & Shriberg, L. D. (1992). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. Journal of Speech and Hearing Research, 35, 1316-1332.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. Journal of Communications Disorders, 11, 499-512.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. (1996). Comprehension of dysarthric speech: Implications for assessment and treatment planning. American Journal of Speech-Language Pathology, 5, 55-66.

Young, M. A. (1969). Observer agreement:
Cumulative effects of rating many samples. Journal of
Speech and Hearing Research, 12, 135-143.

Appendix A
Informed Consent

I, _____, consent to serve as a listener in this research project on measuring speech intelligibility of children.

I understand that the study involves giving a percentage of speech intelligibility to 12 speech samples. The study will require a gross estimation procedure and an orthographic transcription procedure to assess speech intelligibility. Both procedures result in a score that yields a percentage of intelligibility. It will take a total of approximately 4 hours to listen to all 12 tapes.

I understand that participation in this study will present no physical, social, economic, or other risks except for the possible inconvenience of coming to a central location to participate in the study. All data obtained during the course of the study will remain confidential. Published data and public records will not reveal my name.

It has been explained to me that the purpose of the study is to learn if the estimation procedure is an accurate and reliable measurement of speech intelligibility. I may not receive any direct benefit from taking part in this study, but my participation may help to increase knowledge that may benefit others in the future.

Heather Clarke has offered to answer any questions I may have about the study and what is expected of me in the study. I understand that I am free to withdraw from participation in this study at any time without jeopardizing my relationship with Portland State University.

I have read and understand the foregoing information and agree to participate in this study.

Date: _____ Signature: _____

If you experience problems that are the result of your participation in this study, please contact the Chair of the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 105 Neuberger Hall, Portland State University, (503) 725-3417.

Appendix B

Listening Instructions

Listener name: _____ Listener #: _____
 Number of years experience as SLP in public schools _____

You will be listening to 15 connected speech samples. At the end of each sample, you will have approximately 1 minute to assess what you have heard, and assign a percentage of intelligibility between 0% and 100% for that speaker. You are required to make a gross estimation based purely on what you have heard. **No tracking, notation or transcription of any kind is allowed.** Put your estimated percentage of intelligibility in the space provided which corresponds with the speech sample presented.

Your name will not be used in any publications or public files related to this study. You will be referred to by listener number only. Do you have any questions?

Sample 1	_____	%
Sample 2	_____	%
Sample 3	_____	%
Sample 4	_____	%
Sample 5	_____	%
Sample 6	_____	%
Sample 7	_____	%
Sample 8	_____	%
Sample 9	_____	%
Sample 10	_____	%
Sample 11	_____	%
Sample 12	_____	%
Sample 13	_____	%
Sample 14	_____	%
Sample 15	_____	%

Appendix C

Orthographic Transcription Instructions

You will be listening to 12 connected speech samples. On a separate page for each sample, write down the subject # and sample # in the blanks at the top of each page. Listen to each utterance and write down or type the words you think the child said. If you do not understand a word, please put an X where the word should be. If you understand part of a word, write that part of the word down along with an X (i.e., Xing). You do not need to write down fillers, such as um, mm-mm, uh huh, etc. You are encouraged to guess the words said. **You may listen to each utterance a maximum of three times.** If you have any questions, please call Heather Clarke at (360) 574-8522. Please notify Heather when the transcriptions have been completed so that arrangements may be made to have them picked up from you.

Appendix D

Raw Data by ListenerPercentage of Words Understood in Connected Speech
Gross Estimation Procedure

S#	Spk.#	L1	L2	L3	L4	Mean	Median	Range
1	36	95	100	95	90	95	95	10
2	10	70	80	85	75	78	78	15
3	38	60	90	65	60	69	63	30
4	9	40	65	50	50	51	50	25
5	13	55	50	65	65	59	60	15
6	16	30	25	35	40	33	33	15
7	14	30	20	50	70	43	40	50
8	8	20	5	20	40	21	20	35
9	33	60	100	80	80	80	80	40
10	17	45	85	60	80	67	70	40
11	44	70	95	75	75	78	75	25
12	43	15	2	30	30	19	23	28
Mean						58	57	27

Note: S = Sample; Spk. = Speaker; L = Listener.

Appendix E

Raw Data by ListenerPercentage of Words Understood in Connected Speech
Orthographic Transcription Procedure

S#	Spk.#	L1	L2	L3	L4	Mean	Median	Range
1	36	96	83	97	87	91	92	14
2	10	77	87	71	79	79	78	16
3	38	93	87	86	69	84	87	24
4	9	60	56	64	67	62	62	11
5	13	47	47	67	67	57	57	20
6	16	56	38	36	47	33	43	15
7	14	58	59	81	75	44	67	23
8	8	33	31	32	44	35	33	13
9	33	73	65	62	66	67	66	11
10	17	62	67	68	76	68	68	14
11	44	69	63	74	77	71	72	14
12	43	17	20	22	33	23	21	16
Mean						60	62	16

Note: S = Sample; Spk. = Speaker; L = Listener.

Listener Estimations for Samples Rated Twice

Spk.#	Listener 1		Listener 2		Listener 3		Listener 4	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
	Difference		Difference		Difference		Difference	
16	30	30	25	65	35	50	40	50
	0		40		15		10	
43	15	25	2	5	30	25	30	30
	10		3		5		0	
33	60	70	100	100	80	85	80	95
	10		0		5		15	
Mean Difference								
	6.66		14.33		8.33		8.33	

Note: Spk. = Speaker; 1st = the first presentation of the connected speech samples for gross estimation of speech intelligibility; 2nd = the second presentation of the connected speech samples for gross estimation of speech intelligibility; Difference = difference between the 1st gross estimation and the 2nd gross estimation.

Raw Data by Listener

Percentage of Words Understood in Connected Speech
Comparison of Gross Estimation and
Orthographic Transcription Procedures

S#	Spk.#	Listener 1 Gross Ortho Difference		Listener 2 Gross Ortho Difference		Listener 3 Gross Ortho Difference		Listener 4 Gross Ortho Difference	
1	36	95	96	100	83	95	97	90	87
		+1		-17		+2		-3	
2	10	70	77	80	87	85	71	75	79
		+7		+7		-14		+4	
3	38	60	93	90	87	65	86	60	69
		+33		-3		+21		+9	
4	9	40	60	65	56	50	64	50	67
		+20		-9		+14		+17	
5	13	55	47	50	47	65	67	65	67
		-8		-3		+2		+2	
6	16	30	56	25	38	35	36	40	47
		+26		+13		+1		+7	
7	14	30	58	20	59	50	81	70	75
		+28		+39		+31		+5	
8	8	20	33	5	31	20	32	40	44
		+13		+26		+12		+4	
9	33	60	73	100	65	80	62	80	66
		+13		+35		-18		-14	
10	17	45	62	85	67	60	68	80	76
		+17		-18		+8		-4	
11	44	70	69	95	63	75	74	75	77
		-1		-32		-1		+2	
12	43	15	17	2	20	30	22	30	33
		+2		+18		-8		+3	
Mean Difference									
		14		18		11		6	

Note: S = Sample; Spk. = Speaker; Gross = Gross Estimation Procedure; Ortho = Orthographic Transcription Procedure.

Appendix H

Raw Data by Listener in Random Speaker Order

Percentage of Words Understood in Connected Speech
Orthographic Transcription Procedure

S#	Listener 1		Listener 2		Listener 3		Listener 4	
	Spk.#	%	Sub.#	%	Sub.#	%	Sub.#	%
1	36	96	13	47	17	68	33	66
2	8	33	38	87	36	97	44	77
3	44	69	17	67	10	71	38	69
4	13	47	36	83	33	62	13	67
5	16	56	43	20	14	81	16	47
6	10	77	14	59	44	74	43	33
7	14	58	44	63	43	22	8	44
8	38	93	16	38	8	32	9	67
9	9	60	8	31	13	67	17	76
10	17	62	10	87	38	86	10	79
11	43	17	9	56	9	64	14	75
12	33	73	33	65	16	36	36	87

Note: S = Sample; Spk. = Speaker; % = Percentage of Words Understood.