

Development and Validation of a Brief Assessment of Preschoolers' Articulation

Jason L. Anthony^{1*}, Martha Dunkelberger² and Rachel G. Aghara¹

¹Children's Learning Institute, University of Texas Health Science Center at Houston, USA

²Department of Communication Sciences and Disorders, University of Houston, USA

*Corresponding author: Jason L. Anthony, Children's Learning Institute, University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 2377, Houston, TX 77030-1501, USA, Tel: 15124713434; E-mail: Jason.L.Anthony@uth.tmc.edu

Received: August 22, 2014; Accepted: October 31, 2014; Published: November 07, 2014

Copyright: © 2014 Anthony JL, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Objective: The Houston Sentence Repetition Test of Articulation (HSRTA) was developed as a screener and brief outcome measure of articulation abilities of 3- to 5-year-old children. The HSRTA employs a sentence repetition task, which theoretically combines all of the advantages of the traditional citation method of assessing articulation with many of the advantages of the continuous speech method. The aim of this study was to examine the psychometric properties of the new measure.

Methods: A sample of 175 children was assessed twice, with approximately five months between assessment waves. The sample was ethnically diverse and ranged in age from 2 years and 11 months to 5 years and 4 months (mean age=4 years 6 months, SD=5 months). At each wave, children were administered the HSRTA and standardized tests of speech, language, and memory.

Results: The HSRTA demonstrated good internal consistency at both assessment waves (alphas=.84 and .86, respectively). Similarly, factor analysis clearly indicated it indexed a single latent ability. The HSRTA demonstrated moderate stability across the five month time span ($r=.57$, $p<.0001$). The new measure demonstrated convergent validity with a standardized articulation test ($r_s=.71$ and $.68$, $ps<.0001$) and discriminant validity with standardized vocabulary and auditory memory tests (r_s from $-.32$ to $-.47$). The HSRTA demonstrated internal consistencies and test-retest reliabilities that were equivalent to those of a standardized, norm referenced test of articulation, but the HSRTA was more sensitive to the effects of time ($F[1,160]=11.26$, $p<.01$).

Conclusion: Psychometric analyses indicated that the new measure is a reliable, valid, and sensitive tool for assessing individual differences in articulation skills among 3- to 5-year-old children. Collectively, results indicate the HSRTA surpasses minimum standards for a screener and brief outcome measure. Potential uses for researchers and practitioners are discussed.

Keywords: Assessment; Articulation; Sentence repetition; Preschool; Speech sound disorder; Early childhood

Introduction

Children's articulation skills have traditionally been assessed using one of two methods. The citation method requires children to produce a single word utterance, usually elicited through picture naming, object naming, or verbal repetition. In contrast, the continuous speech method requires children to produce words in connected speech. Continuous speech samples are usually elicited through conversation, storytelling, or story retelling. Each of these two assessment methods have important advantages and disadvantages that should be considered when one is planning a research study or an articulation assessment. In this paper, we discuss the tradeoffs between these two methods of assessing children's articulation, and we propose that a less common method (i.e., sentence repetition) may minimize tradeoffs and prove particularly useful in a variety of contexts. Finally, we describe and evaluate a new test based on the sentence repetition method.

When selecting which tests to use to assess children's articulatory functioning, practitioners and researchers must consider and weigh a

number of factors. Some of the issues to take into account include the purpose of assessment, qualifications of the examiner, breadth and depth of phoneme sampling, ease of phonetic transcription, psychometric properties of tests, oral language competencies of the child, and amount of time available for testing. How one prioritizes the answers to these questions will often dictate which method of assessment is chosen, because the citation method and the continuous speech method have distinct advantages and disadvantages.

Advantages of the citation method of assessing articulation

When breadth of phoneme sampling is of primary importance, the citation method offers a clear advantage over the continuous speech method. Phonemes are the individual sounds in a given oral language. The citation method allows for systematic assessment of all phonemes in a given language in all appropriate positions of words. This is achieved by test stimuli being carefully selected by the tests' authors such that the stimuli elicit each individual speech sound and each consonant blend in all of the positions that these sounds appear in the language. This commonplace design feature of articulation tests that employ the citation method ensures that any particular phonemes that a child struggles to produce will be stimulated for production and evaluation. A thorough and controlled sampling of phonetic targets is

necessary to reach reliable conclusions concerning the typicality of a child's speech production.

In contrast, breadth of phoneme sampling cannot be guaranteed when using a continuous speech method. When children provide a spontaneous speech sample, they tend to self-select words and sentence structures that are well within their articulatory and linguistic capabilities [1-3]. In other words, children may avoid saying words that include sounds they have difficulty saying and children may avoid saying complex sentences that include more advanced grammatical and syntactic rules. This avoidance of phonemes can render a continuous speech sample limited in scope and variety. For example, when Andrews and Fey [4] employed structured conversations aimed at eliciting particular target words in children's continuous speech, they found that none of their participants produced all of the target words in their continuous speech samples. Similarly, Hura and Echols [3] found that when their participants were administered a nonsense word repetition task the children tended to omit later developing phonemes, such as /r/, /v/ and /z/. Compounding children's tendency to favor phonetic constructions with which they are proficient, 3- to 8-year-old children with Speech Sound Disorder (SSD) also tend to produce shorter, less grammatically complex utterances than children with typical speech development [5], which further limits the number of sounds one can evaluate. Flipsen argued this may be due to increased co-morbidity rates of speech and language delays.

Another important advantage of the citation method of articulation assessment over the continuous speech method is ease of transcription of phonemes. Phonetic transcription is known to be an imperfect and unreliable endeavor [6]. For example, Gruber [7] reported inter-rater reliability of transcription of conversational speech samples by experienced, practicing speech-language pathologists at 89% for consonants and 82% for vowels. Those levels of agreement far exceed the levels we have observed in our own research. Students, who recently completed graduate level courses in phonetics and speech sound disorders, correctly transcribed 80% of consonants and 65% of vowels when listening to a continuous speech sample [8]. Because of the inherent challenges of phonetic transcription of spontaneous connected speech, researchers and practitioners may prefer assessment methods that provide a known stimulus to transcribe, often referred to as a "closed stimulus". A closed set of picture stimuli or object stimuli arguably allows the assessor to concentrate more on the child's production because less cognitive resources are needed for transcription, given that most of the child's production will correspond to expectancy. As such, when reliability of phonetic transcription is paramount and/or when efficient use of assessment time and scoring time is important, one may prefer to administer an articulation test that employs the citation method over one that employs the continuous speech method.

All too often, a need for efficient assessment is paramount. For example, researchers designing an assessment battery for a given study must consider the number of constructs in need of measurement, the limited attention spans of young children, budgetary constraints, and time constraints imposed by administrators in the case of school-based research. A speech-language pathologist working in a private practice must weigh third party reimbursement rates with the amount of time required of an evaluation. Parallel considerations exist for heads of special education in school districts. When brevity is paramount, one would generally prefer the citation method over the continuous speech method, as the time required to administer, transcribe, score, and interpret an articulation test based on the citation method may be as

much as three times less than that needed to administer, transcribe, score, and interpret a continuous speech sample [9].

Finally, when the purpose of assessment is classification or to determine eligibility for special services then the citation method offers some distinct advantages. Diagnosis and classification are contexts in which there is a strong need to contextualize a child's articulation accuracy relative to the population at large. For example, fair access to special education services in the public schools or to rehabilitation services in the private sector can only be assured through the use of a uniform classification system that is applied universally. These uniform classification systems are, by necessity, predicated on normative data. Detailed, normative data concerning the typical sequence of phoneme acquisition and other phonological processes (e.g., final consonant deletion, fronting, cluster reduction, etc.) based on assessments that employ the citation method are readily available and included in virtually all of the commercially prepared articulation assessments. Normative data from assessments employing the continuous speech method are less readily available and arguably less reliable, given children's tendency to self-select word structures and sentence structures with which they are more proficient [1-3]. The index yielded from a continuous speech sample that is commonly used to compare children's articulation abilities is Percent Consonants Correct (PCC). Although easily calculated and widely used, the PCC metric has limitations for inter-individual comparisons because it is susceptible to item characteristics that vary from speech sample to speech sample. For example, different children's speech samples may differ greatly in the relative proportions of early-, middle-, and late-developing sounds that they attempt. Thus, normative data on PCC in the context of spontaneous speech samples may need to be referenced cautiously when classification or determination of eligibility for special services is the goal of assessment.

In summary, the citation method of assessing children's articulation holds a number of advantages over the continuous speech method for practitioners and researchers alike. Specifically, it allows for a broad and systematic assessment of phonemes in each word position, improved accuracy of phonetic transcription, more efficient administration, and more reliable use of norm-referenced scoring. However, the continuous speech method holds some distinct advantages over the citation method that are also important to note.

Advantages of the continuous speech method of assessing articulation

One clear advantage of the continuous speech method is depth of phoneme sampling. Tests that employ the continuous speech method evaluate individual phonemes more frequently and in a greater variety of contexts. For example in a single speech sample, the sound /s/ may be produced and scored numerous times as an isolated consonant in a variety of positions in words, as part of numerous initial consonant blends, and as part of numerous final consonant blends. As such, the increased number of "items" in continuous speech samples is likely to increase this method's reliability for scoring individual sounds, all other things being equal. The increased number of contexts in which a sound is evaluated in a continuous speech sample may also increase this method's sensitivity, validity, and clinical relevance. Similarly, this method is more likely than the citation method to provide valuable information about children's phonological processes, such as fronting, stopping, final consonant deletion, and cluster reduction.

In contrast, tests that employ the citation method rarely sample individual phonemes in as many contexts. Instead, most articulation tests employing the citation method evaluate each phoneme in only one to four contexts (e.g., at the beginning of a word, in the middle of a word, at the end of a word, and maybe as part of a consonant blend) [10-12]. Because the citation method provides a relatively shallow assessment of articulation, it may be more likely to yield false negative classifications than the continuous speech method. False negatives are cases classified as typical even though they are truly disordered. The hypothesis that tests employing the citation method may yield more false negatives is consistent with findings that children tend to produce emerging sounds more accurately when assessed using citation methods than continuous speech methods [13]. Similarly, children evince fewer phonological processes (e.g., fronting, stopping, final consonant deletion, etc.) when evaluated using citation methods than they do when evaluated using continuous speech methods [4]. Thus, the continuous speech method is arguably more sensitive to children's articulation errors and to phonological processing errors.

Another advantage that might lead an assessor to select the continuous speech method is its ecological validity. Many consider the ecological validity of the continuous speech method to be better than that of the citation method because people generally communicate in spoken sentences rather than single word utterances. Shriberg and colleagues [14] assert that continuous speech samples are adequately robust for psychometric analysis and research purposes. Additionally, Kent, Miolo and Bloedel [15] demonstrated that a variety of analyses can be carried out on continuous speech samples including comparison to established norms (e.g., the PCC metric) [14], phonetic contrast analysis, phonological process analysis, and word level or continuous speech level analysis of phonetic accuracy. The variety of metrics available from a continuous speech sample offers researchers and practitioners the opportunity to evaluate more dimensions of children's articulation skills, which allows them to ask more questions, test more hypotheses, and more closely monitor development of various articulation skills.

Finally, a researcher's or clinician's desire to obtain indices of communicative competence beyond phonological accuracy may lead her or him to decide to use a continuous speech task. For example, continuous speech samples yield invaluable information concerning prosody, overall intelligibility, morphosyntactic complexity, and semantic complexity [16]. In contrast, a citation task by its very design cannot assess any of these oral language competencies. Thus, the argument can be made that a continuous speech sample offers a clinically efficient method of gathering data that addresses a variety of oral language competencies beyond articulation.

In summary, research pertaining to articulation assessments suggests that both the citation method and the continuous speech method offer distinct advantages and disadvantages, such that one method's strengths tend to be the other method's weaknesses and vice versa. Many studies have shown that data gathered by these two methods of elicitation are somewhat disparate [4,13,15-17]. Used in combination, therefore, the two methods could well inform a practitioner or researcher who recognizes their differences and uses the data appropriately to answer clinical and theoretical questions [9,18,19]. Consequently, the seemingly ideal situation would be to use tests based on both methods to achieve a comprehensive evaluation of a child's articulation. However, this would be a very lengthy assessment that may not be feasible in school, clinical, or research settings.

Sentence repetition as a viable alternative method of assessing articulation

A possible middle ground with few compromises is the elicitation of a continuous speech sample using a sentence repetition method. The sentence repetition method requires children to repeat sentences that are spoken by an examiner. Examiners then score the accuracy of each consonant and consonant blend produced. To avoid floor and ceiling effects, the sentences should be designed such that their difficulty matches the developmental levels of the target population by imposing certain phonetic features, word features, and sentence features. For example, phonetic features such as proportion of early-, middle- and late developing sounds, proportion of consonant blends, and frequency of occurrence of phonemes in speech can be manipulated to influence the difficulty of the test and corresponding appropriateness for children of different ability levels. Likewise, word features and sentence features (e.g., word frequency, age of acquisition of vocabulary words, average number of syllables per word, Mean Length of Utterance (MLU)) can all be manipulated with the same purpose in mind. Theoretically, a well-designed sentence repetition test of articulation should allow a researcher or practitioner to achieve goals of efficiency, scoring accuracy, ease of phonetic transcription, breadth and depth of phonetic sampling, while still potentially eliciting phonological processes and dysfluencies. The standardized administration procedure and "close set" of responses would appear to have the potential to yield a highly reliable measure that could yield reliable normative data. In short, the sentence repetition method combines all the benefits of the citation method with nearly all the benefits of the continuous speech method. The only apparent short comings of the sentence repetition method are that it yields less useful information about prosody, morphosyntax, or semantics than can be derived from a continuous speech sample.

Although only a few studies have evaluated sentence repetition methods of articulation assessment, the results thus far demonstrate good convergent validity and reliability. Johnson, Weston, and Bain [20] compared 4- to 6-year-olds' structured conversational speech samples to their performances on a sentence repetition test. The sentences were designed such that they consisted of three to seven words of developmentally appropriate vocabulary. The phonemes represented in the sentences were weighted based on Shriberg and Kent's [6] report on the proportional occurrence of phonetic classes in first grade children. The results showed that the PCC measures of both tasks yielded clinically and statistically equivalent results.

Gordon-Brannon and Hodson [16] used ten simple, active, declarative sentences of five words each in a repetition task that was paired with visual stimuli to determine if speech intelligibility classifications could be more easily applied in a known context. The sentences included late developing phonemes, such as fricatives, affricates, and liquids. Gordon-Brannon and Hodson showed that the known context and morpho-syntactic cues assisted their examiners in accurately producing a phonetic or orthographic transcription of their participants' utterances.

Although a few sentence repetition tests of articulation were available when we conducted the present study [16,20], none were well suited for our purposes. Some existing sentence repetition tests were too lengthy, such as the Sounds-in-Sentences subtest of The Goldman-Fristoe Test of Articulation-2 (GFTA-2). Others included item content well beyond the abilities of our target population either because they sampled too many late-developing phonemes or the sentences were too morphosyntactically complex. Consequently, we developed a brief,

developmentally appropriate sentence repetition test, called the Houston Sentence Repetition Test of Articulation (HSRTA).

Purpose of the present study

This study aimed to evaluate the psychometric properties of HSRTA. We hypothesized that although a brief measure, HSRTA would demonstrate good internal consistency and that children's performances on the measure would be well explained by a single latent ability. We also expected HSRTA to demonstrate good convergent validity via large and significant correlations with a widely accepted, standardized test of articulation. Because children's performances on sentence repetition tasks can be influenced by skills other than articulation ability, it was also necessary to evaluate discriminant validity with proximal skills. As such, we hypothesized that HSRTA would be more highly correlated with standardized articulation tests than with standardized tests of vocabulary and short-term memory. Finally, we expected HSRTA to be sensitive to developmental changes that occur over a five month time span.

Method

Participants

Educational context

The current study was embedded in a larger project that evaluated the efficacy of a low intensity book exchange program and a low intensity parent education program, neither of which had any impact on children's speech development. Classroom-level inclusion criteria were (a) full day preschool programming, (b) all classroom instruction provided in English, (c) most or all children in classrooms were native English speakers, (d) most children in classrooms were 4 years of age, and (e) enrollment in the Texas Early Education Model (TEEM). TEEM emphasizes frequent, intensive, and ongoing professional development for early childhood educators, on-site mentoring, regular monitoring of children's academic progress, and choice among a list of research-based curricula. TEEM also requires integration among early childhood education service delivery systems. As such, about an equal number of federally funded Head Start classrooms, state-funded public school prekindergarten classrooms, and privately funded child care classrooms participated. The 23 Houston-based classrooms that participated in the current study were among the approximately 2500 classrooms across Texas that participated in TEEM during the 2006/2007 school year.

Child demographics

Active parental consent was obtained to assess children's speech, language, and literacy skills. From the pool of 271 consented children who attended the 23 participating classrooms, 8 children were randomly selected from each classroom to be included in the program evaluation and the present study. This translated into a sample size of 175 children. When first tested in late fall, the sample of 175 children ranged in age from 2 years and 11 months to 5 years and 4 months, with a mean age of 4 years and 6 months and a standard deviation of 5 months. Most children were 4-year-olds ($n=135$). Eighty-nine children were female (51%) and eighty-six children were male (49%). The ethnic breakdown of the sample was 58.5% African-American, 22.6% Hispanic American, 13.4% Caucasian, and 5.5% multiracial. Children's performances on standardized tests of expressive vocabulary, receptive

vocabulary, short-term auditory memory, and nonverbal ability revealed average to low average abilities at both assessment waves.

Study design

This correlational study involved administration of an assessment battery to participating children at two points in time. Specifically, testing occurred in late fall and mid spring of the 2006/2007 school year. In addition to the HSRTA, another measure of articulation was administered at both assessment waves to establish convergent validity and to examine relative reliability of the two articulation tests. Because children's vocabulary and auditory memory skills were the most likely candidates that may confound HSRTA, standardized measures of vocabulary and auditory short-term memory were administered at both assessment waves to establish discriminant validity (i.e., to demonstrate that HSRTA is more a test of articulation than a test of vocabulary or memory). It is noteworthy that we selected a test of auditory memory that employed a sentence repetition task because it would be a very stringent test of the discriminant validity of HSRTA.

Measures

Articulation

Houston Sentence Repetition Test of Articulation (HSRTA)

This measure required children to repeat sentences spoken by an examiner. The examiner spoke the sentences with typical inflections (e.g., rising for interrogative forms, falling for declarative forms). If a child omitted any whole words from his or her reproduction of the stimulus sentence, then that sentence was readministered by the examiner. Consonant omissions, substitutions, and additions were scored as errors. Consonant distortions were not counted as errors, according to the assumption that consonant distortions still reflect participants' awareness of the contrastive function of the target phoneme [21]. The total number of consonant errors made for each sentence was recorded.

Stimuli for HSRTA consisted of 15 sentences (Table 1) that contained a total of 80 consonant phonemes. The 24 consonant sounds in English can be divided into three distinct developmental classes, based on the consistency with which they are produced by 3- to 6-year-old children with typical development and by 3- to 6-year-old children with delayed speech development [22,23]. According to Shriberg and colleagues, preschool children accurately produce the "early 8" consonant sounds in running speech more than 75% of the time. The "middle 8" consonant sounds are produced accurately in running speech between 25% and 75% of the time. The "late 8" consonant sounds are produced accurately less than 25% of the time. This developmental classification of consonant sounds informed the design of HSRTA. Specifically, we sampled 30 early consonant phonemes, 30 middle consonant phonemes, and 20 late consonant phonemes. Loading the sentences with early and middle occurring consonant sounds allowed more in-depth sampling of the phonemes that were likely to provide the most information about individual differences in articulation among our preschool aged participants.

Frequency of phoneme occurrences and word structures were also considered in the design of stimuli for HSRTA [6]. The most commonly occurring phonemes in adult's and children's speech (/n/, /t/, /s/ and /d/) comprised 31% of the sampled consonant sounds. The least commonly occurring consonant sound (/ʒ/) was not sampled at all. Additionally, a majority (75%) of the words used were

monosyllabic and the remainder were predominantly disyllabic, as suggested by Flipsen and colleagues [1].

Hello there.
Happy Birthday!
Pick it up.
Meet the family.
Go away, Shoo!
That was mean.
Just you wait!
No more fighting!
Very nice cats!
I'm helping.
It's for me!
Can we go swimming?
Aren't you a sweet baby?
I'm running to school.
I'm going to the zoo!

Table 1: Items on the Houston Sentence Repetition Test of Articulation.

Finally, sentences of HSRTA were purposefully constructed to conform to the lower end of the typical range of MLU for preschool children, in order to minimize potentially confounding effects of auditory short-term memory. Specifically, sentences ranged from two to seven morphemes in length with a MLU of 3.93. Eleven of the fifteen sentences were simple sentences in greeting (2), imperative (4) or declarative (5) forms. Two sentences were interrogatives (i.e., questions) and two sentences were complex sentences with embedded prepositional phrases. Using simple syntactic and morphological forms should have minimized the demands on participants' memory, as all forms of sentences used in HSRTA are typically mastered by Brown's Late Stage V [24].

Goldman-Fristoe Test of Articulation-2 (GFTA-2)

The Sounds-in-Words subtest employed a picture naming task to elicit single-word utterances. To administer the subtest, examiners presented a picture and asked "What is this?" and children typically produced the target word. Specific methods of cuing are used if children do not spontaneously produce the target word. Target words were short, high-frequency words that contained at least one targeted consonant or a targeted initial consonant cluster. Specifically, the articulation of 23 individual consonant sounds was evaluated from 1 to 3 times across initial, medial, and final positions. All items were administered to all children. Standardized administration and scoring procedures were followed.

Vocabulary

Expressive One-Word Picture Vocabulary Test (EOWPVT)

The EOWPVT employed a picture naming task to elicit nouns, verbs, and adjectives. Specifically, examiners presented children with colored line drawings that depicted an action, object, category, or concept. Children were asked to label each drawing. Prescribed cuing methods were used for elicitation if children responded to the wrong part of an illustration or if they provided a response at the wrong level of abstraction. Standardized administration and scoring procedures were followed [25].

Receptive One-Word Picture Vocabulary Test (ROWPVT)

The ROWPVT used a multiple choice, picture pointing task to assess receptive vocabulary. To administer, examiners stated a word and children were required to point to one of four illustrations that corresponded to the stimulus word. Standardized administration and scoring procedures were followed [26].

Auditory short-term memory

Woodcock-Johnson Psychoeducational Battery, Revised (WJ-R)

The Memory for Sentences subtest required an examinee to repeat a sentence that was spoken by an examiner. According to standardized procedures, verbatim responses were awarded two points, responses involving only one error were awarded one point, and responses involving more than one error were awarded zero points. Articulation errors were not penalized and testing was discontinued after four consecutive zero point responses, all in accord with standard administration procedures. Stimuli ranged from 1 to 24 morphemes in length with a MLU of 6.8 [27].

Procedures

The speech measures (i.e., articulation measures) and cognitive measures (i.e., measures of auditory memory, nonverbal ability, and vocabulary) were administered by separate assessment teams. The assessment team who administered the articulation measures was composed entirely of speech-language pathology students, including one undergraduate student, two post baccalaureate students, and eight graduate students. All speech-language pathology students had already completed course work and labs in phonetics and speech sound disorders. These examiners attended a 3-day training workshop led by the second and third authors.

The assessment team who administered the cognitive measures was composed of highly experienced research staff. All of these examiners had undergraduate or graduate degrees. Many were retired teachers. One was a retired pediatrician, and one was a retired developmental psychologist. Examiners who administered the cognitive battery attended a 4-day training workshop led by the first author.

The two assessment teams responsible for testing during Wave 1, in late fall, were also responsible for testing during Wave 2, in mid spring. To ensure procedural reliability, all examiners attended refresher training sessions in preparation for Wave 2, and all examiners were required to demonstrate competence on all measures in their battery prior to both assessment waves. This was accomplished through individual test-outs with the first author.

Children were tested individually at their preschools. Testing took place in locations that school administrators designated for testing. Testing was typically conducted in 20 to 60 minute sessions. The length of the testing sessions was determined on a per child basis depending on his or her attention span and desire to continue. Children were given general verbal praise (e.g., "Good job", "Nice

working”, “I’m having fun with you”), physical praise (e.g., high fives), and tangible reinforcements (e.g., stickers) for participating in the testing.

Results

Preanalysis data inspection

Descriptive statistics indicate that the sample performed similarly on the two articulation measures at both assessment waves (Table 2). Specifically, sample means were 11.1 errors and 11.4 errors at Wave 1

on the HSRTA and GRTA-2 respectively and 7.3 errors and 8.6 errors at Wave 2 respectively. The two articulation measures also demonstrated moderate positive skewness at both assessment waves, indicating that in general the sample found both articulation tests relatively easy but that ceiling performances were infrequent. Moreover, the two articulation measures evidenced identical skewness at both waves. All of the above findings coincide with expectancies based on the sample’s age, in light of the average standard scores obtained at both time points on the GFTA-2. The two articulation tests were highly inter-correlated and moderately correlated with measures of vocabulary and memory (Table 2).

	M	SD	Skew	GFTA-2	EOWPVT	ROWPVT	WJ-R
Wave 1							
HSRTA	11.15	7.19	1.28	0.71**	-0.34**	-0.35*	-0.47**
GFTA-2	11.35	8.09	1.26		-0.25*	-0.25*	-0.31**
EOWPVT	37.98	12.75	0.87			0.76**	0.58**
ROWPVT	43.79	13.06	0.08				0.58**
WJ-R	30.36	4.45	-0.08				
Wave 2							
HSRTA	7.30	4.98	1.48	0.68**	-0.32**	0.41**	-0.46**
GFTA-2	8.59	6.80	1.47		-0.28*	-0.38**	-0.34**
EOWPVT	44.52	13.11	0.67			0.78**	0.64**
ROWPVT	51.64	13.08	-0.07				0.64**
WJ-R	31.43	5.05	-0.35				

Table 2: Descriptive Statistics and Correlations at Wave 1 and Wave 2. *p<.001; **p<.0001; HSRTA=Houston Sentence Repetition Test of Articulation; GFTA-2=Sounds-in-Words subtest of Goldman Fristoe Test of Articulation 2nd edition; EOWPVT=Expressive One-Word Picture Vocabulary Test; ROWPVT=Receptive One-Word Picture Vocabulary Test; WJ-R=Memory for Sentences subtest of Woodcock Johnson-Revised.

Internal consistency of the HSRTA

The HSRTA demonstrated good internal consistency at both assessment waves (Cronbach’s alphas=.84 and .86 for Wave 1 and Wave 2, respectively). These internal consistency estimates were impressive given that they were based on only fifteen items. That is, because of the positive association between number of items and Cronbach alpha coefficients, the internal consistency estimates of .84 and .86 are probably underestimates of the true internal consistency of the HSRTA, which actually sampled 80 consonant phonemes. For the sake of comparison, the 72 item GFTA-2 also demonstrated good internal consistency (Cronbach’s alphas=. 89 and .87 for Waves 1 and 2, respectively).

Children’s scores on the fifteen items of the HSRTA were subjected to exploratory factor analysis to examine how many latent abilities were driving children’s performances on the new test. Principle factor analysis of the fifteen items from the first assessment wave (N=175) found the first factor explained 91% of the variance in children’s scores. This factor had a large eigen value of 4.3, and it included sizable loadings (Lambdas=.35 to .65) from all but one item. The second largest factor had an eigen value of only 0.53, and it included only one appreciable loading of .41; however, that item loaded more strongly on

the first factor. In short, a one factor solution was clearly superior based on all rules of thumb (e.g., eigen values greater than 1.0, 70% variance explained, and interpretability of factors). These results are consistent with the notion that HSRTA was measuring a single latent ability, presumably articulation ability.

Test-retest reliability and sensitivity to time of the HSRTA

To examine test-retest reliability and sensitivity to change across time, the HSRTA was administered at two time points, with approximately five months of time elapsing between assessment waves. Some degree of change was expected in participant scores due to the developmental progression of skills over the five month time period. Therefore, we expected to find only moderate test-retest correlations. Indeed, the five month test-retest correlation of the HSRTA was significant at .57, p<.0001. For the sake of comparison, the five month test-retest correlation of the GFTA-2 was comparable at .66, p<.0001. The difference between the test-retest correlation of the HSRTA and the test-retest correlation of the GFTA-2 was not significant, t=1.25, p>.20.

Both HSRTA and GFTA-2 demonstrated sensitivity to the five month interval of time, as indicated by tests of the difference between

raw scores obtained at Wave 1 and Wave 2 ($t_s=8.7$ and 6.0 , $p_s<.0001$, for HSRTA and GFTA-2, respectively). In an effort to compare the two articulation tests' sensitivity to time, the raw scores obtained on each measure at Wave 1 were subjected to a z-score transformation, resulting in sample means on both measures of zero and sample standard deviations on both measures of one. Participant's raw scores obtained at Wave 2 were then rescaled to the metric of Wave 1. This was accomplished by subtracting the observed mean of the sample obtained at Wave 1 from the observed raw score that a given child obtained at Wave 2, and then dividing the result by the observed standard deviation of Wave 1. In short, the sample as a whole demonstrated a .56 standard deviation improvement in scores on the HSRTA and a .37 standard deviation improvement in scores on the GFTA-2.

To statistically compare the two articulation tests' sensitivity to time, we performed a repeated measures mixed effects analysis. The mixed model nested z-scores on both articulation tests at both waves within children, which yielded four observations per child, in order to control for the fact that observations of the same child are not independent. An unstructured residual covariance matrix was included in the model to account for the unequal variances of the four measurements as well as the correlations between them, which ranged from .52 to .76 ($p_s<.0001$). Independent variables in the model included a categorical variable called Measure, which had values of HSRTA or GFTA-2, a categorical variable called Wave, which had values of zero at Wave 1 and values of one at Wave 2, and a Measure by Wave interaction. The significant main effect of Wave ($F[1,155]=73.17$, $p<.0001$) was expected given results of the t-tests described above. More importantly, there was a significant Measure by Wave interaction ($F[1,160]=11.26$, $p<.01$) that indicated the effect of Wave on children's performances was reliably greater when children's performances were quantified by the HSRTA.

Convergent and discriminant validity of the HSRTA

The HSRTA demonstrated good convergent validity with the Sounds-in-Words subtest of the GFTA-2. Specifically, Table 2 shows that correlations of the HSRTA with the GFTA-2 were high, positive, and significant at both assessment waves ($r_s=.71$ and $.68$, $p_s<.0001$, for Wave 1 and Wave 2, respectively). Table 2 also shows that HSRTA demonstrated good discriminant validity, as correlations of the HSRTA with the EOWPVT, ROWPVT and WJ-R were uniformly moderate, negative, and significant at Wave 1 ($r_s=-.34$, $-.35$, and $-.47$, respectively, $p_s<.0001$) and at Wave 2 ($r_s=-.32$, $-.42$, and $-.46$, respectively, $p_s<.0001$). It is important to note that the negative relations of the HSRTA with the measures of discriminant validity were expected as they were a consequence of different scoring practices: The HSRTA and the GFTA-2 were scored in terms of the total number of errors, whereas the EOWPVT, ROWPVT, and WJ-R were scored in terms of the total number correct responses.

We then performed significance tests of the difference between dependent correlations to verify that the HSRTA was more highly correlated with the GFTA-2 than it was with measures of more distal constructs. At Wave 1, the HSRTA was significantly more highly correlated with the GFTA-2 than it was with the EOWPVT, ROWPVT, and WJ-R Memory for Sentences ($t_s=5.7$, 5.6 , and 4.1 , respectively, $p_s<.001$). At Wave 2, the HSRTA was significantly more highly correlated with the GFTA-2 than it was with the EOWPVT, ROWPVT, and WJ-R Memory for Sentences ($t_s=5.1$, 4.2 , and 3.4 , respectively, $p_s<.001$).

Discussion

The present study was designed to evaluate the psychometric properties of a new measure of preschool children's articulation abilities. The Houston Sentence Repetition Task of Articulation (HSRTA) was developed to serve the purposes of a screener and brief outcome measure for use with children aged 3- to 5-years. By employing a sentence repetition task, the HSRTA capitalizes on most advantages of the citation method and the continuous speech method of articulation assessment. Consequently, the HSRTA arguably allows for a broad and deep assessment of children's articulation in an ecologically valid context (i.e., connected speech). The closed stimulus set makes scoring and phonetic transcription, if desired, relatively easy and reliable. The standardized procedures, closed stimulus set, and ease of scoring make the HSRTA, and other articulation tests that employ the sentence repetition method, amenable to standardization and reliable usage of norm-referenced data. Regarding psychometrics, the new measure generally demonstrated impressive reliability, validity, and sensitivity to change especially in light of its brevity.

The HSRTA demonstrated good internal consistency at both assessment waves, and moderate test-retest reliability across a five month time span. Moreover, the HSRTA demonstrated internal consistency values and a test-retest reliability coefficient that were equivalent to those of the widely used Sounds-in-Words subtest of the GFTA-2, which is a standardized and nationally normed articulation test that is used by thousands of speech-language pathologists in research, school, rehabilitation, and private practice settings. Collectively, these results indicate that the HSRTA surpasses minimum standards for adequate reliability as a screener and brief outcome measure.

One of the most exciting results of the present study was that not only was HSRTA found sensitive to articulation development that occur over a five month interval of time but that HSRTA was actually found more sensitive than the Words-in-Sounds subtest of the GFTA-2. We conjecture that the increased sensitive of the HSRTA was probably due to our purposeful sampling of phonemes that matched the developmental levels of our sample. Specifically, of the 80 consonant sounds evaluated by the HSRTA, 30 were early developing phonemes, 30 were middle developing phonemes, and 20 were late developing phonemes. Relatively higher proportions of early and middle developing consonant sounds permitted in-depth sampling of the phonemes likely to provide the most information about individual differences in articulation among our preschool aged participants. In contrast, if one were to develop a sentence repetition test of articulation for older children, then one would want to include proportionately more late developing phonemes and proportionately fewer early developing phonemes.

Keeping in mind that the Words-in-Sounds subtest was also found sensitive to children's articulation development over the five month time period, the significant interaction between time and articulation test indicated that HSRTA is likely to be sensitive to improvements in articulation that occur over an even shorter time span. This is promising news for researchers and practitioners who monitor preschool children's speech development over brief time periods or who evaluate the efficacy of relatively brief interventions. Based on the pattern of findings and the magnitude of the difference in sensitivities of the HSRTA and the Words-in-Sounds subtest, one would expect the HSRTA to be sensitive to typical articulation development that occurs over a three or four month time span. However, future research will

need to confirm or refute that expectation and identify the minimal amount of typical development to which the HSRTA is sensitive.

Convergent validity of the HSRTA was evidenced through significant correlations with the Sounds-in-Words subtests of the GFTA-2 at both assessment waves. It is noteworthy that significant validity coefficients were obtained despite the fact that these two articulation measures employ different methods of assessment: The HSRTA employs the sentence repetition method, and the Sounds-in-Words subtest employs the citation method. Thus, their association and corresponding validity coefficient were not inflated by shared method variance.

To evaluate the discriminant validity of HSRTA, we administered to our participants measures of proximal constructs that would be most likely to confound performances on the HSRTA. Specifically, we examined the extent to which children's performances on the HSRTA were related to their vocabulary and memory abilities. Only 12% of the variance in children's performances on the HSRTA could be explained by children's vocabulary scores. These results indicate that we successfully incorporated into the sentence stimuli vocabulary that was well within our participants' abilities.

A more noteworthy 22% of the variance in children's performances on the HSRTA could be explained by children's memory scores. This result was somewhat surprising given that the sentences included in the HSRTA were much less complex and shorter (MLU=3.9) than those included in the test of auditory short term memory (MLU=6.8). However, all of our preschool participants reached the stop criteria on the memory test long before administration of the longest and most complex sentences on that test. As such, the average length and complexity of sentences from the memory test that were actually administered would have been much smaller. It appears that the discriminant validity of the HSRTA could be improved by shortening and making less complex a couple of the longest sentences on the test. Whereas minimizing the potentially confounding role of memory is important, some degree of association will be unavoidable given that the HSRTA and the memory test both employ a sentence repetition task. Despite the shared method variance, correlations of children's performances on the memory test and the HSRTA were significantly smaller than correlations of HSRTA with the other articulation measure (i.e., Sounds-in-Words from the GFTA-2). These results in combination with those of the factor analysis that demonstrated HSRTA performances were well explained by a single latent ability indicate that HSRTA is primarily a test of articulation and not primarily a measure of memory or oral language.

In summary, psychometric analyses indicated that our new measure is a reliable, valid, and sensitive tool for assessing individual differences in articulation skills among 3- to 5-year-old children. These findings extend prior research that demonstrated good convergent validity between a different sentence repetition test of articulation and a continuous speech test of articulation in 4 to 6-year old children [20]. As such, the HSRTA may fill an important gap given that currently available articulation tests that employ the sentence repetition method are only appropriate for older children or take much longer to administer.

The present findings have direct implications for researchers and practitioners who have need to evaluate children's articulation abilities. For example, we have used the HSRTA in our research to help identify the roles of articulation, phonological awareness, and phonological representation in the development of young children's

emergent literacy [28,29]. Alternatively, Speech-Language Pathologists (SLP) employed in various settings could use the HSRTA as a screener to help them identify preschool children who may be in need of further evaluation aimed at diagnosis and treatment planning. However additional research that compares the diagnostic efficacy of various cut-scores is needed to provide users of HSRTA with an optimal screening criterion.

Limitations

Although there are no major threats to the internal validity of the study that limit the conclusions that we can draw concerning the internal consistency, validity, and sensitivity of the HSRTA, there were some noteworthy limitations that do preclude drawing other potentially important conclusions. First, the five month interval of time that elapsed between administrations of the HSRTA, although reasonable for examining sensitivity to time, was too long to be considered a reasonable evaluation of test-retest reliability. More appropriate time frames from which to evaluate test-retest reliability of a measure of articulation would range from approximately one week to one month, when little change in ability is expected in the absence of speech therapy and when practice effects would be negligible. Given the current study's five month time frame, it was no surprise that the test-retest correlation of the HSRTA was only moderate in size.

Second, failure to include in the study an articulation measure that employed the continuous speech method precluded examination of the efficiency and utility of the HSRTA relative to this other commonly used method of assessing young children's articulation. Future research will need to establish the degree of agreement between PCC achieved on the HSRTA and that achieved in a continuous speech sample. In the absence of such direct empirical support, we must extrapolate from prior research that has demonstrated equivalence of scores obtained from the continuous speech method and the sentence repetition method in 4- to 6-year-old children [20]. Nonetheless, the present study did provide sufficient evidence to support the convergent validity of the HSRTA with a standardized articulation test that employed the citation method.

Third, because psychometric analyses based on classical test theory are sample specific, it is important to note limitations of the generalizability of the findings. Given that approximately two thirds of the sample was recruited from facilities that have financial need-based eligibility criteria, the present sample included a relatively large proportion of children from lower socioeconomic strata. Additionally, the present sample was comprised of a large proportion of children from ethnic minority backgrounds. Collectively, such demographic characteristics are associated with increased risk for learning difficulties. Indeed, most of the present sample demonstrated low average or average language and memory abilities. However, the vast majority of the present sample scored in the average range on the nationally normed test of articulation. As such, results of the present study can be comfortably generalized to 3- to 5-year-old children from socioeconomically disadvantaged backgrounds and possibly to 3- to 5-year-old children from the general population as well. However, the present findings are not generalizable to children outside the age range of 3 to 5 years or to children with exceptionalities, such as hearing impairment.

Fourth, because the present study did not evaluate the HSRTA on a clinical sample in the context of diagnostic and therapeutic services provided by certified speech-language pathologists, the present study was unable to address the HSRTA's diagnostic efficacy or sensitivity to

effects of treatment. Although we believe the sentence repetition method of articulation assessment is well suited to evaluate treatment progress if the item content includes a large sampling of the phonemes targeted in a particular child's therapy, we did not construct the HSRTA to serve such a purpose and, as such, we do not advocate that the HSRTA be used for such a purpose. Likewise, HSRTA was neither developed as or should be used as a diagnostic tool. Instead, HSRTA was developed to efficiently and accurately index individual differences in global articulation ability among 3- to 5-year-old children and to be sensitive to changes in global articulation ability that occur over the course of a five month time span.

Conclusions

In summary, the HSRTA is a reliable, valid, and sensitive tool for researchers or practitioners who need to efficiently index differences among preschool children's global articulation ability or index change in that ability over at least a five month period of time. As such, results of the present study strongly support use of HSRTA as a general outcome measure, and they imply HSRTA has the potential to also serve as a screener. However, future research will need to specifically evaluate the HSRTA as a screener for identifying preschool age children in need of more comprehensive assessment. In addition to reporting the psychometric properties of this new measure of articulation, the present study exemplified the utility of a relatively underutilized method of assessing children's articulation, and it highlighted the particular phonological, semantic, and morphosyntactic features that one needs to consider when constructing a sentence repetition test of articulation. As such, we hope that this study piques interest among academics and professionals and sparks development of similarly well-constructed sentence repetition tests of articulation for other populations and purposes.

References

1. Flipsen P Jr, Hammer JB, Yost KM (2005) Measuring severity of involvement in speech delay: segmental and whole-word measures. *Am J Speech Lang Pathol* 14: 298-312.
2. Goffman L (1999) Prosodic influences on speech production in children with specific language impairment and speech deficits: Kinematic, acoustic and transcription evidence. *Journal of Speech Language and Hearing Research* 42: 1499-1517.
3. Hura SL, Echols CH (1996) The role of stress and articulatory difficulty in children's early productions. *Developmental Psychology* 32: 165-176.
4. Andrews N, Fey ME (1986) Analysis of the speech of phonologically impaired children in two sampling conditions. *Language, Speech and Hearing Services in the Schools* 17: 187.
5. Flipsen P (2006) Syllables per word in typical and delayed speech acquisition. *Clin Linguist Phon* 20: 293-301.
6. Shriberg LD, Kent RD (2003) *Clinical Phonetics*. (3 ed.), Boston, MA: Allyn and Bacon.
7. Gruber FA (1999) Probability estimates and paths to consonant normalization in children with speech delay. *J Speech Lang Hear Res* 42: 448-459.
8. Anthony JL (2007) [Accuracy of phonetic transcription by SLP students in Raising a Reader project]. Unpublished raw data.
9. Masterson JJ, Bernhardt BH, Hofheinz MK (2005) A comparison of single words and conversational speech in phonological evaluation. *Am J Speech Lang Pathol* 14: 229-241.
10. Fudala JB (2000) *Arizona articulation proficiency scale*. (3rd ed.), Los Angeles, CA: Western Psychological Services.
11. Hodson BW (2004) *Hodson assessment of phonological patterns*. (3rd ed.), Austin, TX: Pro-Ed.
12. Lippke BA, Dickey SE, Selmar JW, Soder AC (1997) *Photo articulation test*. (3rd ed.), Austin, TX: Pro-Ed.
13. Morrison JA, Shriberg LD (1992) Articulation testing versus conversational speech sampling. *J Speech Hear Res* 35: 259-273.
14. Shriberg LD, Austin D, Lewis BA, McSweeney JL, Wilson DL (1997) The percentage of consonants correct (PCC) metric: extensions and reliability data. *J Speech Lang Hear Res* 40: 708-722.
15. Kent RD, Miolo G, Bloedel S (1994) The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology* 3: 81.
16. Gordon-Brannon M, Hodson BW (2000) Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology* 9: 141.
17. Kwiatkowski J, Shriberg LD (1992) Intelligibility assessment in developmental phonological disorders: accuracy of caregiver gloss. *J Speech Hear Res* 35: 1095-1104.
18. Bernhardt BH, Holdgrafer G (2001) Beyond the basics II: Supplemental sampling for in-depth phonological analysis. *Language, Speech and Hearing Services in the Schools* 32: 28.
19. Wolk L, Meisler AW (1998) Phonological assessment: a systematic comparison of conversation and picture naming. *J Commun Disord* 31: 291-310.
20. Johnson CA, Weston AD, Bain BA (2004) An objective and time-efficient method for determining severity of childhood speech delay. *Am J Speech Lang Pathol* 13: 55-65.
21. Shriberg LD, Kwiatkowski J, Hoffmann K (1984) A procedure for phonetic transcription by consensus. *J Speech Hear Res* 27: 456-465.
22. Shriberg LD (1993) Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *J Speech Hear Res* 36: 105-140.
23. Shriberg LD, Kwiatkowski J (1994) Developmental phonological disorders. I: A clinical profile. *J Speech Hear Res* 37: 1100-1126.
24. Owens RE (2005) *Language development: An introduction*. (6 ed.), Boston, MA: Allyn and Bacon.
25. Brownell R (2000a) *Expressive One Word Picture Vocabulary Test*. (2000 ed.), Novato, CA: Academic Therapy Publications.
26. Brownell R (2000b) *Receptive One Word Picture Vocabulary Test*. (2000 ed.), Novato, CA: Academic Therapy Publications.
27. Woodcock R, Johnson M (1989) *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM Teaching Resources.
28. Anthony JL, Greenblatt Aghara R, Dunkelberger MJ, Anthony TI, Williams JM, et al. (2011) What factors place children with speech sound disorders at risk for dyslexia? *American Journal of Speech-language pathology* 20: 146-160.
29. Anthony JL, Williams JM, Aghara R, Dunkelberger M, Novak B, et al. (2010) Assessment of individual differences in phonological representation. *Reading and Writing: An Interdisciplinary Journal* 23: 969-994.