6-10-1996

A Study of the Correlation between the Articulation Competence Index (ACI) and the Percentage of Words Understood in the Continuous Speech of 4- and 5-year-olds of Varying Phonological Competence

Susan Coll Mitchell
Portland State University

10.15760/etd.7042

THESIS APPROVAL

The abstract and thesis of Susan Coll Mitchell for the Master of Science in Speech Communication: Speech and Hearing Science were presented June 10, 1996, and accepted by the thesis committee and the department.

COMMITTEE APPROVALS: [redacted]

Mary Gordon-Brannan, Chair

[redacted]

Lisa Letcher-Glembo

[redacted]

Marjorie Terdal
Representative of the Office of Graduate Studies

DEPARTMENT APPROVAL: [redacted]

Rhea Paul, Chair
Department of Speech Communication

****

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by [redacted] on 21 August 1996

ABSTRACT

An abstract of the thesis of Susan Coll Mitchell for the Master of Science in Speech Communication: Speech and Hearing Science, presented June 10, 1996.

Title: A Study of the Correlation Between the Articulation Competence Index (ACI) and the Percentage of Words Understood in the Continuous Speech of 4- and 5-Year-Olds of Varying Phonological Competence

Intelligibility refers to how recognizable a speaker's words are to the listener. Severity, a broader but closely related concept, incorporates intelligibility, disability, and handicap. Many factors influence intelligibility, including speech sound production, voice, and prosody, as well as a number of linguistic and contextual factors.

Clinicians and researchers in the field of speech-language pathology require accurate measures of intelligibility and severity to assess and describe communicative functioning and to measure change over time. Determining the most accurate and efficient measurement

approaches has been the focus of recent attention in the field.

This study was a preliminary investigation of the relationship between the Articulation Competence Index (ACI), a severity metric, and the percentage of words understood in continuous speech, the standard measure of intelligibility. Specifically, the study addressed the research question:

Is there a significant correlation between the Articulation Competence Index (ACI) and percentage of words understood in samples of continuous speech of 4- and 5-year-olds with varying levels of phonological competence?

Subjects were thirty 4- and 5-year-olds from the Portland metropolitan area. Four listeners calculated percentage-of-words scores for each child's 100-word speech sample. These scores were compared to ACI scores calculated by the investigator for each of the samples.

The data were analyzed using the Pearson product-moment correlation (Pearson $\underline{r}$). A moderately strong correlation ($\underline{r}$ = .71 to .81) was found between the ACI and percentage of words understood. Squaring the correlation coefficients resulted in values for $\underline{r}^2$ of .50 to .66, indicating that the ACI accounts for more than half the variability of continuous speech intelligibility.

The results suggest that the ACI does reflect the intelligibility component of severity. However, concerns regarding methodology of this study, specifically the limited number of samples used in examining intra- and inter-rater reliability, should be considered when evaluating the results.

A STUDY OF THE CORRELATION BETWEEN THE ARTICULATION COMPETENCE INDEX (ACI) AND THE PERCENTAGE OF WORDS UNDERSTOOD IN THE CONTINUOUS SPEECH OF 4- AND 5-YEAR-OLDS OF VARYING PHONOLOGICAL COMPETENCE

by

SUSAN COLL MITCHELL

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
1996

ACKNOWLEDGEMENTS

I would like to express my appreciation to my thesis advisor, Mary Gordon-Brannan, especially for her flexibility and willingness to accommodate my hectic schedule. I would also like to thank the other members of my thesis committee, Lisa Letcher-Glembo and Marjorie Terdal, for the suggestions that helped shape my work into its final form. Thanks, also, to Mary T. Withers and Ellen Rueller, who helped so much with the logistics of my project (i.e., letting me in and out of the lab!). Very special thanks go to Traci Giacchero, who gave so many hours of her time and made the listening task fun.

I also wish to express my sincerest thanks to my family; each of you helped me in your own, unique way. My daughter, Rosemary, with her supportive phone calls and handwritten notes of encouragement, is a most considerate and compassionate young woman. My son, Ocie Scott, who took time from his own graduate studies to advise me on statistical matters, is a wise and patient teacher. And, most of all, I thank my husband, Ocie, who helped me with many small details, and whose support and good will are responsible for enabling our whole family to achieve our educational goals. Many, many thanks.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION AND STATEMENT OF PURPOSE

Introduction

Intelligibility refers to how recognizable a speaker's words are to the listener. It can be said that intelligibility of the spoken message is largely responsible for the effectiveness of an individual's verbal communication. Many factors influence intelligibility, such as loudness of the vocal signal, rate of speech, and intonation. Prosody, linguistic complexity, and the presence of articulation errors also can affect the intelligibility of the spoken message. The many influences on intelligibility contribute to the difficulty of quantifying the concept.

Because increased intelligibility is often the goal of intervention for speech, clinicians require accurate measures of intelligibility levels to describe this important aspect of communication, to prepare profiles of clients' communicative functioning, to establish the need for intervention for speech, and/or to measure change over the course of treatment. Investigators in the field of speech and language research require valid ways of measuring intelligibility in order to operationalize the construct of intelligibility for purposes of research and reporting.

A number of ratings, scales, and other measures of intelligibility have been proposed and employed for clinical and research purposes (Kent, Miolo, & Bloedel, 1994). Some of these gauge overall intelligibility based on subjective ratings or impressions of listeners (Kent, 1992). Others address particular aspects of speech that contribute to intelligibility or the lack of it, such as vocal quality, fluency, or the presence of phonological deviations (Hodson & Paden, 1981; Shriberg & Kwiatkowski, 1982; Weiss, 1982). Determining the efficiency and accuracy of these approaches has been the focus of recent attention in the field.

In 1982, Shriberg and Kwiatkowski proposed Percentage of Consonants Correct (PCC) to measure the degree of severity of children's speech disorders. They explored the validity of this measure by comparing PCC scores and a) an ordinal rating system based on severity of phonological involvement, and b) intelligibility measured in percentage of words understood in samples of continuous speech. The results of these investigations indicated a significant positive correlation between PCC and the rating system, but only a moderate correlation between PCC and percentage of words understood (Shriberg & Kwiatkowski, 1982).

In 1993, Shriberg proposed the Articulation Competence Index (ACI), which adjusts PCC to reflect the relative proportion of distortions, errors more common beyond the preschool years, to quantify more accurately the severity of involvement over a lifetime in speakers with developmental phonological disorders. Shriberg and Kwiatkowski (1982) defined severity as a concept incorporating intelligibility, disability, and handicap. One test of the validity of the ACI would be to examine how closely ACI scores are correlated with scores derived from procedures that measure these components of severity.

In addition to the previously cited work by Shriberg and Kwiatkowski (1982), a number of other authors have examined intelligibility measured in percentage of words understood in continuous speech by unfamiliar listeners, and have judged this to be the standard against which other methods of describing intelligibility can be measured for accuracy (Bernthal & Bankson, 1993; Gordon-Brannan, 1993; Kwiatkowski & Shriberg, 1992). The degree to which the ACI correlates with this standard has not yet been examined in the literature.

Purpose

The purpose of this study was to investigate the relationship between the Articulation Competence Index (ACI) and percentage of words understood in continuous speech. The study was, in its design, a limited and preliminary study, and as such it will contribute limited and preliminary information regarding how well the ACI reflects the intelligibility component of severity. It is hoped that the experiences of the investigator will provide direction for more comprehensive and definitive efforts in determining the validity and utility of the ACI.

Specifically, this study addressed the question of how closely the ACI is correlated with percentage of words understood in continuous speech by unfamiliar listeners, as a standard for measuring intelligibility in 4- and 5-year-olds with varying levels of phonological competence. The research question posed was:

Is there a significant correlation between the Articulation Competence Index (ACI) and percentage of words understood in samples of continuous speech of 4- and 5-year-olds with varying levels of phonological competence?

The research question was formulated as the null hypothesis:

There is not a significant correlation between the Articulation Competence Index (ACI) and percentage of words understood in the continuous speech of 4- and 5-year-olds with varying levels of phonological competence.

CHAPTER II

REVIEW OF THE LITERATURE

This study explored the relationship between the Articulation Competence Index (ACI) and percentage of words understood in samples of continuous speech of 4- and 5-year-olds. A review of relevant literature will address definitions of the term "intelligibility" and the significance of the concept. The conceptualization will be further developed through an examination of factors that contribute to intelligibility of the spoken message. This will be followed by an examination of procedures for measuring intelligibility, particularly in preschool populations. The literature review will conclude with a discussion of the importance of investigating the validity intelligibility measures for clinical and research purposes.

Defining Intelligibility

Gordon-Brannan (1993) defined intelligibility as "the degree to which a person's speech is understood by a listener" (p.7). Hodson and Paden (1981) described unintelligible children as those who experience "extreme difficulties in making themselves understood" (p. 370).

In their text, Bernthal and Bankson (1993) addressed intelligibility as the predominant measure of the efficiency of an individual's competence in the use of speech.

Kent, Miolo, and Bloedel (1994) indicated their conviction regarding the importance of intelligibility and maintained general agreement with Subtelny's (1977) contention that "Intelligibility is considered the most practical single index to apply in assessing competence in oral communication" (p. 183). However, Kent et al. indicated a corresponding lack of agreement regarding how intelligibility should be measured. In this, they appeared to agree with Gordon-Brannan (1993, 1994), who offered the concise definition of intelligibility cited above, but noted the difficulty of defining the term operationally.

Severity, a measure of the degree to which a person's speech differs from that of adults in the linguistic community, is a concept closely related to intelligibility (Billman, 1986). In defining and measuring intelligibility, it is important both to recognize the similarity and to maintain the distinction between the two terms. Shriberg and Kwiatkowski (1982) and Gordon-Brannan (1993) clarified the distinction in noting that severity is the more general term that incorporates intelligibility.

Related Factors

Though speech sound errors often are viewed as major determinants of intelligible speech, a wide range of factors also have been considered as potential influences on intelligibility. Table 1 provides a comprehensive selection of these items, originally listed in an optional portion of the Weiss Intelligibility Test (Weiss, 1982).

Table 1

Factors that Influence Intelligibility (Weiss, 1982)

Adventitious sounds	Pronunciation
Articulation	Rate
Communicative disfluency	Redundancy
Inflection	Resonation
Juncture	Rhythm
Mean length of utterance	Semantics
Morphology	Stress
Morphophonemics	Syntax
Pauses	Voice quality
Physical Posture	Intensity
Pitch	Pragmatics

Most of these factors, either individually or collectively (as part of a superordinate category, such as "suprasegmentals," "voice," or "prosody"), have been investigated to determine their effect on or association with intelligibility (Hodson & Paden, 1981; Kwiatkowski & Shriberg, 1992; Shriberg, 1993; Weston & Shriberg, 1992). Brief descriptions of these investigations follow.

Contextual and Linguistic Factors

Weston and Shriberg (1992) conducted two studies that revealed positive associations between intelligibility and a number of contextual and linguistic variables, including utterance length, fluency, phonological complexity, and grammatical form. They also found that the position of a word within an utterance, as well as word position relative to other unintelligible words, is associated with intelligibility. The authors concluded that articulatory elements alone cannot provide a complete explanation for lack of intelligibility, and should not be the exclusive focus in intelligibility assessment.

The speaker-listener dyad. Not all factors contributing to intelligibility are associated entirely with the speaker. Some authors have stressed the importance of considering the speaker-listener dyad in both defining and measuring intelligibility (Connolly, 1986; Kent, 1993; Kent et al., 1994; Weston & Shriberg, 1992). They have noted that failure to receive a spoken message may result from the listener's inability to decode the message or extract cognitive meaning from it for one or more reasons, including lack of familiarity with the speaker or the topic, or lack of listener attention at the moment of the communicative attempt. Connolly (1986) referred to the listener's inability to discern the intended meaning of a statement as "indeterminability," and maintained that intelligibility was one element of this broader concept (p. 372). Kent (1993) agreed that viewing intelligibility solely as an attribute of the speaker or the message is a narrow conceptualization which is "always incorrect" (p. 225).

Predictability. Kent (1993) noted an important element that arises in the speaker-listener dyad when the speaker is a child. When a speaker uses the conventional adult speech and language patterns of a common linguistic community, the listener can employ predictive strategies to enhance perception of the spoken message. Children's verbal productions tend to be more variable than those of adults. This increased variability has a detrimental effect on predictability which can make the listener's task more difficult, thus potentially compromising intelligibility.

Articulation and Phonological Factors

While stressing the variety of social and linguistic elements affecting the speaker-listener dyad, Connolly (1986) and Kent (1993) also acknowledged that articulatory and phonological competence of the speaker are important elements in determining intelligibility in the limited sense, and they cited a number of articulatory and phonological factors that can affect success in producing a spoken message. Among these factors were lack of phonological contrasts, degree of distance between a target and its actual production, and the frequency and consistency of the speech sound error (Connolly, 1986; Kent et al., 1994).

While speech sound production is regarded as an important component of intelligibility, authors have reported that articulation errors and phonologic deviations influence intelligibility to various degrees, depending on the type, but not necessarily the frequency, of the error or deviation (Kent, 1992; Shriberg, 1993). For example, Kent et al. (1994) noted that an individual with a pervasive lisp can be quite intelligible, and that a speaker with a phonologic disorder also can be highly intelligible to listeners familiar with the particular phonological pattern.

Hodson and Paden (1981) studied phonological processes present in the speech of unintelligible 4-year-olds and normally developing children of the same age. The study revealed that intelligible and unintelligible children could be distinguished by use of specific phonological patterns. For example, in attempting to produce "th," children who were more intelligible usually substituted other continuant sounds (e.g., /f/, /s/, /v/, or /z/), while children who were less intelligible substituted /t/ or /d/, indicating use of a stopping process. The authors concluded that use of particular phonological patterns by each group indicated differing strategies for dealing with phoneme classes, and noted that specific strategies were associated with the overall intelligibility of the individuals who used them. The phonological processes most often used by unintelligible children were cluster reduction, stridency deletion, stopping, final consonant deletion, fronting, backing, syllable reduction, prevocalic voicing, and glottal replacement. The authors did not indicate the relative degree to which each of these processes contributed to reduced intelligibility. However, Billman (1986) reported that, for children in a similar study, backing and prevocalic singleton omission had the greatest negative impact on intelligibility, and that liquid deviations, while common, were not significantly correlated with intelligibility.

Intelligibility Measures

The many factors influencing intelligibility are equaled by the variety of approaches to measuring intelligibility levels. In this section, selected measures will be reviewed, with emphasis on those measures examined in this study, namely percentage of words understood in continuous speech as well as the Articulation Competence Index (ACI) with its component, Percentage of Consonants Correct (PCC). The ACI and PCC assess accuracy of phoneme production, an important factor in intelligibility. However, it should be noted that PCC and the ACI may be more accurately described as measures of severity of involvement than intelligibility per se. Kent et al. (1994) justified including PCC and ACI in their comprehensive review of intelligibility measures because of the close relationship between severity of articulatory involvement and intelligibility, and the reliance of PCC and ACI on phonemic factors. This review also will describe additional measures that have been compared to percentage of words understood and PCC in other research.

Scales and Ratings

Two commonly used methods, equal-appearing interval scales and direct magnitude estimation (DME), both involve evaluating word-, sentence-, or conversation-level speech samples by assigning a number to indicate the level of intelligibility or severity perceived by a listener. Interval scaling procedures represent intelligibility as a continuum, while DME rates intelligibility relative to a selected standard. A number of authors have noted problems regarding the psychometric properties of such procedures (Gordon-Brannan, 1993; Kent, 1992; and Kent et. al, 1994; Schiavetti, 1992). Despite these shortcomings, Connolly (1986) offered that rating scales might be the only practical measurement tool with highly unintelligible speakers whose utterances cannot be sufficiently glossed so that targets can be identified.

Percentage of Words Understood

In their text, Bernthal and Bankson (1993) stressed the value of intelligibility data derived from samples of connected speech. They recommended that, since accurate speech sound production in conversation is the goal of phonological intervention, evaluation of these productions in continuous speech should be a component of any evaluation. Kwiatkowski and Shriberg (1992) concluded that valid assessment of intelligibility must be based on scores derived from samples of continuous speech in order for such assessment to reflect the interaction of factors related to language, speech, voice and prosody.

One subtest of the Weiss Intelligibility Test involves calculating the percentage of intelligible words in a 200-word sample of contextual speech, which is averaged with percentage of intelligible single-word productions to yield the overall intelligibility score

(Weiss, 1982). Other authors have used percentage of words understood in samples of spontaneous speech as a standard for comparison to other intelligibility or severity-level measurements. For example, Gordon-Brannan (1993) found that the percentage-of-words measure was highly correlated with four other measures: (a) percentage of imitated single words understood, (b) percentage of words understood in imitated sentences, (c) listener ratings of intelligibility, using a 7-point scale, and (d) Phonological Deviation Average (PDA).

Phonological Deviation Average (PDA)

The Phonological Deviation Average (PDA), also referred to as the Phonological Deviation Score (PDS), is derived from phonological deviation scores yielded from the Assessment of Phonological Processes-Revised (APP-R; Hodson, 1986). Administration of this instrument involves elicitation and narrow phonetic transcription of 50 spontaneous single-word or short utterances as the child names objects or pictures. The stimuli contain all the American English phonemes, including consonant sequences. The child's productions are analyzed, and an average of occurrence of 10 basic phonological deviations is computed. This average is used to assign a severity level, using a formula that also takes into account the child's chronological age. Garrett and Moran (1992) found that PDS was highly intercorrelated with four other measures: (a) percentage of consonants correct in single words, (b) percentage of consonants correct in connected speech, (c) perceptual ratings by untrained listeners, and (d) perceptual ratings by graduate students in speech-language pathology. Gordon-Brannan (1993) found that PDA was one of four measures highly correlated with intelligibility expressed as percentage of words understood in samples of continuous speech.

Percentage of Consonants Correct (PCC)

Shriberg and Kwiatkowski (1982) proposed Percentage of Consonants Correct (PCC) as a measure of severity of involvement, encompassing disability, intelligibility, and handicap. To calculate PCC, numbers of incorrectly and correctly articulated consonants in 1-minute samples of continuous speech are counted, and a percentage is derived. Based on this percentage, a severity level, ranging from mild to severe is assigned. Sampling and scoring rules for determining PCC are provided in Appendix A.

In a study involving sixty 3- to 9-year-old children with developmental phonological delays, PCC scores were only moderately correlated ($\underline{r} = .42$; $\underline{r}^2 = 18\%$) with intelligibility measured as percentage of words understood (Shriberg & Kwiatkowski, 1982). However, intelligibility and PCC were more highly correlated with severity ratings than eight other variables: loudness, (vocal) quality, phrasing, stress, rate, age, sex, and average words per utterance (Shriberg & Kwiatkowski, 1982). Though PCC analyzes speech sound productions at the phoneme level, it requires only a determination that consonants are either produced correctly or incorrectly, and does not analyze the nature of the consonant errors present in a speech sample.

The Articulation Competence Index (ACI)

Shriberg (1993) noted that "single, specific articulatory distortion errors" are the type of speech sound error usually seen in older children and adults (p. 106). He proposed the ACI, which adjusts PCC scores in favor of consonant distortions, as a better device than PCC for testing individuals' articulatory competence repeatedly over a lifetime. Criteria for scoring distortion errors for purposes of calculating the ACI are provided in Appendix B. The ACI metric is based on PCC and the Relative Distortion Index (RDI), and, like PCC, the resulting score is used to assign a severity level. The RDI is a percentage calculated by dividing the total distortion errors in a 1-minute sample of continuous speech by the total number of consonant errors (including distortions) in the sample. The ACI percentage is derived through the following formula:

$$ACI = \frac{PCC + RDI}{2}$$

It is important to note that a particular scoring exception is utilized when PCC scores are 95% or higher. In such cases, ACI scores would be inordinately low, unless errors were in the form of distortions. For example, an individual who correctly produced 95% of the target consonant sounds and did not make any distortion errors would receive an ACI score of only 47.5, while an individual who correctly produced only 80% of all consonants and made only distortion errors would receive an ACI score of 90%. To account for this discrepancy, Shriberg (1993) adopted the practice of using PCC scores in place of ACI scores in such cases. Therefore, an individual with a PCC score of 95% would have an ACI score of 95%, and an individual who correctly articulated all consonants in a 1-minute speech sample would receive an ACI score of 100%.

Summary

The varied approaches to evaluating and measuring intelligibility have been described and classified under the major headings of "impressionistic" and "quantitative" (Kent, 1993; Kent et al., 1994). Impressionistic statements, such as, "The client is highly unintelligible," clearly require subjective judgments regarding the intelligibility of a speaker. However, other approaches, such as rating scales or percentage of intelligible words, though somewhat more objective, also rely to a degree on listener judgment in deriving numerical scores. This does not necessarily discount any one procedure, but it does make correlation of various methods critical, both in establishing their validity and in selecting measurement approaches that meet specific clinical or research needs.

Shriberg (1993) proposed the Articulation Competence Index (ACI) to measure the severity of speech disorders in individuals from two years of age through adulthood. Shriberg and Kwiatkowski (1982) defined severity as a concept that incorporates the constructs of intelligibility, disability, and handicap. It is important, therefore, to explore the validity of the ACI as a severity metric by examining the relationship between the ACI and other procedures that measure those components of severity.

Bernthal and Bankson (1993) and Weiss (1982) have considered intelligibility data obtained from samples of connected speech as highly valid. Gordon-Brannan (1993), Kwiatkowski and Shriberg (1992), and Shriberg and Kwiatkowski (1982) have compared percentage of words understood in samples of spontaneous speech to other intelligibility and severity metrics, and have supported the use of percentage-of-words scores to operationalize the construct of intelligibility in such comparisons.

Examining the degree of correlation between the ACI and the percentage of words understood in continuous speech would offer important information regarding the validity of the ACI as a measure of severity; however, such a correlation has not yet been investigated and reported in the literature.

CHAPTER III

METHOD

This study examined the relationship between the standard measure of intelligibility, percentage of words understood, and the Articulation Competence Index (ACI) (Shriberg, 1993), when used to evaluate the continuous speech of 4- and 5-year-olds with varying levels of phonological competence. Because of time considerations in this preliminary study, extensive use was made of data and speech samples collected as part of a previous study by Gordon-Brannan (1993), entitled, "Speech Intelligibility Assessment of Young Children with Varying Levels of Phonological Proficiency/Deficiency." Therefore, details regarding procedures utilized in that study will be discussed. For clarity, procedures from the Gordon-Brannan study will be classified in headings by the abbreviation G-B.

Subjects

Subjects for this study were 30 of the 48 children, with varying levels of intelligibility, recruited from preschools and speech-language pathology caseloads in the Portland, Oregon metropolitan, area, who participated in the doctoral research of Gordon-Brannan (1993). As the current study was a preliminary effort to examine the correlation between the ACI and percentage of words understood, it was decided to include only the minimum number of subjects required to lend sufficient power in statistical tests to be used in analyzing the resulting data. The decision to include only the minimum number of subjects was based on anticipation of the amount of time required to complete extensive training in use of the ACI and the complexity of the listening and scoring tasks.

Selection Criteria in the G-B Study

Subjects ranged in age from 4:0 (years:months) to 5:6 (mean = 4:7). They were selected from a group of 57 children, screened for hearing and receptive language deficits, using pure tone audiometry and the Test of Auditory Comprehension of Language-Revised (TACL-R) (Carrow-Woolfolk, 1985). The test manual for the TACL-R affirms that the test effectively differentiates persons who have language comprehension deficits from those who do not. Children who had hearing levels of 35 dB HL or better bilaterally and those who scored at the 10th percentile or above on the TACL-R were considered free of significant hearing and language problems. It should be noted that only three children had mild hearing losses, indicated by pure tone averages no higher than 35 dB HL bilaterally, while the remainder passed a hearing screening at 20 dB HL. Parent report and direct observation by the investigator were used to identify and exclude potential subjects with obvious neurological, motor, and/or laryngeal or resonance deviancy that could affect speech (Gordon-Brannan, 1993).

At the conclusion of the study, Gordon-Brannan listed the subjects, identified by previously assigned subject numbers, in descending order reflecting degree of intelligibility, as measured by percentage of words understood in continuous speech. The resulting list, therefore, represented a continuum of intelligibility levels. This listing was divided into 4 groups of 12 subjects each, with the first group containing the 12 most intelligible subjects, the second group containing the next most intelligible, and so forth.

Subject Selection for the Current Study

The 30 subjects for this study were selected from Gordon-Brannan's (1993) list of 48 subjects, through stratified random sampling. In this process, 7 subjects from each of the intelligibility levels were randomly selected, with the remaining 2 subjects selected at random from the entire listing.

Procedures

G-B Sample Collection

Gordon-Brannan taped 100-word continuous speech samples from each of the subjects, using picture cards and a children's book, The Relatives Came (Rylant & Gammell, 1985), to elicit conversation. Though the elicitor's utterances also were recorded, care was taken not to make comments that would serve as hints as to the content of a child's speech.

Instrumentation

G-B instrumentation. The samples for each child were recorded in an acoustically treated room. A Sharp SX D200 digital audiotape recorder and an AKG, Model C451, capacitor flat microphone were used to make the recordings. The investigator and subject were seated at a cloth-covered table with the microphone placed on foam or in a microphone stand set on the table, approximately 6" from the subject's mouth. The children's caregivers were given the option of remaining in the room while the speech samples were obtained. A Panasonic camcorder, VHS Reporter, Ag-10 was used to make video recordings for subsequent viewing by the caregivers, should their assistance be required in glossing the samples at a later time. The 100-word continuous speech samples were later dubbed in random order onto digital and analog audiotapes.

In listening to the recorded speech samples to determine the percentage of words understood, listeners played back the analog tapes at home on their own analog tape recorders of various models.

Instrumentation in the current study. Because the listening task to determine ACI scores requires finer discrimination, listening sessions to collect the ACI data were conducted in a closed room, using a Denon digital audiotape recorder (Model DTR-80P) connected to a Sony table-top speaker (Model SRS-150) to play back the digital tapes.

Transcripts

Orthographic transcriptions of the 100-word speech samples were prepared by Gordon-Brannan and research assistants. A parent or caregiver of the more unintelligible children attempted to provide verification of the gloss from the videotaped and/or audiotaped recordings. Words that remained unintelligible were represented in the transcriptions by an X or a blank line. The completed transcripts were used as scoring keys for calculating percentage of words understood in the Gordon-Brannan study as well as for calculating the ACI scores in the current study. A sample of a portion of one transcript appears in Appendix C.

Scoring

Calculating percentage of words understood in the G-B study. Four speech-language pathology graduate students at Portland State University inspected the stimulus materials, listened to the tapes, and made orthographic transcriptions of the continuous speech samples. Each listener's orthographic transcription of a speech sample was compared to the scoring key for that sample. Gordon-Brannan calculated the percentage of words understood by each listener for each subject, following the method outlined by Kwiatkowski and Shriberg (1992), that is, dividing the total number of words understood in a continuous speech sample by the total number of words in the sample. Scores determined by each of the four listeners for each of the subjects are included in Appendix D.

Calculating the ACI for the current study. As training in the listening and scoring task, the investigator reviewed the criteria for determining the ACI and for distinguishing distortions from other consonant errors (Shriberg, 1993). These criteria are provided in Appendixes A and B. It is important to note that sounds that are not standard productions of target phonemes, but are not recognized as another distinct phoneme (e.g., dentalized sibilants) are categorized as distortions, as are "all potential additions" (Shriberg, 1993, p. 132).

In training, special attention was given to distinguishing clinical from non-clinical distortions, as the latter are not considered distortion errors for purposes of calculating the ACI. Non-clinical distortions include: a) palatalized /s/ [/s̡/], b) retroflexed /s/ [/s̯/], c) deletion of initial /h/ in unstressed pronouns, and d) substitution of a glottal stop for /t/ in word-final position (Shriberg, 1993).

The investigator also practiced scoring samples, using transcripts and speech samples of subjects from the original Gordon-Brannan (1993) study not selected to be included in the current investigation. After completing 3 to 4 hours of training and practice, the investigator listened to the 30 samples included in this study and calculated the ACI for each. In accordance with the ACI scoring rules (Shriberg, 1993), words that were unintelligible to the investigator were not scored, even though the gloss of these words might have been provided on the scoring keys.

To score the speech samples, the investigator listened to the taped continuous language samples and recorded consonant errors on a copy of the scoring key for each sample. Each consonant articulation error was indicated by marking a diagonal line across the letter representing the target sound. When the error was a distortion, a second diagonal line was drawn intersecting the first, forming an X. Vowel distortions (e.g., derhotacized /ɝ/ or /ɚ/; notably raised, lowered, fronted, or backed vowels or diphthongs; and/or vowels of notably lengthened or shortened duration) were indicated by circling the orthographic representation of the vowel. A portion of a marked sample is provided in Appendix C. The marks were counted at a later time, and the ACI was calculated, following the procedure formulated by Shriberg (1993).

$$ACI = \frac{PCC + RDI}{2}$$

The variables in the formula for ACI are defined as follows:

$$PCC = \frac{\text{TOTAL CORRECTLY ARTICULATED CONSONANTS}}{\text{TOTAL CONSONANTS}}$$

$$RDI = \frac{\text{ERRORS DUE TO DISTORTIONS}}{\text{TOTAL ARTICULATION ERRORS}}$$

Reliability

Percentage-of-words scoring in the G-B study.

Inter-rater reliability was established between the four listeners from the original Gordon-Brannan (1993) study through the Pearson product-moment correlation (Pearson *r*). Because the listeners were permitted to listen to each speech sample as many as three times, intra-judge reliability was not determined.

<u>The Articulation Competence Index (ACI) scoring</u>. Inter-rater and intra-rater reliability in assigning the Articulation Competence scores were addressed through the Pearson product-moment correlation (Pearson *r*). Because it was anticipated that the listening and scoring tasks were likely to require considerable time, only 20% of the samples were used in correlations to assess reliability, though it was understood that the small sample size would reduce the power of the statistical outcome, yielding limited or questionable results.

The investigator scored 6 of the samples twice, with the second presentation of those samples occurring at least 24 hours after the first. A comparison of the two sets of scores for the six samples was used as a measure intra-judge reliability. Another graduate student nearing completion of the Speech and Hearing Sciences Program at Portland State University participated in training sessions in scoring and calculating the ACI. The student then listened to 6 of the speech samples and calculated the ACI for each. Inter-judge reliability was addressed by correlating the two sets of scores for the six samples calculated by the principal investigator and the second graduate student. In addition, an item analysis of the two sets of marked transcripts was conducted to further evaluate inter-judge reliability.

Analysis

All statistical analyses were performed using SPSS for Windows (SPSS, 1993). A confidence level of .05 was established for all statistical analyses.

The research question, regarding the relationship between percentage of words understood and the ACI, was addressed following a simple correlational design. After the computation of ACI for each of the 30 speech samples, the correlation between the independent variable, intelligibility as measured by percentage of words understood, and the dependent variable, the ACI, was calculated using the Pearson-r.

CHAPTER IV

RESULTS AND DISCUSSION

The purpose of this study was to investigate the degree to which the Articulation Competence Index (ACI) reflects intelligibility in the speech of 4- and 5-year-olds of varying phonological competence. This was addressed by correlating ACI scores calculated from recorded samples of continuous speech obtained from 30 children, with the percentage of words understood in those samples. Prior to presenting the results of the study, reliability data will be offered. The chapter will conclude with a discussion of the results of this study and an anecdotal account of the investigator's experiences in collecting the data, which provides information for consideration in designing more comprehensive studies to assess the validity of the ACI.

Reliability

<u>Reliability of Percentage of Words Understood Data in the G-B Study</u>

The Pearson product moment correlation (Pearson *r*) was used to examine inter-rater reliability among the four listeners who determined the percentage of words understood in each of the 30 samples. Reliability coefficients for the percentage-of-words scores ranged from .87 to .94. All individual correlations between each of the six pairs of listeners were significant ($p < .001$), indicating that the four listeners were generally in agreement in determining percentage of words understood. This level of agreement also suggests that the subjects were similarly intelligible to the four listeners. A correlation matrix for the percentage-of-words-understood measure is provided in Table 2. Because the listeners were permitted to listen to each sample as many as three times, intra-rater reliability was not determined.

Table 2

Correlation Matrix for Percentage of Words Understood

Listener	Listeners			
	A	B	C	D
A	1.00			
B	.88	1.00		
C	.87	.93	1.00	
D	.88	.93	.94	1.00

Note: $p < .001$

Reliability The Articulation Competence Index (ACI) Scoring

After completing training and practice in the Articulation Competence Index (ACI) scoring procedures, the investigator calculated the ACI for the 30 samples. Several days after the first scoring, the investigator listened to and re-scored 6 of the samples. At that time, a second graduate student, who had participated in the earlier training and practice sessions, also listened to and scored the same 6 samples.

Intra-rater reliability in ACI scoring. The Pearson *r* was used to examine intra-rater reliability between the first and second sets of ACI scores calculated by the investigator. The resulting coefficient (*r*) of .96 was significant ($p < .05$). While the sample size ($n = 6$) used in calculating the correlation limits the power of the resulting statistic and may give cause to question the validity of the procedure to address intra-rater reliability in this study, the strong correlation between the two sets of scores indicates that the investigator was consistent in scoring the samples.

Inter-rater reliability in ACI scoring. Inter-rater reliability was examined through the Pearson *r*, calculated from the two sets of ACI scores, that is, those assigned by the investigator and those calculated by the second graduate student. The two sets of ACI scores are provided in Appendix D. The resulting coefficient (*r*) of .94 was significant ($p < .05$). Because the correlation was determined based on a small sample size ($n = 6$), the power of the resulting statistic is reduced, which may limit the validity of the procedure to assess inter-rater reliability in this study. The outcome of $r = .94$ indicates that the two scorers were in close agreement in assigning ACI scores.

Results

The research question investigated was: Is there a significant correlation between the Articulation Competence Index (ACI) and percentage of words understood in samples of continuous speech of 4- and 5-year-olds with varying levels of phonological competence?

The research question was addressed through use of the Pearson r. Because this was a preliminary investigation of the validity of the ACI, a single scorer determined the ACI scores used in all correlations. The reader should be mindful of this element of the study design in evaluating the resulting correlations. The correlation coefficients for percentage of words scores assigned by each of four listeners and ACI scores calculated by the investigator are provided in Table 3. The coefficients (rs) ranged from .71 to .81. All individual correlations were significant ($p < .001$), indicating a moderately strong correlation between percentage of words scores and the ACI scores. Squaring the correlation coefficients to further assess the degree of relationship between the two measures yielded values for r^2 ranging from .50 to .66. These values indicate that the dependent variable, the ACI, accounts for more than half of the variability in intelligibility in continuous speech, as measured by percentage of words understood. Individual r^2 values are included in Table 3. Raw data in the form of percentage of words and ACI scores for each subject are provided in Appendix E.

Table 3

Correlation Matrix for the Articulation Competence Index (ACI) and Percentage-of-Words Understood Scores

		Percentage of Words Understood by Listeners			
	ACI	A	B	C	D
ACI (r)	1.00	.71	.76	.73	.81
ACI (r^2)	1.00	.50	.58	.53	.66

Note: ACI (r) indicates Pearson r correlation between the ACI scores and percentage-of-words scores. ACI (r^2) indicates squared values for the correlation between ACI scores and percentage-of-words scores.

Discussion

This study was a preliminary investigation of the correlation between the Articulation Competence Index (ACI) and the percentage of words understood in continuous speech samples of 4- and 5-year-old children of varying phonological competence. The percentage of words understood was used as the standard measure of intelligibility. The ACI is a measure of severity, which, according to Shriberg (1993), encompasses intelligibility, disability, and handicap. A correlation between scores for the two measures, when used to assess samples of continuous speech, indicates the degree to which the ACI reflects the intelligibility component of severity. The results demonstrate that there is a moderately strong relationship between the two measures and that the parameters of speech measured by the ACI account for more than half the variability in continuous speech intelligibility.

Reliability in Percentage of Words Understood in the G-B Study

Because listeners were permitted to listen to the speech samples three times, intra-rater reliability was not examined. Correlations to determine inter-rater reliability were moderately high, indicating that the four listeners were in general agreement in determining the percentage of words understood in each sample. These results also suggest that the subjects were similarly intelligible to the listeners.

To further examine the degree of agreement among the four listeners in assigning percentage-of-words scores, the relative range of scores for each subject was determined through an additional statistical operation. This involved

This involved dividing the standard deviation of the four percentage-of-words scores for each subject (the population standard deviation) by the mean of the four scores. Analysis of the resulting relative range of scores for each subject revealed that the four listeners differed by less than 5% in assigning scores for 15 of the subjects. For each of 10 subjects, the range differed by less than 10%, while scores for each of 4 subjects differed by less than 15%. The range of scores for the single remaining subject varied by 16%. This analysis suggests that, though the listeners were not in complete agreement in determining the actual percentage of words understood in each sample, they were not widely disparate in assigning percentage-of-words scores. The lack of extreme disparity supports the evidence provided by the correlation coefficients, which indicated that the listeners were in general agreement in determining percentage of words understood. That a degree of divergence in scores was observed is not surprising, however, given the complexity of the interaction between partners in the speaker-listener dyad, described by Connolly (1986) and Kent (1993).

Reliability in the ACI Scoring

Correlation coefficients for intra- and inter-rater reliability in ACI scoring were very high, indicating that agreement in determining ACI scores. Because only 6 of the 30 samples were included in calculating the correlations to assess intra- and inter-rater reliability in ACI scoring, the results are limited in statistical power. Because of this, the validity of the procedures to determine intra- and inter-rater reliability in ACI scoring may be subject to question. The ACI scores subjected to statistical analysis to determine intra- and inter-rater reliability were also examined through other means, including rank ordering and item analysis. While these methods do not increase the power of the statistical correlations, they do offer some additional information regarding similarities and discrepancies between sets of ACI scores determined in this study.

Intra-rater reliability in ACI scoring. The two sets of 6 samples included in intra-rater reliability testing were ranked in descending order by ACI scores assigned by the investigator in initial and second scoring sessions. The rank order was identical for both sets of scores, while differences between first and second sets of scores ranged from only 3 to 6 percentage points, indicating consistency in scoring. An item analysis of errors noted on the speech sample transcripts used in the first and second scorings did not reveal any pattern in discrepancies. However, including all 30 samples in repeat scoring might have revealed discrepancies. However, including all 30 samples in repeat scoring might have revealed patterns that were not observed in the limited selection of 6 samples.

Inter-rater reliability in ACI scoring. The investigator assigned higher ACI scores to 4 of the 6 samples than did the second scorer. However, no pair of scores differed by more than 10%. When the two sets of ACI scores determined by the investigator and the second scorer were ranked in descending order, there were differences in the first three places in the ranking, while there was total agreement in the last three places of the order. It should be remembered, however, that in the 3 sets of scores that were not ranked identically, just as in the entire 6 sets, no pair of scores differed by more than 10%. While the examination of the ranked scores shows close agreement, conclusions drawn from this examination are limited by the small number of samples.

Analysis of the two sets of scored samples revealed that the second scorer judged fewer consonants as correct on 4 of 6 samples, though only one set of scores differed in identification of consonant errors by more than 6%. The greatest discrepancy in scoring correct consonants was an 18% difference for Subject 2, whose resulting ACI scores of 30 and 28, assigned by the investigator and the second scorer, differed by 7%. Both scorers also ranked Subject 2 in last place by ACI scores.

While the numbers of errors judged as distortions by the two scorers did not reveal any pattern, one consistent discrepancy in scoring distortions was revealed by an item analysis of the scored transcripts. This discrepancy occurred in scoring /s/ distortions. In the 6 samples, the second scorer found 9 instances of /s/ distortions that were not scored as errors by the investigator, with 5 of these discrepancies occurring in the sample for Subject 1. Both scorers were aware that, according to the ACI scoring rules, palatalized /s/ (/s/) and retroflexed /s/ (/s/) are non-clinical distortions and, therefore, are not scored as errors. However, in informal discussion during training, the investigator observed that she tended to attribute other questionable /s/ productions to the recording quality, while the second scorer did not. In this area, it could be said that the investigator disregarded the scoring instruction to "score as incorrect unless heard as correct" (Shriberg & Kwiatkowski, 1982, p. 260). It is apparent that the two scorers maintained their respective approaches to scoring /s/ productions throughout the data collection process. Presenting all 30 samples for ACI scoring by additional scorers might have revealed more scoring patterns or discrepancies than were apparent in the 6 samples.

Correlation Between Percentage of Words Understood and the Articulation Competence Index (ACI)

Correlations between percentage of words understood by each of the four listeners and the Articulation Competence Index (ACI) scores assigned by the investigator were moderately high, indicating that the ACI does reflect intelligibility in continuous speech.

The correlation between the ACI scores and percentage of words understood (illustrated by the respective *r*-values of .71, .76, .73, and .81; and values for r^2 of .50, .58, .53, and .66) is higher than the moderate correlation (r = .42; r^2 = 18%) between Percentage of Consonants Correct (PCC) and percentage of intelligible words reported by Shriberg and Kwiatkowski (1982). This suggests that the ACI more accurately reflects the intelligibility component of severity than does PCC. This appears to be supported by a later study by Shriberg (1993), who concluded that the ACI was the more sensitive severity metric by demonstrating that the ACI provides better separation of speech-delayed 4- and 5-year-olds from speech-normal children of the same age than does PCC.

The results of this study in the form of a moderately high correlation between the ACI, a measure of the severity of articulatory or phonological involvement, and percentage of words understood in continuous speech, the standard measure of intelligibility, provide empirical support for researchers who have noted the importance of articulatory and phonological factors in intelligibility (Gordon-Brannan, 1993; Hodson & Paden, 1981; Kent, 1993; Weiss, 1982). In this study, the $\underline{r}^2$ values derived from the correlations between the ACI and percentage-of-words data show that, though articulatory and phonological factors account for more than 50% of the variability in speech intelligibility, 40 - 50% of the variability must be accounted for by other factors. The literature describes a number of influences on intelligibility that may, individually or in combination, constitute these factors. Weiss (1982) offered a comprehensive list of possible influences on intelligibility, and some of these, such as contextual and linguistic elements, also have been addressed by Connolly (1986), Kent (1993), and Weston and Shriberg (1992). Shriberg (1993) acknowledged the importance of examining suprasegmental elements in speech evaluation, by proposing the ACI as only one part of a larger assessment system, that includes measures of voice and prosody. The contextual, linguistic, and suprasegmental factors examined by these authors are potential sources of variability in speech intelligibility not accounted for by articulatory and phonological factors.

Caution should be exercised in drawing conclusions from this study regarding the validity of the ACI as an validity of the reliability data are only one area of question. Other considerations involve differences in methodology between this study (including data used from the Gordon-Brannan study) and the previously cited studies addressing the ACI and its component, PCC. Shriberg and Kwiatkowski (1982) and Shriberg (1993) used percentage-of-intelligible words data from on-line transcription of children's speech and from recordings of the same speech samples, presented one utterance at a time with no repetitions. In contrast, the percentage-of-words-understood data used in this study were obtained from as many as three presentations of each recorded utterance (Gordon-Brannan, 1993). Also, the recordings used in this study differed in that those in the Shriberg and Kwiatkowski (1982) and Shriberg (1993) studies were "compressed" by removing pauses, while the samples recorded by Gordon-Brannan (1993) included such pauses, as well as comments from the interviewer. The impact of these differences, particularly on the number of words, utterances, target consonants, and linguistic cues in this sample, is not clear at this time.

Additional Considerations

This study was, in its design, preliminary and limited, so certain methodological factors, particularly in the area of reliability, should be kept in mind in considering the results. It is important also to consider the anecdotal report of the investigator regarding her experience in ACI scoring.

As anticipated, scoring the samples was a time-consuming task that required a great deal of concentration and effort. In initial listening sessions, it took approximately 2 hrs to score three 100-word samples. With practice, however, the time required was less than half that, that is, 20 to 30 min per sample. This investigator, as well as the second scorer, made a determined effort to be thorough and accurate in evaluating the recorded samples and in interpreting the instructions for ACI scoring. Both individuals reported, however, that, though they gave as much thought as possible to scoring each sample, they continued to question their judgment, particularly regarding scoring co-articulated speech sounds. Both scorers reported resolving this dilemma by listening to each sample as many times as necessary to satisfy themselves that they had done the best they could. Both scorers indicated, however, that they never reached a point where they believed that they had scored a sample flawlessly. This indicates the complexity of the ACI scoring task, which should be considered in evaluating the results of this study and in designing others to address the validity of the ACI.

CHAPTER V

SUMMARY AND IMPLICATIONS

Summary

Intelligibility, which refers to how recognizable a speaker's words are to the listener, is largely responsible for the effectiveness of an individual's spoken message. Severity, a broader but closely related concept, incorporates intelligibility, disability, and handicap. Many factors influence intelligibility, including speech sound production, voice, and prosody, as well as linguistic and contextual factors. The variety of influences on intelligibility contribute to difficulty in quantifying the concept.

Clinicians and researchers in the field of speech-language pathology require accurate measures of intelligibility and severity to assess and describe communicative functioning and to measure change over time. Intelligibility and severity have been measured by a number of scales and rating systems, as well as by subjective impressions of listeners. Other measurement procedures address particular aspects of speech production, such as articulation, phonology, vocal quality, or fluency. Determining the efficiency and accuracy of these approaches has been the focus of recent attention in the field.

This study was a preliminary investigation of the relationship between the Articulation Competence Index (ACI), a severity metric, and the percentage of words understood in continuous speech, the standard measure of intelligibility. Specifically, the study addressed the research question:

Is there a significant correlation between the Articulation Competence Index (ACI) and percentage of words understood in samples of continuous speech of 4- and 5-year-olds with varying levels of phonological competence?

Subjects for the study were thirty 4- and 5-year-olds from the Portland metropolitan area. Four listeners calculated percentage-of-words scores for each child's 100-word speech sample. These scores were compared to ACI scores calculated by the investigator for each of the samples.

The data were analyzed using the Pearson product-moment correlation (Pearson $\underline{r}$). A significant positive correlation ($\underline{r}$ = .71 to .81) was found, indicating a moderately strong correlation between the ACI and percentage of words understood. Squaring the correlation coefficients resulted in values for $\underline{r}^2$ of .50 to .66, indicating that the independent variable, the ACI accounts for more than half the variability of continuous speech intelligibility.

The results of the study suggest that the ACI does, to a significant degree, reflect the intelligibility component of severity. However, concerns regarding methodology, particularly the limited number of samples used in examining intra- and inter-rater reliability in this study, should be considered when evaluating the results.

It is suggested that the limitations of this study be considered in designing future studies to establish the validity of the ACI as an instrument for use in longitudinal studies to quantify severity of articulatory and phonological involvement in individuals over their lifetimes.

Implications

The results of this study have implications for clinical practice as well as for research. Some of these considerations regard the use of the Articulation Competence Index (ACI) itself, while other considerations involve more indirect inferences to be drawn from the outcome of this study. It is, therefore, important to keep in mind that the results suggest that the ACI does reflect intelligibility, and that the parameters of speech production assessed by the ACI do not account for all the variability in intelligibility.

Clinical Implications

The author has reported her experiences in calculating the Articulation Competence Index (ACI) scores for 30 subjects and has noted the complexity of this task. At the present time, it appears unlikely that the ACI, a time-consuming and as yet not thoroughly validated measure, will have widespread clinical utility.

The most significant consideration for clinical practice arising from this study involves the indication that the parameters of speech measured by the ACI, that is, phoneme production, account for little over half of the variability in intelligibility. Though this was a preliminary study, the results suggest that the clinician would do well to consider that, since speech sound production is not the only determinant of intelligibility, speech sound production should not be the sole focus of assessment and resulting treatment.

Further research regarding the validity of the ACI, particularly as a predictor of a young child's future articulatory or phonological competence also will have potential benefits for clinical practice. These benefits will involve treatment planning, particularly in the area of target selection. A more complete discussion of these considerations follows.

Research Implications

This was a preliminary study with inherent methodological limitations. While the results appear to suggest that the Articulation Competence Index (ACI) does, to some degree, reflect the intelligibility component of severity, more thorough studies are indicated to yield more conclusive results regarding the validity of the ACI as a measure of severity.

Shriberg (1993) has reported that ACI scores provide excellent separation of speech-delayed from speech-normal preschoolers. A question arises, however, regarding whether the ACI is the most efficient and effective means of identifying speech-delayed children. Studies comparing results and examiner experiences from administration of traditional articulation and phonological assessment instruments to ACI scores would be helpful in making this determination.

Hodson and Paden (1981) reported that speech error type, rather than frequency of speech sound errors, had the greater effect on intelligibility. As ACI scoring involves both error types (distortions versus omissions, substitutions, and additions) and frequency of errors (in deriving the Percentage of Consonants Correct (PCC) and

Relative Distortion Index (RDI) components), studies designed to isolate these elements could either challenge or provide support for Hodson and Paden's conclusions.

Shriberg (1993) proposed the ACI as a means to quantify severity of involvement in individuals from 2 years of age through adulthood. He also stated his intention that examination of data obtained through longitudinal studies utilizing the ACI as one component of a 10-part Speech Disorders Classification System would provide a means of charting the progression of developmental phonological disorders that are initially manifested during the preschool years. Information from such longitudinal studies would aid in understanding developmental phonological disorders, and would be particularly useful in predicting outcomes. Specifically, understanding which early manifestations of developmental phonological disorders are most easily resolved could be of use in treatment planning, particularly in selecting clients and determining targets for intervention. It is important to note, however, that the ACI is only one component of the larger Speech Disorder Classification System, proposed to offer such predictive value. Determining the validity and usefulness of the ACI and the other components of the larger system remains to be addressed by future research.

Increasing intelligibility is often the goal for speech remediation. Increased understanding of intelligibility is a goal for research in the area of speech communication. Determining the most effective and efficient means of measuring intelligibility, as an entity or as a component of the larger concept of severity, is vital to the clinical practice of speech-language pathology, and to the research efforts that help form the knowledge base for that field.

References

Bernthal, J. E., & Bankson, N. W. (1993). Articulation and phonological disorders (3rd Ed.). Engelwood Cliffs, NJ: Prentice Hall.

Billman, K. S. (1986). Phonological processes and intelligibility of spontaneous utterances in young children. Unpublished master's thesis, San Diego State University, San Diego, CA.

Carrow-Woolfolk, E. (1985). Test for auditory comprehension of language-revised. Allen, TX: DLM.

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

Garrett, K. K., & Moran, M. J. (1992). A comparison of phonological measures. Language, speech, and Hearing Services in Schools, 23, 48-51.

Gordon-Brannan, M. (1993). Speech intelligibility assessment of young children with varying levels of phonological proficiency/deficiency. Unpublished doctoral dissertation, The Wichita State University, Wichita, 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. (1986). The assessment of phonological processes revised. Austin, TX: Pro-ed.

Kent, R. D. (Ed.). (1992). Intelligibility in speech disorders. Amsterdam: John Benjamins.

Kent, R. D. (1993). Speech intelligibility and communicative competence in children. In A. P. Kaiser & D. B. Gray (Eds.), Enhancing children's communication: Research foundations for intervention (Vol. 2, pp. 223-239). Baltimore: Brookes Publishing Co.

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.

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

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

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. Journal of Speech and Hearing Research. 36, 105-140.

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.

SPSS for Windows. (1993). Chicago: SPSS Inc.

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

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

APPENDIX A

Procedures to Calculate Percentage of Consonants Correct (PCC), According to Shriberg and Kwiatkowski (1982)

The following procedures are used to calculate Percentage of Consonants Correct (PCC):

<u>Sampling Rules</u>

1. Consider only intended (target) consonants in words. Intended vowels are not considered.
   a. Addition of a consonant before a vowel, e.g., on [hɔn] is not scored because the target sound /ɔ/ is a vowel.
   b. Post-vocalic /r/ [feir] *fair* is a consonant, but stressed and unstressed vocalics [ɝ], [ɚ], as in *furrier* [fɝiɚ] are vowels.
2. Do not score target consonants in the second or successive repetitions of a syllable, e.g., *ba-balloon*. Score only the first /b/.
3. Do not score target consonants in words that are completely or partially unintelligible or whose gloss is highly questionable.
4. Do not score target consonants in the third or successive repetitions of adjacent words unless articulation changes. For example, the consonants in only the first two words of the series [kæt], [kæt], [kæt] are counted. However, the consonants in all three words are counted if the series were [kæt], [kæk], [kæt].

Scoring Rules

1. The following six types of consonant sound changes are scored as incorrect:
   a. deletions of a target consonant;
   b. substitutions of another sound for a target consonant, including replacement by a glottal stop or a cognate;
   c. partial voicing of initial target consonants;
   d. distortions of a target sound, no matter how subtle;
   e. addition of a sound to a correct or incorrect target consonant, e.g., *cars* said as [karks].
   f. initial /h/ deletion (he [i]) and final n/ŋ substitutions (ring [rin]) are counted as errors only when they occur in stressed syllables; in unstressed syllables they are counted as correct, e.g., feed her [fidɚ]; or running [rʌnin].
2. Observe the following:
   a. The response definition for children who obviously have speech errors is "score as incorrect unless heard as correct." This response definition assigns questionable speech behaviors to an "incorrect" category.
   b. Dialectal variants should be glossed as intended

in the child's dialect, e.g., *picture* "piture"; *ask* "aks", etc.

c. Fast or casual speech sound changes should be glossed as the child intended, e.g., *don't know* "dono"; *and* "n", etc.

d. Allophones should be scored as correct, e.g., *water* [waɾɚ], *tail* [teɾ l̩].

Calculation of PCC

The percentage of Consonants Correct (PCC) for a speech sample is calculated by the formula:

$$\text{PCC} = \frac{\text{Number of Correct Consonants}}{\text{Number of Correct Plus Incorrect Consonants}} \times 100$$

APPENDIX B

Categories of Distortions Used in Calculating the Articulation Competence Index (ACI), Adapted from Shriberg (1993)

The following outlines the types and categories of distortions used in calculating the Articulation Competence Index (ACI). A more thorough description is available in Shriberg (1993).

Articulatory distortions comprise four subtypes:

(a) *non-clinical* versus *clinical* speech-sound errors, and

(b) *uncommon* versus *common*, based on occurrence during different ages of normal speech development.

Nonclinical Distortions

Shriberg (1993) defines nonclinical distortions as "speech-sound differences of allophones that are due to dialectal or idiolectal differences in linguistic background or speech-motor constraints" (p. 132). These are not considered distortion errors in calculating the ACI. Examples include:

1. palatalized /s/ ([s̡], sometimes called a "hissy s");
2. retroflexed /s/ ([ʂ], sometimes called a "whistling s");
3. deletion of initial /h/ in unstressed pronouns;
4. substitution of glottal stop for /t/ in word-final position.

Clinical Distortions

These are further classified as "common" or "uncommon," according to the speaker's age.

Common Distortions. For purposes of calculating the ACI, these are always scored as distortions, regardless of the speaker's age. Included are:

1. labialized /l/ or /r/,
2. velarized /l/ or /r/,
3. lateralized voiced or voiceless sibilant fricatives or affricates,
4. derhotacized /r/, /ɝ/, or /ɚ/,
5. dentalized voiced or voiceless sibilant fricatives or affricates.

Uncommon Distortions. This classification includes distortions that may involve both consonants and vowels, and all such errors are scored as distortions for purposes of calculating the ACI. The four classes of uncommon distortions are:

1. Weakly articulated consonants.
2. Imprecise consonants and vowels.
    a. on-glides or off-glides (epenthetics) on consonants or vowels/dipthongs, excepting epenthetic stops on nasals (see below)
    b. notably lowered, raised, fronted, or backed vowels/dipthongs

c. notably lengthened or shortened durations of consonants and vowels

d. notably aspirated stops

e. notably frictionalized stops and fricatives

f. notably pharyngealized velar stops

3. Failure to maintain oral/nasal contrasts.

a. nasal emissions

b. denasalized nasal consonants (and epenthetic stops) in the absence of upper respiratory involvement

c. nasalized consonants (i.e., /m/-like sound replacing /b/ or /p/; /n/-like sound replacing /d/, /t/, or /l/

d. nasalized vowels/dipthongs in contexts other than those appropriate for assimilative nasality

4. Notable failure to maintain appropriate voicing. "Full" voicing errors (saying /s/ instead of /z/) are treated as substitution errors, not distortions. The following voicing errors are only scored as distortions in children over 5 years old, and then only when noticeable and consistent in several speech sounds and sound classes.

a. notable nonaspiration of prevocalic voiceless stops

b. notable partial voicing of voiceless stops, fricatives, and affricates

c. notable partial devoicing of voiced stops, fricatives, and affricates.

Calculation of the Articulation Competence Index (ACI)

Scores are obtained through the following formula:

$$ACI = \frac{PCC + RDI}{2}$$

The variables in the formula for ACI are defined as follows:

$$PCC = \frac{\text{TOTAL CORRECTLY ARTICULATED CONSONANTS}}{\text{TOTAL CONSONANTS}}$$

$$RDI = \frac{\text{ERRORS DUE TO DISTORTIONS}}{\text{TOTAL ARTICULATION ERRORS}}$$

APPENDIX C

Portion of an Orthographic Transcription of a Speech Sample, With Markings Used in Calculating the Articulation Competence Index (ACI)

SCORE SHEET SUBJECT:1/SAMPLE:4

/	X	O	Utterance	
1		1	Hugging each other. hʌgɪn itʃ ʌðɚ	(5)
1	1		It's this page my thumb's on. ɪts ðɪs peɪdʒ maɪ θʌmz an	(11)
2	1		They're in the house. ðeɪr ɪn ðə haʊs	(6)
	1		Having a party. hævɪn ə partɪ	(6)

Note: Markings used in scoring the samples include: / = consonant error; X = distortion; O = vowel distortion. The total marks for each utterance were entered in respective columns to the left of each utterance, and the numbers in each column were added to calculate the total errors of each type.

A full phonetic transcription of each utterance was required to accurately calculate the number of target consonants in each sample. The number circled to the right of each phonetically transcribed utterance indicates the number of target consonants in that utterance. The circled numbers for all the utterances in a sample were added to yield the total number of target consonants in each sample.

APPENDIX D

Articulation Competence Index (ACI) Scores Calculated by Each of Two Scorers

Subject - Sample	Scorer A	Scorer B
1 - 04	63	68
2 - 23	30	28
3 - 22	54	48
4 - 40	70	63
5 - 38	49	43
6 - 33	60	65

Note: Scorer A was the investigator. Scorer B was the second graduate student who participated in the procedures to address intra-rater reliability.

APPENDIX E

Raw Data: Scores for Percentage of Words Understood and the Articulation Competence Index (ACI)

Subject/	Percentage of Words Understood				ACI
Sample	Listener A	Listener B	Listener C	Listener D	
01-04	92	95	97	89	70
02-23	72	59	58	65	33
03-22	70	71	68	76	51
04-40	86	79	76	84	76
05-38	68	77	71	67	44
06-33	92	91	89	86	63
07-10	87	80	87	90	45
08-32	70	69	59	70	51
09-14	91	89	92	88	64
10-39	91	92	97	93	77
11-15	90	92	92	91	97
12-30	67	63	76	69	48
13-07	83	83	83	83	67
14-48	98	97	92	94	95
15-27	56	53	36	53	38
16-28	82	74	76	70	52
17-05	92	86	98	98	99
18-11	59	73	64	82	56
19-16	49	69	62	64	44
20-25	92	95	82	85	62
21-36	98	99	99	97	97
22-43	51	45	40	52	34
23-47	89	85	81	87	66
24-12	65	72	78	77	56
25-26	85	88	83	87	61
26-01	75	91	92	87	70
27-17	70	82	83	80	53
28-03	37	46	35	49	51
29-18	74	84	84	78	52
30-13	92	98	96	96	97