
Comparisons of Children's Single-Word Articulation Proficiency, Single-Word Speech Intelligibility, and Conversational Speech Intelligibility

Comparaisons de l'habileté à articuler les mots-phrases, de l'intelligibilité de la parole à mots-phrases et de l'intelligibilité de la conversation normale chez les enfants

Mark A. Stimley, PhD and Georgia Hambrecht, PhD

Indiana State University
Terre Haute, Indiana

ABSTRACT

To examine the relationship between single-word articulation proficiency and speech intelligibility of children, the Arizona Articulation Proficiency Scale (AAPS; Fudala & Reynolds, 1986) and the single-word speech intelligibility subtest of the Assessment of Intelligibility of Dysarthric Speech (AIDS; Yorkston & Beukelman, 1981) were administered to 15 children who were receiving speech-language therapy for remediation of articulation or phonological disorders. Conversational speech samples were also collected using an interview technique (Evans & Craig, 1992) and analyzed for speech intelligibility. Results indicated that a significant moderate positive correlation existed between AAPS single-word articulation proficiency scores and AIDS single-word speech intelligibility scores. Results also indicated nonsignificant positive correlations between AAPS scores and conversational speech intelligibility scores and between AIDS single-word multiple choice speech intelligibility scores and conversational speech intelligibility scores. Findings are discussed in terms of (a) factors that may account for the magnitude of the correlation between articulation proficiency and speech intelligibility and (b) clinical implications of these findings.

ABRÉGÉ

Afin d'examiner le rapport entre l'habileté à articuler les mots-phrases et l'intelligibilité de la parole chez les enfants, on a administré l'*Arizona Articulation Proficiency Scale* (AAPS; Fudala & Reynolds, 1986) et le sous-test d'intelligibilité de la parole à mots-phrases de l'*Assessment of Intelligibility of Dysarthric Speech* (AIDS; Yorkston & Beukelman, 1981) à 15 enfants qui recevaient une thérapie orthophonique pour remédier à des troubles d'articulation ou phonologiques. On a également recueilli des échantillons de conversation normale au moyen d'une technique d'entrevue (Evans & Craig, 1992) qui ont été analysés pour en déterminer l'intelligibilité. Les résultats ont montré qu'il existe une corrélation positive modérée significative entre les scores d'habileté à articuler des mots-phrases AAPS et les scores d'intelligibilité des mots-phrases AIDS. Les résultats ont également relevé des corrélations positives non significatives entre les scores AAPS et les scores d'intelligibilité de la conversation normale et entre les scores d'intelligibilité AIDS de mots-phrases à choix multiples et les scores d'intelligibilité de la conversation normale. Les conclusions sont étudiées en fonction (a) des facteurs pouvant expliquer l'importance de la corrélation entre l'habileté à articuler et l'intelligibilité de la parole et (b) les incidences cliniques de ces conclusions.

KEY WORDS: articulation • speech intelligibility

Accurate measurement of speech intelligibility (i.e., that aspect of oral speech-language output that allows a listener to understand what a speaker is saying; Nicolosi, Harryman, & Kresheck, 1996) is a critical component for the assessment of children with phonology disorders (Gordon-Brannan, 1994). In 1982, Shriberg and Kwiatkowski observed that practicing speech-language pathologists (SLPs) typically make subjective judgements concerning the percentage of intelligible words in connected speech as their primary measure of speech intelligibility. Since that time, the number of methods to measure speech intelligibility has grown (e.g., Fudala & Reynolds, 1986; Schiavetti, 1984; Shriberg & Kwiatkowski, 1982; Weiss, 1982). Unfortunately, our observations indicate that many speech-language pathologists are not

regularly using these methods, but instead are continuing to rely on their subjective judgements. The practice of using subjective judgements as primary speech intelligibility measures is particularly distressing because they are unreliable in very young children (Rvachew, Russell, & Rafaat, 1993) and less stable than more objective systematic measures of speech intelligibility for children with moderate or severe disordered phonologies (Gordon-Brannan & Sugarman, 1994). An exception to this tendency to use subjective judgements to describe intelligibility occurs when an examiner administers the Arizona Articulation Proficiency Scale (AAPS; Fudala & Reynolds, 1986) as an articulation inventory and as part of the standard scoring procedures obtains an estimate of speech intelligibility.

The AAPS, a test of children's articulation, provides the SLP



with an inventory of the consonants and vowels that a child produces correctly and a method for estimating the child's speech intelligibility. AAPS total scores (also known as articulation proficiency scores) and subsequent estimates of speech intelligibility are based on the misarticulations that children experience and the probable frequency of occurrence in North American speech of the misarticulated sounds. AAPS total scores and speech intelligibility estimates decrease as children experience more misarticulations and as these misarticulations occur on sounds used with greater frequency in the language (e.g., misarticulation of final /t/ sounds reduces the total score and the speech intelligibility estimate more than misarticulation of final /k/ sounds because final /t/ occurs four times more often in the language than final /k/).

Although AAPS total scores may range anywhere along a continuum from zero to 100, these scores are not used to estimate speech intelligibility in terms of the percent of words correctly understood by the listener. Instead of percentages, the AAPS manual provides a table for interpreting AAPS scores into descriptive statements about speech intelligibility (e.g., scores that range from 85 to 94.5 are interpreted as indicating that "speech is intelligible, although noticeable in error," while scores that range from 70 to 84.5 are interpreted as indicating that "speech is intelligible with careful listening"). Although two studies (Schissel & James, 1979a, 1979b) concluded that the AAPS's sampling technique underestimated the severity of articulation disorders and therefore may also underestimate the severity of speech intelligibility problems, no study has compared AAPS performance and its estimate of speech intelligibility with an independent objective measure of speech intelligibility.

One of the few commercially available objective measures of speech intelligibility is the Assessment of Intelligibility of Dysarthric Speech (AIDS; Yorkston & Beukelman, 1981). Although the AIDS was designed to assess the effects of acquired articulation disorders called the dysarthrias on adult speech intelligibility, examination of the methods that the AIDS uses to collect speech samples (word imitation) and quantify single-word speech intelligibility (in terms of percent of correct words identified by a naive listener) indicates that this test could be used to assess the effects of developmental articulation disorders on speech intelligibility. Recent findings of Morris, Wilcox, and Schooling (1995) also support the use of similar methods with children. Administration and scoring procedures for the single-word speech intelligibility subtest of the AIDS may also help control several factors known to affect subjective judgments of speech intelligibility. These factors include such things as speaking rate, length and complexity of words, and the listener's familiarity with the participant's articulation patterns and topic (Shipley & McAfee, 1992; Yorkston & Beukelman, 1981).

The primary purpose of this study is to examine the relationship between articulation proficiency as measured by the

AAPS and speech intelligibility (percent of correctly identified words) as measured by the AIDS. The secondary purpose of this study is to examine the relationships between these single-word articulation and intelligibility measures and conversational speech intelligibility.

Method

Participants

Participants for this study were 15 children (11 males and four females). As shown in Table 1, participants ranged in age from four to 10 years (mean = 5.9 years). At the time of testing, all participants were enrolled in speech-language therapy for articulation or phonological disorders at a university speech-language clinic.

To be included in this study, each participant also needed to (a) possess normal speech sound discrimination (based on report by the participant's clinician) and (b) pass a pure tone hearing screening test bilaterally at 500, 1000, 2000, and 4000 Hz at 25 dB. Hearing screening tests were conducted in the same rooms in which the experimental protocol would later be administered. All participants were also screened to ensure that they had the picture naming and word repetition skills needed to respond appropriately to the AAPS and AIDS stimulus items.

Procedures

Each participant was audiotaped as they were administered the picture test portion of the AAPS and the multiple-choice format of the single-word speech intelligibility portion of the AIDS by the graduate student SLP assigned to the child in the clinic. Administration and scoring of these two tests were consistent with the procedures recommended in the test manuals. Standard conversational speech samples between the participants and their graduate student SLPs were also elicited and audiotaped.

Test presentations were counterbalanced to control for possible order effects. Half the participants were administered the AAPS first, and the other half were administered the AIDS first. Conversation speech samples were collected after the administration of the two tests. Most of the testing and conversational speech sampling was completed within a single 50-minute session. If testing and sampling were conducted on more than one day, the length of time between administrations varied from two to five days depending on when the participant's next regularly scheduled therapy session occurred.

Testing was conducted in a quiet therapy room with both the student SLP and the research assistant present. The student SLP presented the stimulus items from the tests, while the research assistant transcribed responses to the AAPS stimuli, monitored the administration of the tests, controlled the audio-tape equipment, and took contextual notes during the conversational speech sample. Conversational speech samples and responses to test stimuli were recorded in stereo on a Technics RS-B29R



tape recorder using a Sony EMC-155 microphone attached to the collar of the participant's clothing and a Sony ECM-150 microphone placed on a stand on the table in front of the participant.

During the administration of the AAPS, participants were presented with 48 test items that required participants to name pictures and to respond to accompanying questions to elicit 70 target sounds. Recorded responses were transcribed and subsequently scored for correctness of targeted consonants and vowels by the research assistant. AAPS total scores were determined according to the standard procedures and formulae presented in the AAPS test manual.

Before the administration of the AIDS, a unique 50-word stimulus set for each participant was created. Stimulus sets were created by randomly choosing one word from each of the original 12-word sets on the AIDS. During the administration of the AIDS, participants repeated each of the 50 single-word targets following the examiners' model. Pointing cues were used to prompt the child to say each word at a particular time. This cueing procedure was used to allow the research assistant time to turn the tape recorder off during the examiner's presentation of the verbal model and then turn it back on to record the participant's imitated response. All participants completed this test procedure without difficulty.

Following the administration of AIDS, the first author and two graduate students in speech-language pathology, who were unfamiliar with the participants (but who were familiar with the 600 AIDS words), scored the 15 tapes independently of each other. Scoring was completed using the multiple-choice format of the AIDS that required listeners to choose single-words from 12-item multiple-choice lists that best corresponded to the participant's taped productions. Mean AIDS scores for each of the 15 children were then calculated based on the three sets of listener scores.

A conversational speech sample was obtained for each child using an interview format (Evans & Craig, 1992). Each child was given question prompts about his or her family, school, or free-time activity (e.g., Are you in school? Tell me about it.). Once the topic was introduced, the graduate student examiner followed the child's conversational lead. A middle 100 consecutive words were transcribed and scored for speech intelligibility (i.e., the percent of words in the sample that were understood by the listener) using procedures described in Shipley and McAfee (1992).

Reliability

Intrajudge and interjudge point-to-point reliabilities (McReynolds & Kearns, 1983) based on four randomly selected participants were determined for the AAPS. Intrajudge reliability was determined by having the research assistant score the AAPS tapes and then score the four randomly selected tapes

again after a period of three weeks. Point-to-point intrajudge reliability (based on agreement on the occurrence and nonoccurrence of target behaviours on a response-by-response basis) for the AAPS was determined to be 93%. Interjudge reliability was determined by having the second author score four randomly selected AAPS tapes and then compare these scores to the research assistant's first AAPS scores. Point-to-point interjudge reliability for the AAPS was determined to be 92%.

Using the point-to-point method and the total method (McReynolds & Kearns, 1983), AIDS intrajudge reliabilities were determined by having the first author score the AIDS from tape and then rescore four randomly selected AIDS tapes after a period of three weeks. Total method reliability was obtained by summing the smaller scores recorded by this judge for the four participants, dividing this sum by the sum of the larger scores recorded by this judge for the same participants, and multiplying this quotient by 100. Point-to-point and total method intrajudge reliability scores for the AIDS were 86% and 99%, respectively. Given the multiple-choice forced-choice nature of the scoring task that AIDS judges perform, the use of the total method may more accurately reflect the procedure's true reliability than the use of the point-to-point method.

Point-to-point and total method interjudge reliabilities for the AIDS were determined by having the second author score four randomly selected AIDS tapes and then compare these scores to the first author's first AIDS scores using the methods described above. Point-to-point and total method reliabilities for the AIDS were determined to be 80% and 97%, respectively.

Results

AAPS total scores, mean AIDS scores (mean percent of words correctly identified by three listeners), and conversational speech intelligibility scores (percent intelligible words scores) were determined for each participant and are reported in Table 1. Pearson Product Moment correlations were calculated on these sets of scores to determine if there were relationships between the measures.

Results indicated that there was a moderate positive relationship between AAPS scores and AIDS single-word multiple-choice speech intelligibility scores ($r = .57, p < .05$). Within the group, as a participant's articulation proficiency score increased, the participant's speech intelligibility score also tended to increase.

Results also showed the presence of weak positive correlations between AAPS scores and conversational speech intelligibility scores ($r = .49, p > .05$) and between AIDS single-word multiple choice speech intelligibility scores and conversational speech intelligibility scores ($r = .24, p > .05$). These correlations between the two single-word measures and the conversation measure were statistically nonsignificant.

Discussion

Based on these results, there is a significant moderate positive correlation between the two measures of single-word speech intelligibility used in the study. The strength of this correlation indicated that although the speech intelligibility estimates obtained on the AAPS and the intelligibility score on the AIDS are related, there are factors beyond articulation proficiency (as measured by the AAPS) that influence speech intelligibility scores (as measured by the AIDS) even at the single word level.

This moderate correlation supported the position stated by Peterson and Marquardt (1981) that although articulation proficiency and speech intelligibility are related, they are not identical. The correlation between articulation proficiency and speech intelligibility in this study was also congruent with the recent findings of Morris et al. (1995) who compared percentile scores on the Goldman-Fristoe Test of Articulation (GFTA; Goldman & Fristoe, 1986) to percent intelligibility scores obtained on their modified version of the AIDS test called the Preschool Speech Intelligibility Measure (PSIM; $r = .73, p < .01$).

The weaker correlations between both of the single-word measures (AAPS scores and AIDS scores) and the conversational speech intelligibility measure ($r = .49$ and $r = .24$, respectfully) are consistent with previous findings (Healy & Madison, 1987; Morrison & Shriberg, 1992; Yorkston & Beukelman, 1981). These correlations suggest that there are critical differences between articulation and speech intelligibility measures at the single word level and speech intelligibility measured at the conversational level. SLPs are well advised not to rely solely on single-word sampling methods when evaluating a child's overall speech intelligibility abilities. If the SLP is interested in speech intelligibility at the conversational level, the SLP would be best served by measuring it in conversation.

Although both single word to conversational correlations were statistically nonsignificant, examination of the relative strengths of the correlations is revealing. The correlation between the AAPS score and conversational speech intelligibility is slightly stronger than the correlation between the AIDS scores

Table 1. Identification Information and Performance Scores of each Participant.

Participant	Sex	Age years	AAPS Total Score	Mean AIDS Score mean percent correctly identified words score	Conversational Speech Intelligibility percent intelligible words score
1	M	4	78.5	60.7	43
2	M	5	83.5	82.0	50
3	M	5	85.5	54.7	67
4	F	8	95.0	94.0	95
5	M	4	79.0	69.3	44
6	M	4	84.0	77.3	55
7	M	9	90.5	90.7	65
8	F	5	90.0	75.3	70
9	M	5	94.0	63.3	81
10	F	4	69.0	46.0	79
11	M	8	87.5	81.3	83
12	M	4	93.5	56.7	76
13	M	5	81.0	58.7	77
14	M	8	87.0	86.7	68
15	F	10	93.0	96.7	98
Means		5.9	86.1	72.9	70.0
Standard Deviations		2.1	7.2	15.6	16.8

and conversational speech intelligibility. Perhaps the AAPS score which is partially based on the frequency of occurrence of the misarticulated speech sounds within the language is responsible for this slightly stronger correlation. Future research studies that examine the relationship between articulation proficiency and the intelligibility of conversational speech might be improved if articulation proficiency is measured during connected speech tasks that attempt to control for the frequency of occurrence of speech sounds in the language (such as the AAPS sentence reading test does).



Clinical Implications

Results of this study suggest that if an SLP finds it necessary for clinical or administrative purposes to rate children according to single-word speech intelligibility based on an objective measure of percent of words correctly identified by a listener, then the use of a test like the single-word speech intelligibility subtest of the AIDS would be an efficient and reliable method of obtaining objective ratings. This portion of the AIDS also provides the SLP with information about the severity and the nature of a child's speech (articulation-intelligibility) disorder that supplements information obtained from the speech sound inventory, rather than simply providing confirming evidence of the existence of a disorder.

These results also provide support for the argument that SLPs should not attempt to predict conversational speech intelligibility based on these single-word level articulation (AAPS) and speech intelligibility (AIDS) performance measures. Single-word and conversational production tasks are different and warrant independent examination.

Acknowledgments

This work was supported in part by a grant obtained from the Blumberg Center for Interdisciplinary Studies in Special Education, Indiana State University, Terre Haute, Indiana.

We thank Christy Hobson-Meek who coordinated activities and experimental protocol administrations, our student SLPs, and our participants and their parents who kindly volunteered to participate in this study.

Submitted: December 1997

Accepted: January 1999

Please address all correspondence to: Mark A. Stimley, PhD, Department of Communication Disorders and Special Education, Indiana State University, Terre Haute, Indiana 47909, USA.

References

- Evans, J. L., & Craig, H. K. (1992). Language sample collection and analysis: Interview compared to freeplay assessment contexts. *Journal of Speech and Hearing Research, 35*, 343-353.
- Fudala, J., & Reynolds, W. (1986). *Arizona Articulation Proficiency Scale* (2nd ed.). Los Angeles: Western Psychological Services.
- Goldman, R., & Fristoe, F. (1986). *Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service.
- Gordon-Brannan, M. (1994). Assessing intelligibility: Children's expressive phonologies. *Topic in Language Disorders, 14*, 17-25.
- Gordon-Brannan, M., & Sugarman, N. (1994, November). *Measuring intelligibility: Accuracy of estimated percentage of words understood*. Paper presented at the annual convention of the American Speech-Language-Hearing Association, New Orleans, LA.
- Healy, T. J., & Madison, C. L. (1987). Articulation error migration: A comparison of single word and connected speech samples. *Journal of Communication Disorders, 20*, 129-136.
- McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. Baltimore, MD: University Park Press.
- Morris, S. R., Wilcox, K. A., & Schooling, T. L. (1995). The Preschool Speech Intelligibility Measure. *American Journal of Speech-Language Pathology, 4*, 22-28.
- Morrison, J., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research, 35*, 259-273.
- Nicolosi, L., Harryman, E., & Kresheck, J. (1996). *Terminology of communication disorders: Speech-language-hearing* (4th ed.). Baltimore, MD: Williams & Wilkins.
- Peterson, H. A., & Marquardt, T. P. (1981). *Appraisal and diagnosis of speech and language disorders*. Englewood Cliffs, NJ: Prentice-Hall.
- Rvachew, S., Russell, R., & Rafaat, S. (1993, November). *Severity rating: How reliable are we?* Paper presented at the annual convention of the American Speech-Language-Hearing Association, Anaheim, CA.
- Schiavetti, N. (1984). Scaling procedures for quantification of speech, language, and hearing. In R. G. Daniloff (Ed.), *Articulation assessment and treatment issues* (pp. 327-353). Philadelphia, PA: John Benjamins.
- Schissel, R. J., & James, L. B. (1979a). A comparison of children's performance on two tests of articulation. *Journal of Speech and Hearing Disorders, 44*, 363-372.
- Schissel, R. J., & James, L. B. (1979b). An investigation of the assumptions underlying the Arizona Articulation Proficiency Scale (Revised). *Language, Speech, and Hearing Services in Schools, 10*, 241-245.
- Shipley, K. G., & McAfee, J. G. (1992). *Assessment in speech language pathology: A resource manual*. San Diego, CA: Singular.
- 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, A. (1982). *Weiss Intelligibility Test*. Tigard, OR: CC Publications.
- Yorkston, K., & Beukelman, D. (1981). *Assessment of Intelligibility of Dysarthric Speech*. Tigard, OR: CC Publications.

