11-1-1995

The Preschool Speech Intelligibility Measure

Sherrill R. Morris

Original Citation

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

Clinical Focus

The Preschool Speech Intelligibility Measure

Sherrill R. Morris
Kim A. Wilcox
University of Kansas, Lawrence

Tracy L. Schooling
Chesapeake Center, Inc., Alexandria, VA

Documenting changes in speech intelligibility across time is an important but difficult task for speech-language pathologists. This study reports on the development and initial testing of the Preschool Speech Intelligibility Measure (PSIM), a single-word, multiple-choice intelligibility measure. The PSIM is adapted from the Assessment of Intelligibility of Dysarthric Speech (Yorkston & Beukelman, 1981) and is designed to plot changes in children's speech intelligibility across time. This instrument is offered as an addition to the existing array of available speech intelligibility measures.

Impairment of the phonological system is one of the most common communication disorders of young children. In 1971, Milisen estimated that functional articulation disorders make up 75 to 80% of all speech impairments in children. Seventeen years later, the National Institute of Neurological and Communicative Disorders and Stroke concluded that 10 to 15% of all preschoolers had a speech disorder (cited in Shames, Wiig, & Secord, 1994). Despite the relatively high prevalence of phonological problems in young children, no standard procedure exists for quantifying the intelligibility of this group (Kent, Miolo, & Bloedel, 1992, 1994). Even though measures are specifically developed for young children (see Kent et al., 1994, for review) they are not widely used by practicing speech-language pathologists. As a result, the present measure of choice for assessing intelligibility in preschool children appears to be subjective estimation of performance such as assignment of percentage intelligibility (e.g., 50% intelligible, 20% intelligible, etc.) or degree of understandability (e.g., easily understood, understandable if topic is known, some words intelligible, unintelligible) (Peterson & Marquardt, 1994). Kent, Weismer, Kent, & Rosenbek (1989) note that scales such as these offer the "advantage of speed but are only crudely quantitative." Part of the difficulty in assessing intelligibility is the inherent subjectivity of the phenomenon. Although articulatory accuracy is a key component in determining the understandability of a message, several other factors are also important contributors. As a result, one listener may consider a message completely unintelligible; a second listener, aided by contextual cues, may be able to accurately grasp the general meaning of the same utterance; while a third, more familiar listener, may actually understand most of the individual words.

This report describes the development and initial testing of the Preschool Speech Intelligibility Measure (PSIM). The PSIM is a modification of Yorkston and Beukelman's (1980, 1981) single-word speech intelligibility test for dysarthric adults. The PSIM was designed to be a time efficient, reliable, and objective measure of the intelligibility of young children.

PSIM Development

Yorkston and Beukelman (1980) suggested using 600 single words in a multiple-choice format for assessing the intelligibility of adults with dysarthria. Within their procedures, they organized 600 test words into 50 sets of 12 phonetically similar forms. Before each testing session, the examiner compiles a unique 50-word stimulus set by randomly choosing one word from each of the original word sets. The examiner then obtains an audiotaped sample of a client's oral reading of the words and enlists a second examiner to identify the speaker's intended words, using the audiotape and a multiple-choice format. The format consists of all 600 test words divided into 50 sets of 12 words, a target word and 11 other phonetically similar words, with the percentage of target words correctly identified yielding a metric of speech intelligibility. The unique word list used for each test administration controls for listener familiarity of target words, allowing the same listener to serve on repeated occasions without biasing obtained measures.

In adapting Yorkston and Beukelman's protocol for preschool children, two primary concerns were addressed.

First, given that the original word lists were designed for use with adult speakers, some words might possibly be less familiar to young children than others, and this difference in familiarity might affect performance. As a means of addressing this concern, all 600 of Yorkston and Beukelman's words were compared to a published list of word frequencies in the speech of children aged 4 years 6 months to 5 years 0 months (Hall, Nagy, & Linn, 1984). Results of that comparison indicated that most of the 600 test words were rarely found in young children's speech. In fact, nearly half of the words were not even included in the published list of 6,550 most frequently used words. Furthermore, only 25 of the words that did appear had a reported frequency of greater than 300 per one million reported tokens. Thus, the 600 test words were taken to comprise a relatively homogeneous set of unfamiliar words for children.

Another concern in adapting Yorkston and Beukelman's measure was stimulus presentation. The targeted PSIM population was composed of nonliterate children, so reading the target words was not an option for PSIM administration as it had been for adults with dysarthria. Instead, Yorkston and Beukelman's (1981) alternate strategy of imitation was used. That is, the children were asked to repeat each of the 50 words following the examiner's model. In order to prevent the examiner's model from being recorded on the tape and biasing the listener, a custom-made microphone switch was designed. This switch interrupted the microphone input while allowing the tape to continue moving. Thus, by depressing the switch, the examiner was able to present her verbal model without recording it. Additionally, the switch provided the means to crudely control the interstimulus-interval of the resultant listening tape.

Piloting of the PSIM procedures was performed with 15 native English speaking children ranging in ability from phonologically impaired to normal speakers and included three different listeners. The piloting confirmed the feasibility of the modeling strategy and of the multiple-choice listening format for use with preschool children. However, an informal review of the listeners' responses indicated they were less likely to choose some foil words than others in 15 of the 50 word lists. As a result, 20 individual words were deleted from the set of 600 and replaced with words that were more phonetically similar to the other items in the respective 12-word list (Appendix A). Additional pilot testing using the revised word lists with seven different children who had similar speaking skills as the original 15 produced percentage intelligibility scores ranging from 26% to 81%, indicating that the range of sensitivity for the measure was appropriate for the preschool age group.

PSIM Testing

Nineteen native English speaking children, ranging in age from 3 years, 5 months to 5 years, 0 months, were selected for initial testing with the PSIM. Each of the children was enrolled in the Language Acquisition Preschool (LAP) at the University of Kansas. Criteria for enrollment in the preschool includes normal cognitive, motor, and social development (Rice & Wilcox, 1995). As shown in Table 1, the 19 children exhibited a range of speech and language skills.

TABLE 1. Standardized test scores and teacher ratings for the nineteen preschool children. Peabody Picture Vocabulary Test–Revised (PPVT-R) and the Reynell Developmental Language Scales results are reported as scaled scores, the Goldman-Fristoe Test of Articulation results are percentile scores, and the teacher rating is a nominal scale (Appendix C).

Subject	Chronological Age in Months	Gender	PPVT-R	Reynell Rec.	Reynell Exp.	GFTA	Teacher Rating[a]
1	56	M	116	113	90	35	1
2	52	M	105	114	104	64	1
3	57	M	94	106	113	76	1
4	49	F	119	104	107	89	0
5	54	M	109	108	98	32	0
6	60	F	94	112	101	13	1
7	53	F	106	108	95	57	1
8	59	M	99	87	70	14	3
9	59	M	105	106	94	19	3
10	60	M	102	110	102	42	1
11	48	M	122	88	72	5	3
12	57	M	63	72	85	19	2
13	58	M	126	122	94	5	3
14	45	M	88	98	73	16	2
15	48	M	84	66	63	3	3
16	58	M	109	118	73	<1	4
17	41	M	81	90	79	7	3
18	46	M	66	63	63	<1	4
19	41	M	90	90	101	14	2

[a]Scores range from 0 = intelligible to 4 = unintelligible.

Each child was administered the PSIM individually (see Appendix B for PSIM words). Testing began by instructing the child in a game of "copy cat" where the examiner says a word and then the child repeats it. The PSIM procedures were practiced with three familiar words (e.g., the child's name, a sibling's name, and a body part such as "nose") prior to the first stimulus item.

During the administration of the practice items, a cuing procedure, such as a nod or finger point, was introduced to prompt the child to say the word at a specific time. This cue was then used throughout the testing to reduce the possibility that the child would respond before the examiner turned on the microphone. Once the child understood the PSIM procedures, actual test items were presented. For each item, the examiner modeled the word with the microphone turned off. Then the examiner turned on the microphone prior to the child's production of the word.

In an attempt to standardize the time between stimulus presentations, the child and examiner counted to five between items. The counting, along with the examiner's verbal model (both unrecorded), allowed approximately a 7-second pause between stimulus items. During subsequent scoring, this time interval allowed listeners to scan the 12 words in the multiple-choice format and make their best guess as to the target word. As an additional aid to the listeners, the examiner recorded the stimulus number prior to each fifth item, e.g., "Now we are ready for number 10." The inclusion of the item number proved essential when listening to children who were unintelligible, where some productions did not sound like any of the 12 response options. The entire PSIM administration was completed within 15 minutes for each child and scoring was accomplished in 5–10 minutes.

TABLE 2. Percent intelligibility scores on the PSIM for nineteen preschool children.

Subject	Judge A	Judge B	Judge C	Mean
1	90	96	96	94.0
2	92	84	96	90.7
3	92	86	94	90.7
4	94	88	88	90.0
5	90	90	84	88.0
6	82	82	86	83.3
7	78	78	78	78.0
8	80	78	74	77.3
9	74	78	76	76.0
10	76	74	78	76.0
11	68	68	64	66.7
12	66	68	60	64.7
13	58	64	72	64.7
14	64	64	60	62.7
15	50	60	54	54.7
16	44	44	54	47.3
17	46	52	40	46.0
18	44	42	42	42.7
19	40	36	52	42.7
Mean	69.9	70.1	71.0	70.3

Following the test administration, three graduate students in speech-language pathology, who were familiar with the 600 PSIM words, scored the 19 tapes in counter-balanced order within a week. Average PSIM scores reported in Table 2 for the 19 children and 3 listeners ranged from 43% to 94%. Correlations between the judges' PSIM scores were high: Judge A–B, $r = .97$, $p > .01$; Judge A–C, $r = .94$, $p > .01$; Judge B–C, $r = .91$, $p > .01$.

PSIM Validity

Although intelligibility is a composite of many parameters, articulatory precision is central to the understandability of a spoken message. As such, the PSIM scores for the initial 19 subjects were compared to percentile scores on the Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986) and to teacher intelligibility ratings (Kansas Department of Education, 1988) to establish concurrent validity. All of the intelligibility ratings were provided by the LAP teacher. The LAP teacher is a speech-language pathologist and was familiar with all 19 children. All three measures, the PSIM, Goldman-Fristoe Test of Articulation, and the teacher rating, were obtained within a 1-week time period. Results indicated that the PSIM scores were highly correlated with the Goldman-Fristoe Test of Articulation percentile scores and the teacher ratings: PSIM–GFTA, $r = .73$, $p < .01$; PSIM–teacher rating, $r = -.78$, $p < .01$; which established concurrent validity between all measures.

PSIM Reliability

Fifty sets of 12 phonetically similar words yield 12^{50} possible 50-word combination tests. As a result, it is unlikely that even highly experienced listeners could correctly guess a child's responses based solely on their familiarity with the PSIM words. However, given the number of possible combinations, it is possible that not all of the 12^{50} possible tests forms are equal in their difficulty or sensitivity. Information on the test-retest reliability of the PSIM is pertinent to this issue and was obtained by administering the PSIM to 12 subjects aged 3 years 5 months to 5 years 9 months. These children were also enrolled in LAP and exhibited a range of speech and language abilities similar to those of the original 19 speakers. Each of the twelve children was administered two different randomized lists of fifty PSIM words within a 1-week period. Two graduate students in speech-language pathology, who were familiar with the 600 PSIM words, listened to all 24 of the administrations within a 2-week period, and comparisons were made between the two administrations (Table 3).

As shown in Table 3, the differences in PSIM scores between the first and second administrations ranged from 0 to 16 percentage points, Mean = 4.8%, for one judge; and from 0 to 14 percentage points, Mean = 6.3%, for the second judge. In 18 of the 24 cases, scores from the two PSIM administrations were within 6% (or three test items) of each other. A correlation coefficient indicated a positive correlation between the two randomized PSIM administrations for each of the judges: Judge 1, $r = .92$, $p < .01$; and

Judge 2, r =.94, $p < .01$.

Intrajudge reliability was assessed by having the first of these two listeners score each of the 24 tapes a second time 1 month after the initial scoring. This resulted in two separate comparisons, one from the 12 children's first PSIM administration and the second from the second PSIM administration (Table 4).

Difference scores from repeated ratings of the same PSIM administration by the same listener ranged from 0 to 6 percentage points or 3 items (Mean = 3.5%) for the first administration and 0 to 6 percentage points (Mean = 2.8%) for the second administration scores. Correlations between the two ratings of each of the two administrations were high: Administration 1, $r = .97$, $p < .01$; and Administration 2, $r = .98$, $p < .01$.

TABLE 3. Percent intelligibility scores for repeated administrations of the PSIM, scored by two judges.

	Listener One			Listener Two		
Subj.	Administration		Difference	Administration		Difference
	One	Two		One	Two	
1	66	70	4	66	68	2
2	76	90	14	82	82	0
3	72	56	16	68	58	10
4	84	80	4	84	90	6
5	94	96	2	96	92	4
6	38	38	0	38	32	6
7	48	54	6	48	38	10
8	96	96	0	94	96	2
9	76	70	6	70	58	12
10	88	86	2	88	84	4
11	94	92	2	94	88	6
12	82	84	2	68	82	14

TABLE 4. Percent intelligibility scores for repeated administrations of the PSIM. All judgments were made by the same listener.

	Administration One			Administration Two		
Subj.	Rating		Difference	Rating		Difference
	One	Two		One	Two	
1	66	70	4	70	70	0
2	76	82	6	90	92	2
3	72	72	0	56	50	6
4	84	90	6	80	86	6
5	94	90	4	96	94	2
6	38	42	4	38	38	0
7	48	44	4	54	48	6
8	96	98	2	96	96	0
9	76	70	6	70	68	2
10	88	94	6	86	86	0
11	94	94	0	92	88	4
12	82	82	0	84	78	6

Inter- and intrajudge reliability ratings for the PSIM were high, with all correlations being above .90. These high scores indicate that different judges give the same child similar PSIM scores when listening to the same sample. Therefore, if it is not possible to have the same listener score pre- and post-test measures, gains in single-word intelligibility as measured by the PSIM should not be greatly affected when two different listeners are used. Additionally, listeners score children similarly on different forms of the PSIM, confirming the appropriateness of using multiple forms of the measure when assessing a child's progress across time.

Discussion

This paper describes the development and initial testing of the Preschool Speech Intelligibility Measure, designed to provide an objective measurement of the single-word intelligibility of preschool children with a wide range of speech-language abilities. Continued use of the PSIM is supported by several of the findings of this project. First, all 43 children in the piloting and initial testing actively participated in the elicitation task. The children understood the imitation task and the cuing procedure after only three trial items and participated with little need for encouragement or adjustment in the procedure. Perhaps counting aloud during the interstimulus interval assisted in keeping the child engaged in the task for the entire test administration. Second, the adult listeners produced relatively unbiased intelligibility scores from the resultant tapes. The phonetic similarity of the words in each 12-item set, as well as the numerous test forms available, made it difficult for listeners to simply "guess" the child's response. Instead, the listener had to concentrate on the child's productions and make judgments concerning their intelligibility. Third, the procedures yielded high inter- and intrajudge reliability ratings. All correlations of PSIM scores between and within judges were above .90.

The results reported here may be taken to indicate a relatively restricted range in PSIM scores in this group, as there were no scores below 43%. However, more recent use of the PSIM in clinical settings with younger children who have phonological impairments and older children with Down syndrome, hearing impairment, and other handicapping conditions indicate that some children do score lower on the PSIM, some as low as 7%. Therefore, the PSIM does appear to test a large range of intelligibility scores in children with a variety of speech impairments.

An initial concern during the development of the PSIM was the use of unfamiliar words as a means of measuring intelligibility in young children, although homogeneity of the words is probably more important for assessment purposes than familiarity. The advantage of using unfamiliar words is that a much larger set of words is available for selection of stimuli. Additionally, concern regarding children's ability to correctly articulate unfamiliar words is diffused by the strong correlations found between the PSIM and the Goldman-Fristoe Test of Articulation, $r = .73$, $p < .01$. This high correlation indicates that children who have difficulty correctly producing familiar words on the Goldman-Fristoe also have difficulty producing unfamiliar words on the PSIM.

For years, investigators have debated the validity of single-word measures in terms of their ability to represent performance on connected speech tasks (Yorkston & Beukelman, 1978; Healy & Madison, 1987; Morrison & Shriberg, 1992; Weston & Shriberg, 1992). Taken together, the high correlations obtained between the PSIM and both the Goldman-Fristoe Test of Articulation and the teacher ratings of intelligibility confirm the validity of the PSIM in terms of both single-word data and informed judgments based on long-term exposure to a child's speech. Moreover, the PSIM appears to extend rather than merely duplicate articulation information. For example, subjects 2, 3, and 4 all achieved mean scores of 90% on the PSIM, whereas their Goldman-Fristoe percentile scores varied by 25 points (64, 76, and 89, respectively.) A similar pattern is observed between the PSIM and the teacher rating scale. For example, subjects 8, 9, 11, 13, 15, and 17 were all rated 3 on the teacher rating scale. These five children show a substantial range on the PSIM: 77%, 76%, 67%, 65%, 55%, and 46%, respectively. The increased range in the PSIM relative to the teacher ratings allows clinicians to assess a child's progress with more sensitivity than the teacher rating scale provides, which is mainly used as a judgment of severity and not as a means to plot progress.

Shriberg and his colleagues (Kwiatkowski & Shriberg, 1992; Morrison & Shriberg, 1992; Weston & Shriberg, 1992) have argued for the use of connected speech samples in assessing speech intelligibility because of the inherent limitations of single-word productions for representing natural, communicative productions. In large part, we agree with this assessment. However, conversational speech samples also have limitations. The primary limitation is lack of control of the speech material for assessment. Conversational speech, by its very nature, does not allow one to predict or control the targeted items. As a result, post-test situations are never identical to pretest situations. Further, in most cases, the speech samples must be transcribed by a trained professional, a process which is labor intensive and may be influenced by listener familiarity.

Although it is appealing to envision a single speech intelligibility measure with the power to represent a speaker's overall communicative effectiveness, the inherent complexity of intelligibility makes it unlikely that a single measure will emerge in the foreseeable future that is appropriate for all speakers, listeners, messages, and contexts. Instead, those interested in measuring speech intelligibility must identify their specific evaluation aim or purpose. If the purpose is to describe a child's natural speech relative to her peers, then starting from a connected speech sample may be the best choice. However, for other purposes, a different corpus may be equally or more effective. Our belief is that the PSIM has potential as a within-subject measure of change in intelligibility across time. The control obtained by the 50-word elicitation format allows for the direct comparison of different administrations of the test to the same child. Since each administration uses different words, there should be little learning effect for either speaker or listener across administrations. In fact, Yorkston and Beukelman (1981) suggest the use of listeners who are already familiar with the protocol and the 600 words in order to reduce test-retest variability of the measure. This type of control is difficult to achieve for measures derived from conversational samples.

It is likely that the best assessment of phonological abilities in young children consists of a composite of individual measures. The use of multiple measures in the assessment of young children allow speech-language pathologists to assess different aspects of the phonological system and determine the best possible treatment plan for a specific child. Measures such as the PSIM are an appealing addition to current available measures in that they provide speech-language pathologists with an objective and systematic measure of intelligibility and assess the child's ability to imitate unfamiliar words, a skill not assessed by most other tests. Additionally, the ease and quickness of PSIM administration and scoring make it an appealing addition to the speech-language pathologist's testing protocol.

Acknowledgments

The authors would like to thank Betty Bunce, Angela Rodakavich, and Jennifer DeVault for their assistance in data collection and analysis. Additionally, we thank Kathy Siren and Mabel Rice for their insightful contributions during the initial development of this measure and Peter Flipsen, Jr. and Audrey Weston for their comments on an earlier version of this paper. We are also grateful to the children and listeners who participated in this study. This work was supported by NIH grant # HO29D90046 and KECRI grant # HO24U80001.

References

Goldman, R., & Fristoe, M. (1986). *Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service.

Hall, W. S., Nagy, W. E., & Linn, R. (1984). *Spoken words: Effects of situation and social group on oral word usage and frequency*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Healy, T., & Madison, C. (1987). Articulation error migration: A comparison of single word and connected speech samples. *Journal of Communication Disorders*, *20*, 129–136.

Kansas Department of Education. (1988). Articulation. *Kansas state guidelines and severity rating scales for speech and language impairments*. Topeka, KS: Author.

Kent, R. D, Miolo, G., & Bloedel, S. (1992, November). *Measuring and assessing the intelligibility of children's speech*. Paper presented at the meeting of the American Speech-Language-Hearing Association, San Antonio, TX.

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*(2), 81–93.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, *54*, 482–499.

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.

Milisen, R. (1971). The incidence of speech disorders. In L. E. Travis (Ed.), *Handbook of Speech Pathology and Audiology* (pp. 619–633). New York: Appleton-Century-Crofts.

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*, *35*, 259–273.

Peterson H. A., & Marquardt, T. P. (1994). Articulation testing.

Appraisal and diagnosis of speech and language disorders (3rd ed., p. 59). Engelwood Cliffs, NJ: Prentice-Hall.

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

Shames, G. H., Wiig, E. H., & Secord, W. A. (1994). *Human communication disorders: An introduction* (4th ed., p. 11). New York: Macmillan.

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

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

Yorkston, K. M., & Beukelman, D. R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders*, *13*, 15–31.

Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of intelligibility of dysarthric speech.* Austin, TX: ProEd.

Received July 25, 1994
Accepted May 10, 1995

Contact author: Sherrill R. Morris, Department of Speech-Language-Hearing, 3031 Dole, University of Kansas, Lawrence, KS 66045

Key Words: preschool, speech intelligibility, clinical assessment

Appendix A

Changes Made to Word Lists

#9 “port” to “torque” (“Port” was reported twice in list nine of the Yorkston and Beukelman (1980) word list.)

#20 “owl” to “Paul”

#22 “creature” to “sweet”

#24 “left” to “lick,” and “less” to “slick”

#26 “super” to “defers”

#28 “so” to “or”

#30 “glib” to “grace”

#31 “praise” to “raced”

#35 “tired” to “bear” and “cart” to “court”

#36 “ought” to “auction,” “off beat” to “otter,” and “offsides” to “awesome”

#38 “below” to “clone”

#40 “closure” to “picture”

#42 “decal” to “delight”

#43 “leaf” to “lid”

#45 “grind” to “grand”

#47 “gang” to “gate”

Appendix B

Preschool Speech Intelligibility Measure Word List

1)	2)	3)	4)	5)
warm	rate	deer	whoop	bitten
store	train	near	boot	hitting
swarm	trait	peer	droop	pitting
for	race	we’re	suit	city
horn	rave	fear	toot	fitting
corn	range	beer	loop	sitting
door	rage	mere	coupe	kitten
torn	trade	steer	snoop	pretty
born	rain	hear	group	mitten
floor	trace	tear /tɪr/	fruit	knitting
storm	rake	cheer	root	written
form	raid	shear	troop	witty

6)	7)	8)	9)	10)
spark	nest	wicker	sort	stretch
heart	said	sicker	court	catch /kɛtʃ/
hark	tread	flicker	cork	said
barn	best	sitter	quart	shread
harp	dread	liquor	port	sketch
part	red	wicked	sport	shed
darn	dead	bitter	torque	dread
dark	bed	ticker	tort	thread
park	dress	litter	fort	threat
cart	rest	glitter	snort	fed
start	bread	quicker	short	red
dart	Ned	slicker	fork	bread

11)	12)	13)	14)	15)
neat	center	born	waste	saddle
Pete	tender	short	why	naval
seam	bumper	wart	weave	gable
beast	fender	horn	wide	table
reap	lender	corn	wives	cable
seat	temper	court	wine	fable
reef	renter	term	with	able
beef	timber	sort	win	rattle
beep	member	form	weep	Mabel
beam	ember	torn	wise	sable
beat	bender	tort	whip	ladle
meat	sender	burn	wipe	stable

16)	17)	18)	19)	20)
tap	boot	fool	pale	tall
bat	bash	panel	paid	stall
vamp	boat	phone	space	wall
sat	bag	fill	page	call
mat	bog	funnel	Spain	shawl
ramp	back	fell	paint	all
rat	bat	fall	sprain	fall
nap	big	find	pain	ball
vat	bought	feel	spray	hall
damp	beg	file	paste	crawl
sap	ban	final	spade	mall
map	bank	foil	pace	Paul

21)	22)	23)	24)	25)
told	beat	train	slick	judge
pull	reach	rain	slip	musk
pole	street	brain	lip	rust
cold	seat	pain	lit	but
bowl	peach	strange	limb	just
mole	sheet	grade	slid	shut
full	each	grain	lick	dusk
fold	neat	rage	list	fudge
bull	teach	range	lid	bust
sold	sweet	gauge	slim	must
soul	preach	sane	slit	rut
mold	screech	grange	lisp	dust

26)	27)	28)	29)	30)
absurd	retract	score	bone	groan
obscure	contrast	torch	boat	grain
reserve	repress	tore	Boone	grope
observe	contract	pour	both	gross
deserve	detract	or	bin	grove
serve	suppress	store	bit	grape
preserve	contact	wore	boast	gleam
occur	compress	more	boot	grave
conserve	subtract	porch	bill	globe
refer	impress	court	booth	grace
converge	intact	scorch	book	glow
defers	depress	shore	bowl	grow

31)	32)	33)	34)	35)
wait	quit	twice	shop	par
rake	sit	wire	shoot	pear
grace	swift	wise	shelf	carp
trace	sip	wide	shoe	car
brace	stiff	wipe	ship	pour
waste	knit	quite	shark	tore
raced	mist	wife	shell	court
rate	whip	twine	shock	tarred
trait	list	wine	shore	carve
pace	slip	white	short	bear
race	wit	why	shot	tear /tɛr/
wake	twist	while	sharp	care

36)	37)	38)	39)	40)
option	sauce	pole	take	creature
often	Ross	soul	lake	picture
otter	law	coal	rake	denture
awful	crawl	blown	mate	rupture
office	flaw	clone	bait	rapture
honest	call	low	late	fixture
author	loss	glow	cape	lecture
awesome	claw	doe (dough)	tape	teacher
awkward	cross	blow	ape	pasture
after	cough	sew (so)	hate	preacher
auction	cloth	bowl	make	mixture
offer	boss	flow	rate	texture

41)	42)	43)	44)	45)
tide	debate	leash	tanner	grand
sigh	delight	lease	planner	green
ride	design	lit	matter	scrub
lie	define	leak	manner	screech
side	defend	lip	mother	stream
high	divine	least	banner	scream
hide	deny	leave	butter	grab
why	decide	live /lɪv/	brother	street
die	defy	lid	sander	scram
eye	divide	limb	bother	scrap
tie	devise	leap	batter	scratch
by	defeat	lean	other	screen

46)	47)	48)	49)	50)
lark	grain	raid	cape	spear
art	grab	made	tape	weird
ark	gauge	faint	pave	seer
spark	glow	vague	cage	beer
march	grape	pay	case	fierce
dark	great	paid	page	pier
park	glass	may	taste	steer
bark	game	day	pace	near
arch	gate	ray	cake	hear
tart	gain	fade	cave	leer
heart	grow	shade	cane	pierce
dart	gag	jade	take	we're

Appendix C

Teacher Rating Scale
Kansas Department of Education, 1988

Please circle the number which most likely describes the intelligibility of the child named above.

0) Conversational speech reflects standard adult patterns.

1) Conversational speech contains some sound production differences. Sound production differences are developmentally appropriate. Spontaneous development of standard adult phoneme production is expected. No therapy indicated.

2) Conversational speech is intelligible although noticeably in error. Use of articulatory shift processes which are inappropriate for age such as voicing deviations, deaffrication, frontal distortions. Sound production reflects common types of distortions or substitutions of later- developing phonemes. Sound productions appear to be developing in normal progression, although delayed up to one year. Sound productions may vary with phonetic context, indicating that spontaneous phoneme development could occur. Periodic observation, but no therapy, indicated unless the child is eight years old, or older.

3) Conversational speech contains words and phrases which are not intelligible. Excessive use (40% or more) of substitution or omission processes which are inappropriate for age, such as velar deviations, stridency deletion, and cluster reduction. (As prevalence of phonological processes increases the resulting decrease in intelligibility would warrant a more severe rating.) Sound productions are not developmentally appropriate and are delayed more than one year. (A delay in phoneme development of more than one year may warrant a rating of "3" despite no decrease in intelligibility.) Spontaneous development of standard adult phoneme production is not expected. Therapy indicated.

4) Conversational speech is intelligible only with knowledge of the context and familiarity with the pupil and pupil's sound system. Excessive use (40% or more) of omission processes (such as syllable reduction, prevocalic and postvocalic sound deletion) or unique processes which are inappropriate for age. Sound productions reflect use of limited number of phonemes or phoneme classes. Sound productions are not developmentally appropriate and are delayed more than one year. Spontaneous development of standard adult phoneme production is not expected. Therapy indicated.