Published in final edited form as:

Clin Linguist Phon. 2006 ; 20(0): 553–561. doi:10.1080/02699200500266455.

The clinical utility of nonword repetition for children living in the rural south of the US

Janna B. Oetting and Lesli H. Cleveland

Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, LA, USA

Abstract

Nonword repetition (NWR) tasks have been shown to minimize cultural biases in language assessment. In the current study, we further examined the clinical utility of NWR with 83 children who lived in the rural south of the US; 33 were African American and 50 were White, with 16 classified as specifically language impaired (SLI) 6-year-olds and 67 classified as either age-matched or younger controls. Main effects were found for group, with the children in the SLI group earning lower NWR scores than the controls. A main effect for syllable length but not race was also documented. The group and syllable length effects could not be explained by differences in the children's articulation abilities or by potential differences in the children's use of vernacular dialect. Discriminant analysis indicated that NWR had a diagnostic accuracy rate of 81% for the 6-year-olds, but sensitivity was low (56%). When combined with scores from one other nonbiased assessment tool, however, the diagnostic accuracy of NWR increased to 90%, with rates of sensitivity and specificity above 80%.

Keywords

Nonword repetition; test biases

Introduction

Language acquisition is linked to a child's cultural upbringing. Unfortunately, most standardized tests that are available to speech-language clinicians have not been developed with socio-culturally neutral (i.e., un-biased) materials. Thus, it is not surprising that when children from minority backgrounds score lower than expected on a standardized tool, cultural biases within the test are often suggested as a possible explanation (Kayser, 1995; Pena & Quinn, 1997; Sattler & Altes, 1984; Teuber & Furlong, 1985, Washington & Craig, 1992). Although many different types of cultural biases can occur within an assessment, the two most often discussed relate to a test's content and its format. Content biases occur when items on a test probe for information from a child that it is inappropriate or unfamiliar to the child's culture. Format biases occur when the unfamiliar or inappropriate aspect of the test relates to the manner in which the test is constructed and/or administered.

The goal of the current study was to evaluate the clinical utility of an experimental task that asks children to repeat nonsense words. Tasks that involve nonword repetition (NWR) are thought to minimize the effects of test content biases and test format biases by making some

aspects of the task familiar to all children while making other aspects of the task equally unfamiliar. The familiar aspect of NWR relates to the phonemic content of the test items. Unlike real words, NWR items can be created with phonemes that are present in a number of languages and dialects. The unfamiliar aspects of NWR relate to the semantic and grammatical content of the test items and the manner in which the test is administered. As nonsense words, they can be created so that they (and the syllables within them) do not carry meaning or grammatical function. In addition, the task of repeating nonsense words is thought to be relatively unfamiliar to all children, regardless of their cultural background. Some scholars further describe NWR as assessing children's processing abilities rather than assessing their knowledge of specific language content (Campbell, Dollaghan, Needleman, & Janosky, 1997).

Although NWR tasks have now been included in a number of experiments, results from five demonstrate the potential of this tool for clinical practice. The first was completed by Campbell et al. (1997). This study included 156 children between the ages of 11 and 14 years, with 31% classified as having majority racial status and 69% classified as having minority status. Among other probes, each child was given one standardized language test, the Oral Language Scale (OLS) from the *Woodcock Language Proficiency Battery-Revised* (Woodcock, 1991) and an NWR task. Results indicated that unlike the OLS scores, the NWR scores did not differ as a function of the children's minority vs. majority racial status.

In four other studies, the clinical usefulness of NWR was examined by comparing the scores of children with language impairments to those of typically developing controls. Dollaghan and Campbell's (1998) study included 40 children who were between the ages of 6 and 9 years, Ellis Weismer, Tomblin, Zhang, Buckwalter, Chynoweth and Jones (2000) included 581 children who were 7, Conti-Ramsden, Botting, and Faragher (2001) included 260 children who were 10, and Conti-Ramsden (2003) included 64 children who were 5. In each of these studies, the children with a language impairment earned lower NWR scores than the controls. Moreover, when Conti-Ramsden and colleagues examined NWR within a discriminant function analysis, it showed fair diagnostic predictability (82%) for 10-year-olds, with a sensitivity rate of 78% and a specificity rate of 87%. For 5-year-olds, the specificity of NWR was also good (100%) but its sensitivity was limited (59%).

In the current work, we further tested the clinical utility of NWR for a group of 4- and 6-year-olds who lived in the rural south of the US. With almost 20% of the children classified as specifically language impaired (SLI), the data allowed us to examine the usefulness of NWR for distinguishing those with impairments from those without. In addition, with 40% of the children identified as African American (AA), the data allowed us to examine whether NWR contained a cultural bias that related to the children's race. The questions guiding the research were: (1) Does NWR distinguish children with SLI from children with typically developing language skills? and (2) does NWR contain a cultural bias that relates to the children's race?

Methods

Participants

Eighty-three children living in a rural area of Louisiana contributed data to the study. There were 33 AA children and 50 White (W) children. All of the children were judged to speak a nonmainstream dialect of English by at least two examiners. Although other, more formal, measures of nonmainstream dialect use were not collected for these children, they were recruited from the same schools as those studied by Oetting and McDonald (2001, 2002). In our earlier studies, all children who were recruited from these schools were documented to speak a variety of either southern White or southern African American English using a blind

listener judgment task and/or detailed coding of vernacular morphology (see Oetting & McDonald, 2002).

Sixteen children were classified as SLI, 36 were classified as typically developing 6-year-olds (6N), and 31 were classified as typically developing 4-year-olds (4N). The criteria for SLI were: (a) currently enrolled in speech-language services in the public schools, (b) designated as exhibiting language skills below his or her peers as determined by the classroom teacher, (c) performed within one standard deviation of the mean on the *Columbia Mental Maturity Scale* (CMMS; Burgmeister, Blum, & Lorge, 1972), (d) performed below one standard deviation of the mean on the PPVT-R (Dunn & Dunn, 1981) and on the syntactic quotient of the *Test of Language Development-Primary* (TOLD-P: 2; Newcomer & Hammill, 1988), (e) did not demonstrate frank neurological impairments or social-emotional deficits per teacher report, and (f) passed a hearing screening within 6 months of the study.

The children in the typically developing groups were recruited from the same schools and classrooms (or day cares and preschools within close proximity to the schools) as those with SLI. All of these children were considered typically developing based on teacher report and none had a history of speech or language impairments. Originally, these participants were recruited as controls for two studies that examined children's word learning abilities, but only 36 of them met the subject selection and subject matching criteria of those studies (see Horohov & Oetting, 2004; Oetting, 2003). Of those who were not included in the vocabulary studies, 25 earned a standard score that was below one standard deviation on either the PPVT-R ($n=16$), TOLD-P: 2 ($n=16$), or CMMS ($n=6$). For the PPVT-R and TOLD-P: 2, these results reflect a 21% fail rate by the controls, a finding that is consistent with previous descriptions of these tools as culturally biased.

Participant profiles are provided in Table I. Also included are standardized z scores and corresponding group percentile scores from the *Goldman-Fristoe Test of Articulation* (GFTA; Goldman & Fristoe, 1986) and standard scores from the Comprehension Subtest of the *Stanford-Binet* (CSSB; Thorndike, Hagen & Sattler, 1986). The GFTA was originally given to the children for descriptive purposes, and the CSSB was given as part of a master's thesis by the second author (Habans, 2000). None of the controls scored below the 10th percentile on the GFTA; six scored one standard deviation on the CSSB. Given the ordinal nature of the percentile scores on the GFTA, each child's percentile on this test was averaged after it was first converted to a z score using a standardization table from Hinkle, Wiersma, & Jurs (1998, p. 633). The GFTA percentiles reported in Table I correspond to each group's averaged z score.

Stimuli

The stimuli were Dollaghan and Campbell's (1998) 16 nonce words (see Table II). Four of the words were one syllable in length while equal numbers of the others included two, three, and four syllables. The nonce words were created with 11 different consonants and nine vowels (three monophthongs and six diphthongs). These phonemes did not include lax vowels or consonants identified as the late eight by Shriberg and Kwiatkowski (1994). All of the words also began and ended with a consonant but not a consonant cluster, and none of the CV or CVC syllables within the words corresponded to an English word. These particular features of the stimuli reduced the risk of the NWR results being confounded by developmental differences in the children's phonological systems or by possible dialectal differences in the children's use of English. Indeed, of the 45 non-mainstream phonological patterns of English that have been identified in the literature (see Bailey, 2001), only three were possible with Dollaghan and Campbell's stimuli. These included glide reduction of /aɪ/

and /ɔɪ/, deletion of unstressed initial and medial syllables, and final consonant devoicing of /b/ and /g/.

The stimuli were recorded onto the audio track of a video-tape. The tape was edited to show a 5- to 7-second blue screen between the words and a flashing white bar in the corner of the screen before each word. Children were asked to listen to each word and repeat it exactly as they heard it. The children's responses were audio recorded for later transcription and scoring. Following Dollaghan and Campbell, NWR was scored by calculating each child's percent of phonemes correct for each syllable length and for the total set of nonce words. Phoneme omissions and substitutions were marked as errors. Phoneme additions were transcribed but not scored.

Reliability

NWR responses from 23 children were independently transcribed and scored by a second examiner (96 phonemes per NWR task X 23 children=2208). Rate of agreement at the phoneme level was above 95% (SLI=95.6%, 6N=96.4%, 4N=96.6%).

Results

Analyses of variance

NWR scores were examined with a mixed ANOVA with group and race as the between-subjects variables and syllable length as the within-subjects variable. Only main effects for group and syllable length were significant; group $F(2, 77)=16.12$, $p<.001$, partial $\eta^2=.30$, and syllable length $F(3, 231)=30.02$, $p<.001$, partial $\eta^2=.28$ (see Table III). These main effects remained even when a group by syllable length ANOVA was rerun with the children's standardized z scores from the GFTA used as a covariate.

The group main effect was explored with Tukey t-tests and these indicated that scores of the children with SLI were significantly lower than both groups of controls. The syllable length main effect was explored with six paired t-tests. These comparisons indicated that scores for the one and two syllable lengths were significantly higher than the scores of the three and four syllable lengths; one vs. three $t(82)=3.53$, $p<.001$, one vs. four, $t(82)=6.82$, $p<.01$, two vs. three $t(82)=5.23$, $p<.001$, two vs. four $t(82)=8.13$, $p<.001$. The three syllable length score was also significantly higher than the four syllable length score, $t(82)=5.57$, $p<.001$.

Discriminant analysis

Discriminant analysis is another way to evaluate the utility of a measure for classifying children from different diagnostic groups. Following the work of Conti-Ramsden (2003) and others, NWR was examined with discriminant analysis on the 52 6-year-olds in the sample. The 4N controls were not included in this analysis because within clinical practice, a diagnosis of language impairment is typically based on the age of the child rather than on language expectations of a younger child. A discriminant function with NWR resulted in 81% of the children being accurately classified as either SLI or typically developing. Although the sensitivity (i.e., the rate at which children with SLI were identified as impaired) of NWR was low at 56%, the specificity (the rate at which children in the 6N group were identified as typical) was high at 92%. In addition, the misclassified children were not disproportionately AA or W. For both racial groups, two children with SLI were misclassified as a 6N case, and for the misclassified 6N children, three were AA and four were W.

To determine whether we could improve the diagnostic accuracy of NWR, we also ran additional discriminant analyses that included other test scores from the battery. Given that

the PPVT-R and TOLD-P2 were used to select all of the children in the SLI group and some of the children in the control groups, scores from these tests were excluded from consideration. Of those remaining, moderate correlations between them and the NWR totals were found (NWR and CSSB: $r=.56$, $p<.01$; NWR and z scores of GFTA: $r=.60$, $p<.01$; CSSB and GFTA: $r=.34$, $p<.05$). A discriminant analysis with NWR and CSSB led to the highest level of diagnostic predictability, correctly classifying 90% of the children (sensitivity=81% and specificity=94%). No other tool or combination of tools of the three listed above (NWR, CSSB, z scores from the GFTA) achieved this level of diagnostic accuracy.

Item analysis

The final analysis examined the NWR error patterns of the children. At the level of the individual, none produced errors that were solely omissions or substitutions. Also, because all of the phonemes except /dʒ, k, u/ were repeated within the stimuli, we were able to confirm that every child correctly produced each phoneme at least once within the stimuli. For the children who made an omission or substitution error with /dʒ, k, u/, their ability to produce these targets was confirmed through visual inspection of the GFTA raw data.

We also searched the data for the three types of error patterns that may have been related to the children's use of a non-mainstream dialect of English (either southern White or southern African American). Seventeen children produced 20 monophthongal variants of /aɪ/ and /ɔɪ/. The SLI group produced five of these patterns, the 6N group produced 11, and the 4N group produced four, with the AA children producing 14 and the W children producing six. With a total of 1245 /aɪ/ and /ɔɪ/ contexts within the data (15 vowels $\times$ 83 children), a finding of 20 indicates that glide weakening of these diphthongs was rare (<2%). Syllable deletion of unstressed initial and medial syllables was also rare as only two instances of this pattern were found in the children's responses. Finally, for final consonant devoicing of /b/ and /g/, 12 examples were found. The SLI group produced one of these examples, the 6N group produced five, and the 4N group produced six, with the AA children producing eight and the W children producing four. With 747 final /b/ and /g/ contexts within the data (9 $\times$ 83 children), a finding of only 12 devoiced productions indicates that this pattern, like the others, was rare (<2%).

At the level of the groups, the phonemes that received the greatest numbers of errors were highly similar across the children. To illustrate this finding, Table IV lists the phonemes on which an error was produced by 20% or more of the children in each group. This table was generated using the children's responses from the first eight words of the stimuli since most of the phonemes (all but /i/ and /u/) within the entire set were included within these items. As can be seen, the rank ordering of these phonemes, from most often in error to least, was relatively similar across the groups.

Finally, for each of the phonemes listed in Table IV, an error analysis was completed. In each case, the patterns of error across the groups were similar. For example, the final consonant /b/ in /naɪb/ was most frequently produced as a /v/ for all three groups, with omissions and substitutions involving /f, d, t/ also occurring infrequently. For medial /tʃ/ in /vætʃaɪp/, the errors were primarily substitutions involving /t/ and /s/, with omissions and substitutions involving /dʒ, θ, ð, ʒ, ʃ/ occurring infrequently. Finally, for the diphthong /ɔɪ/ in /nɔɪtəʊf/, the most frequent error pattern was the production of /oo/, with substitutions involving /aɪ/ and /i/ occurring infrequently. None of these error patterns or others within the data suggested that the NWR results were directly tied to a developmental difference in the children's phonological systems. Moreover, none of these error patterns suggested a racial and/or possible non-mainstream English dialect bias within the stimuli.

Discussion

Results indicated that NWR is affected by a child's diagnostic classification but not a child's race. These findings are consistent with others that have examined NWR, but the current study is the first to document both of these findings in the same study with children who are as young as six years of age. This study is also the first to report these findings while also examining and ruling out phonological (as measured by single word articulation skill) and dialectal (as measured by non-mainstream pattern use) confounds within the results at the level of the individual items. Nevertheless, the results also indicated that for children who are 6-years-old, NWR should not be used as the sole indicator of a child's language ability. Although a discriminant function involving NWR accurately classified 92% of the controls, the classification accuracy of the children with SLI was only 56%. This low sensitivity rate is similar to Conti-Ramsden's (2003) rate of 59% for 5-year-olds. Sensitivity rates in the 50% range are too low to be useful in clinical practice.

When we combined NWR with scores from one other tool, however, the diagnostic predictability of the discriminant function yielded a fair sensitivity rate (81%) and a good specificity rate (94%). Conti-Ramsden (2003) also reported rates of sensitivity and specificity that were above 80% for 5-year-olds when NWR was combined with scores from a second language measure. Thus, the results from of these studies support the use of NWR as long as it is administered with other tools within an assessment.

The results of Conti-Ramsden's research and ours underscore the need for future research to examine the clinical utility of combining different testing batteries within an assessment. In the current study, the second tool included within the discriminant function was the CSSB, whereas in Conti-Ramsden's study, the second tool was a measure of regular past tense marking. Although items on the CSSB are very different from a probe of past tense, the use of these tools within the individual studies was based on previous findings. The CSSB was examined here because this tool has been shown to be culturally appropriate for US children from minority backgrounds (Pena & Quinn, 1997). For Conti-Ramsden's children who lived in the UK, a number of studies in the US and UK have implicated past tense marking as a clinical marker of SLI for children who are Standard English speakers (Tager-Flusberg & Cooper, 1999). Whether similar or dissimilar combinations of tools are needed for the different types of child language learners who live in the US, UK, and elsewhere is an important topic for future study. To examine this issue, cross-validation studies should be completed first to examine the robustness of our findings and those of others.

Acknowledgments

The project was made possible by a LEQSF grant from the LA Board of Regents and an RO3 grant from NIDCD awarded to the first author. The authors would like to thank the teachers, parents and children who participated in the research, and Lesley Eyles, Anita Hall, and Karen Lynch for help with different aspects of the study.

References

- Bailey, G. The relationship between African American vernacular English and White vernaculars in the American South: A sociocultural history and some phonological evidence. In: Landhart, S., editor. Sociocultural and historical contexts of African American English. Philadelphia, PA: John Benjamins; 2001.
- Burgmeister, B.; Blum, H.; Lorge, I. Columbia Mental Maturity Scale. San Antonio, TX: Psychological Corp.; 1972.
- Campbell T, Dollaghan C, Needleman H, Janosky J. Reducing bias in language assessment: Processing dependent measures. *Journal of Speech, Language, and Hearing Research*. 1997; 40:519–525.

- Conti-Ramsden G. Processing and linguistic markers in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2003; 46:1029–1037.
- Conti-Ramsden G, Botting N, Faragher B. Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry*. 2001; 42:741–748. [PubMed: 11583246]
- Dollaghan C, Campbell T. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1136–1146.
- Dunn, L.; Dunn, L. Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Service; 1981.
- Ellis Weismer S, Tomblin J, Zhang X, Buckwalter P, Chynoweth J, Jones M. Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*. 2000; 43:865–878.
- Goldman, R.; Fristoe, M. Goldman-Fristoe Test of Articulation. Circle Pines, MN: American Guidance Services; 1986.
- Habans, L. Unpublished master's thesis. Baton Rouge: Louisiana State University; 2000. Use of the Stanford-Binet in language assessment of southern African American and white children.
- Hinkle, D.; Wiersma, W.; Jurs, S. Applied statistics for the behavioral sciences. Boston, MA: Houghton Mifflin Company; 1998.
- Horohov J, Oetting J. Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics*. 2004; 25:43–67.
- Kayser, H. Intervention with children from linguistically and culturally diverse backgrounds. In: Fey, M.; Windsor, J.; Warren, S., editors. *Language intervention: Preschool through the elementary years*. Baltimore, MD: Paul H. Brookes; 1995. p. 315–331.
- Newcomer, P.; Hammill, D. Test of Language Development-Primary: Second Edition. Austin, TX: ProEd; 1988.
- Oetting, J. Children's use of prepositions to learn verbs. Baton Rouge: Louisiana State University; 2003. Unpublished manuscript
- Oetting J, McDonald J. Non-mainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2001; 44:207–223.
- Oetting J, McDonald J. Methods for characterizing participants' non-mainstream dialect use within studies of child language. *Journal of Speech Language Hearing Research*. 2002; 45:505–518.
- Pena E, Quinn R. Task familiarity: Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in the Schools*. 1997; 28:323–332.
- Sattler J, Altes L. Performance of bilingual and monolingual Hispanic children on the Peabody Picture Vocabulary Test-Revised and the McCarthy Perceptual Performance Scale. *Psychology in the Schools*. 1984; 21:313–316.
- Shriberg L, Kwiatkowski J. Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research*. 1994; 37:1100–1126. [PubMed: 7823556]
- Tager-Flusberg H, Cooper J. Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research*. 1999; 42:1275–1278.
- Teuber J, Furlong M. The concurrent validity of the Expressive One-Word Picture Vocabulary Test for Mexican-American children. *Psychology in the Schools*. 1985; 22:296–273.
- Thorndike, R.; Hagen, E.; Sattler, J. Stanford-Binet Intelligence Scale-Fourth Edition. Chicago, IL: Riverside; 1986.
- Washington J, Craig H. Performances on low-income African American preschool and kindergarten children on the Peabody Picture Vocabulary Test-Revised. *Language, Speech, and Hearing Services in the Schools*. 1992; 23:329–333.
- Woodcock, RW. Woodcock Language Proficiency Battery-Revised. Allen, TX: DLM Teaching Resources; 1991.

Table 1

Description of participants.

	SLI			6N			4N		
	AA	W		AA	W		AA	W	
Number per group	6	10		12	24		15	16	
Age in months	75.17 (5.70)	73.30 (6.34)		71.08 (4.66)	71.54 (4.76)		56.60 (6.92)	54.50 (7.70)	
CMMS ^a	94.50 (5.75)	92.60 (3.95)		102.92 (11.81)	103.04 (9.00)		92.17 (9.95)	101.38 (15.29)	
PPVT-R ^b	71.67 (4.18)	77.10 (5.92)		90.08 (9.60)	101.67 (14.21)		82.73 (13.06)	94.62 (11.86)	
TOLD-P:2 ^c	74.00 (7.34)	71.80 (10.79)		93.42 (5.51)	99.50 (12.35)		87.87 (8.93)	96.06 (9.56)	
GFTA ^d	1.21 (1.33)	-.64 (1.25)		.84 (.85)	1.77 (.78)		.66 (.31)	.68 (.88)	
Average z score									
Percentile corresponding to average z score	89	27		80	96		74	75	
CSSB ^e	83.33 (14.29)	84.80 (4.92)		103.67 (11.90)	108.83 (14.04)		91.20 (8.51)	100.88 (10.28)	

^aStandard score on the Columbia Mental Maturity Scale (Burgmeister, Blum, & Lorge, 1972). Mean=5100; SD = 15.^bPeabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1987). Mean= 100; SD = 15.^cSyntactic quotient of Test of Language Development-Primary: Second Edition (Newcomer & Hammill, 1988). Mean = 100; SD = 15.^dGoldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986). Z scores were used in the analyses. Percentiles that correspond to each group's average z score are reported for descriptive purposes.^eComprehension subtest VI of the Stanford-Binet (Thorndike, Hagen, & Sattler, 1986). Mean = 100; SD = 16.

Table II

Nonword repetition stimuli.

One Syllable	Two Syllable	Three Syllable	Four Syllable
/naɪb/	/teɪvak/	/tʃɪnəɪtaʊb/	/veɪtətʃaɪdəɪp/
/voup/	/tʃouʊvæg/	/naɪtʃouʊveɪb/	/dæʊouʊnəɪtʃɪg/
/taʊdʒ/	/vætʃaɪp/	/dəɪtaʊvæb/	/naɪtʃəɪtaʊvub/
/dɔɪf/	/nəɪtaʊf/	/teɪvəɪtʃaɪg/	/tævətʃɪnaɪg/

Table III

Percentage of phonemes correct at each syllable length.

	One	Two	Three	Four	Total
SL1	70.18 (17.47)	71.13 (14.99)	68.13 (11.77)	59.68 (17.53)	65.84 (11.17)
6N	87.94 (10.57)	88.52 (9.41)	81.38 (12.12)	76.51 (12.01)	81.86 (8.62)
4N	84.90 (10.27)	84.95 (9.22)	78.35 (11.64)	66.50 (16.24)	76.10 (9.80)
Groups Combined	83.38 (13.64)	83.83 (12.32)	77.70 (12.71)	69.53 (16.07)	

Table IV

Item analysis.^a

	Grouped by Language Classification			Grouped by Race	
	SLI	6N	4N	AA	W
/b/ in /narb/	94%	69%	68%	73%	74%
/f/ in /narfaof/	94%	50%	55%	67%	56%
/p/ in /vætfarp/	88%	47%	52%	67%	50%
/v/ in /voop/	50%	22%	26%	33%	26%
/ɔt/ in /narfaotf/	88%	25%	29%	39%	20%
/v/ in /vætfarp/	54%	14%	29%	33%	24%
/k/ in /tervak/	50%	22%	16%	27%	24%
/tʃ/ in /vætfarp/	50%	11%	29%	42%	14%
/g/ in /tʃoovæg/	38%	19%	23%	27%	22%
/ɑ/ in /tervak/	31%	19%	0%	12%	16%
/v/ in /tʃoovæg/	25%	0%	10%	9%	8%

^aPatterns listed from highest to lowest rate of error. Also reported are the percentages of children in each group who made an error on the identified phoneme.