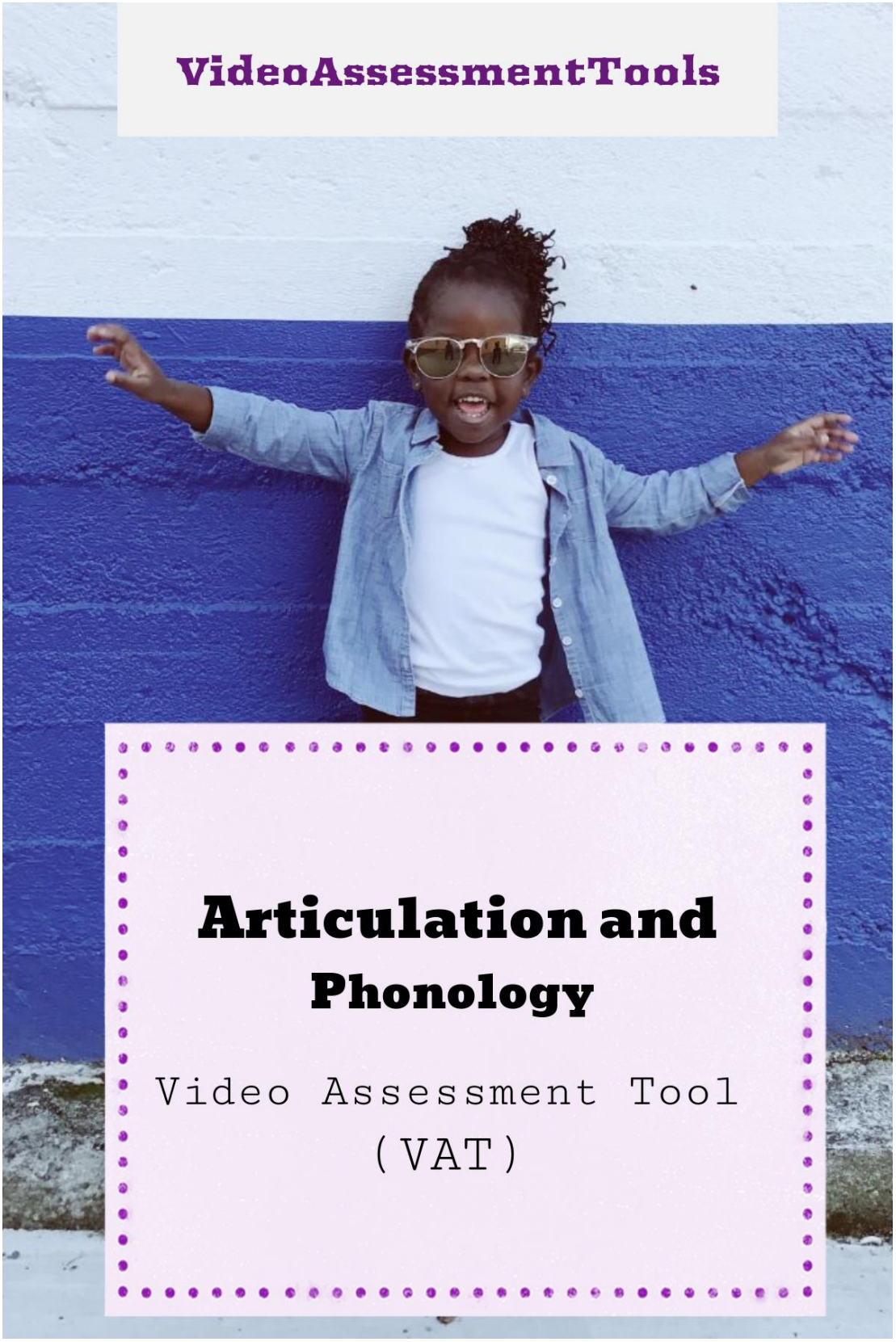
VideoAssessmentTools



Articulation and Phonology

Video Assessment Tool (VAT)

Copyright © 2024 by Lavi Institute

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

TABLE OF CONTENTS

Authors

Adriana Lavi, PhD, CCC-SLP

Cristina Gamarnik, MS, CCC-SLP

Technical Publications Writer

Charlotte Granitto, MS, CCC-SLP

Elaine Lopez

Acknowledgements

Sincere appreciation is extended to the families and children who generously offered their time and effort. Additional thanks goes to the expert panel and speech-language pathologists who field-tested the children.

Chapter

1

Overview of the Assessment

Articulation and Phonology Video Assessment Tool (VAT)

The *Articulation and Phonology Video Assessment Tool (VAT)* is a standardized and norm-referenced articulation and phonology assessment for children and young adults ages 2:9 through 21 years old. It is video based and composed of short video segments where individuals are asked to name or label items while the clinician listens to, and notes any articulatory or phonological errors. The test is broken down into three versions based on age. A separate video test is used for age groups 2:9 – 5:11, 6:0-9:11, and 10:0-21 years old. *The Articulation and Phonology VAT* is an accurate and reliable assessment for speech intelligibility that yields standard scores, percentile ranks, interpretation values, and test-age equivalents. Normative data of this test is based on a nationally representative sample of 1405 (typically developing) children and young adults in the United States.

Articulation and Phonology VAT

The video-based assessment tool is composed of short video segments focusing on 47 (younger group) - 55 (older group) target words. Sounds are located across all positions in initial, medial, and final position of words. Vowels are also assessed.

Testing Format

The Articulation and Phonology VAT is administered from a computer/laptop or tablet. The test is composed of short pre-recorded video segments, which contain 45-55 target words. Individuals are asked to name specific items in the videos. The clinician listens carefully to the production of each word and records any distortion, substitution, omission, etc. of the targeted sounds. The clinician also makes note of any phonological process, such as stopping, fronting, initial consonant deletion, or gliding. The assessment automatically yields a raw score, standard score, percentile rank and interpretation value.

Administration Time

Administration time for the assessment takes approximately 25-30 minutes.

Articulation and Phonology VAT Uses and Purpose

When considering evidence-based practice in relation to assessment tools, it is important to remember that tests are only reliable and valid relative to a purpose. For example, a test may be valid for one

purpose and invalid for another purpose. IDEA states, "(A)(iii) Assessments and other evaluation materials used to assess a child under this section are used for the purposes for which the assessments or measures are valid and reliable."

The Articulation and Phonology VAT should be used to assess children or young adults who have a suspected or previous diagnosis of a speech sound disorder. This tool will aid in the identification or continued diagnosis of an articulation or phonological disorder. Using videos as the subject material provides test-takers with an interactive medium to maintain engagement. The results of the *Articulation and Phonology VAT* provide comprehensive information on articulatory and phonological skills of children and young adults. By utilizing the *Articulation and Phonology VAT*, we are able to develop a better understanding as to how a student's articulation and phonology skills may impact their academic performance and progress in school. It presents with four essential purposes:

a) To help identify articulation impairment and/or phonological deficits and determine the degree of such deficits (e.g., initial IEP based evaluations);
b) To help determine strengths and weaknesses
c) To help document progress during intervention, measure treatment efficacy or re-evaluate articulation/phonological skills and performance as part of triennial IEP based reviews;
d) To help analyze articulation and phonological skills in children and young adults for research purposes

Code of Federal Regulations – Title 34: Education

34 C.F.R. §300.7 Child with a disability. (c) Definitions of disability terms. (11) Speech or language impairment means a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, that adversely affects a child's educational performance.

The Individual's with Disabilities Act (IDEA, 2004) states that when assessing a student for a speech or language impairment, we need to determine whether or not the impairment will negatively impact the child's educational performance. In order to determine whether an articulation or phonology impairment exists, we can collect a speech sample of the individual, and analyze intelligibility and the impact of the impairment on academic success.

Chapter

2

Theoretical Background of the Articulation and Phonology VAT

Articulation and phonological disorders typically occur in preschool and school-aged children between 2:0 and 21:0 years old. When articulation or phonological errors occur, there can be negative impacts to a child's academic performance. Previous research has suggested that students with early speech-language problems are behind their peers in reading, writing, and in other academics areas (Aram & Nation, 1980; King, Jones, Lasky, 1982; Hall & Tomblin, 1978). More specifically, preschool children with speech sound disorders are at a higher risk for difficulties with phonological awareness (PA), which can lead to difficulties with spelling and reading (Peterson, Pennington, Shriberg, & Boada, 2009; Bird, Bishop, Freeman, 1995; Nathan, Stackhouse, Goulandris, & Snowling, 2004). Additionally, children who present with speech sound disorders in kindergarten have been associated with lower literacy outcomes (Overby, Trainin, Smit, Bernthal, & Nelson, 2012). Data from the National Health Interview Survey (2012) estimated that almost half (48.1%) of the children between the ages of 3 and 10 years-old who have been classified as having a communication disorder present with speech sound disorders only (Black, Vahratian, & Hoffman, 2015). Moreover, a recent large-scale study revealed that 18% of 8-year-old children present with unresolved speech sound errors (Roulstone, Miller, Wren, & Peters, 2009) and additional reports suggest that 11% to 40% of children with speech sound disorders also have a related language impairment (Eadie et al., 2015; Shriberg et al., 1999).

There is a clear need for assessment tools that aid in the identification of articulation and phonological disorders because without appropriate assessment and intervention, there can be serious impacts on a child's academic performance. According to the American-Speech-Hearing Association (2016), assessment of speech sound disorders includes the evaluation of accurate productions, specifically, a student's ability to produce sounds in various positions (initial, medial, final) and in different phonetic contexts, sound combinations such as consonant clusters or blends, and syllable shapes such as simple consonant-vowel-consonant (CVC) to more complex consonant-consonant-vowel-consonant-consonant (CCVCC). Assessment should also evaluate sound errors and looks at the type of errors (e.g., omission, substitution, distortion), the consistency of sound errors, and the distribution of errors (e.g., position of the sound in a word) (ASHA, 2016). Lastly, assessment of speech sounds should look at error patterns, specifically, phonological patterns and whether there is a systematic sound change or simplification that affects a class of sounds (e.g., stops), sound combinations (e.g., consonant clusters), or syllable structures (e.g., multisyllabic words) (ASHA, 2016).

Articulation and phonology disorders can have adverse effects on various aspects of language development, as well as academic performance, and peer relationships. For example, a child who feels embarrassed about their speech sounds may avoid social situations or conversations that require them to verbally communicate, which may result in a social language impairment. It is important that speech and language assessments be efficient and accurate to best serve our students. By assessing students with the *Articulation and Phonology VAT*, speech-language pathologists can better identify those individuals who have a suspected or an existing diagnosis of a speech sound disorder.

Contextual Background for Articulation and Phonology VAT

A speech sound disorder is an umbrella term that refers to the difficulty, or combination of difficulties, with perception, production, and/or phonological representation of speech sounds and speech segments (American-Speech-Hearing Association [ASHA], 2016). When these speech sound disorders do not have a known cause, they are referred to as either articulation or phonological disorders. Difficulties in articulation may result in sound distortions, substitutions, and omissions of individual speech sounds (ASHA, 2016). Phonological errors are predictable and result from difficulties in the comprehension and use of a speech sound system and it's governing rules (Bauman-Waengler, 2004). For example, a child with a phonological disorder may engage in final consonant deletion or fronting of speech sounds. The current assessment tool is composed of target words that address articulatory and phonological speech sound errors. Table 1.1 reviews common phonological processes.

Table 1.1 Phonological Processes

Phonological Process	***Definition/Example***
Backing	An alveolar sound (e.g., /t/ and /d/) is substituted with a velar sound (e.g., /k/ and /g/)
Fronting	A velar or palatal sound (e.g., /k/, /g/, and /ʃ/) is substituted with an alveolar sound (e.g., /t/, /d/, and /s/)
Gliding	An /r/ becomes a /w/, or /l/ becomes a /w/ or /j/ sound
Stopping	A fricative (e.g., /f/ or /s/) or affricate (e.g., /tʃ/) is substituted with a stop consonant (e.g., /p/ or /d/)
Assimilation	A consonant sound starts to sound like another sound in the word
Reduplication	A complete or incomplete syllable is repeated
Cluster Reduction	A consonant cluster is reduced to a single consonant
Initial Consonant Deletion	The initial consonant in a word is left off
Final Consonant Deletion	The final consonant in a word is left off
Syllable Deletion	The weak syllable in a word is deleted

Chapter

3

Administration and Scoring Procedures

The following testing guidelines represent specific administration, scoring, and interpretation instructions for the *Articulation and Phonology VAT*. These procedures are considered best professional practice required in any type of standardized assessment as described in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014). Strict standardized administration procedures must be followed to obtain reliable and accurate results. The Standards for Educational and Psychological Testing specifically emphasizes the importance of adhering to specific standardization procedures (Standard 6.1) and documenting deviations from the standardization procedures (Standard 6.3).

Examiner Qualifications

Professionals who are formally trained in the ethical administration, scoring, and interpretation of standardized assessment tools, who hold appropriate educational and professional credentials, may administer the *Articulation and Phonology VAT*. Qualified examiners include speech-language pathologists, clinical fellows and graduate students in speech-language pathology. It is a requirement to read and become familiar with the administration, recording, and scoring procedures before using this test.

Confidentiality Requirements

As described in Standard 6.7 of the Standards for Educational and Psychological Testing (AERA et al., 2014), it is the examiner's responsibility to protect the security of all testing material and ensure confidentiality of all testing results.

Eligibility for Testing

The *Articulation and Phonology VAT* is appropriate to use for individuals between the ages of 2:9 and 21:0 years of age. This assessment tool is particularly helpful for individuals who are suspected of or who have been previously diagnosed with articulation and phonological disorders.

Testing Time

Administration of the *Articulation and Phonology VAT* takes approximately 25-30 minutes.

Additional Testing Considerations and Procedures

A. Seating arrangement is important when administering this test because both the examiner and the student need to be able to see the videos. The examiner must be able to face the student during testing in order to closely observe his/her use of articulation.

B. Administer the test in a quiet, comfortable environment with no distractions. Stop testing if the student appears to be tired or is unwilling to participate.

C. It is important to elicit the examinees' best effort on each test and on each item presented. This can be achieved by establishing rapport with the examinee before the testing begins and by providing praising prompts when needed.

D. Because this is not a timed test, examinees should be allowed time to respond. However, if no response is provided within 10 seconds of presentation of an item, the clinician should prompt the examinee to imitate the target word.

E. If the examiner has reasons to believe that the testing results are invalid, such as poor attention span that is noticeably different from those expected, or student showing sign of being ill, retest at a later time.

Repetition of Video Test Items

Repetition of videos is allowed to the reasonable extent needed to elicit the target word.

Repetition of Item Questions

Repetition of item questions is allowed when the examinee appears to not understand the question or requests a repetition. If the examinee does not provide a response after the second reading, proceed to imitation prompts.

Prompting Rules

The Articulation and Phonology VAT test items were designed to be easily recognizable by young children and older students. However, when responses are ambiguous or other than the target word, the examiner should attempt to elicit the target word by giving prompts. For example, if the target word is *spider*, the examiner may prompt by saying, "*It's an insect that builds webs, what is it*?" If the examinee does not provide the target word after the prompt, the examiner should proceed with imitation prompting such as, "*This is a spider. What is it?*". It is important to note that the use of prompting or imitation does not affect the scoring.

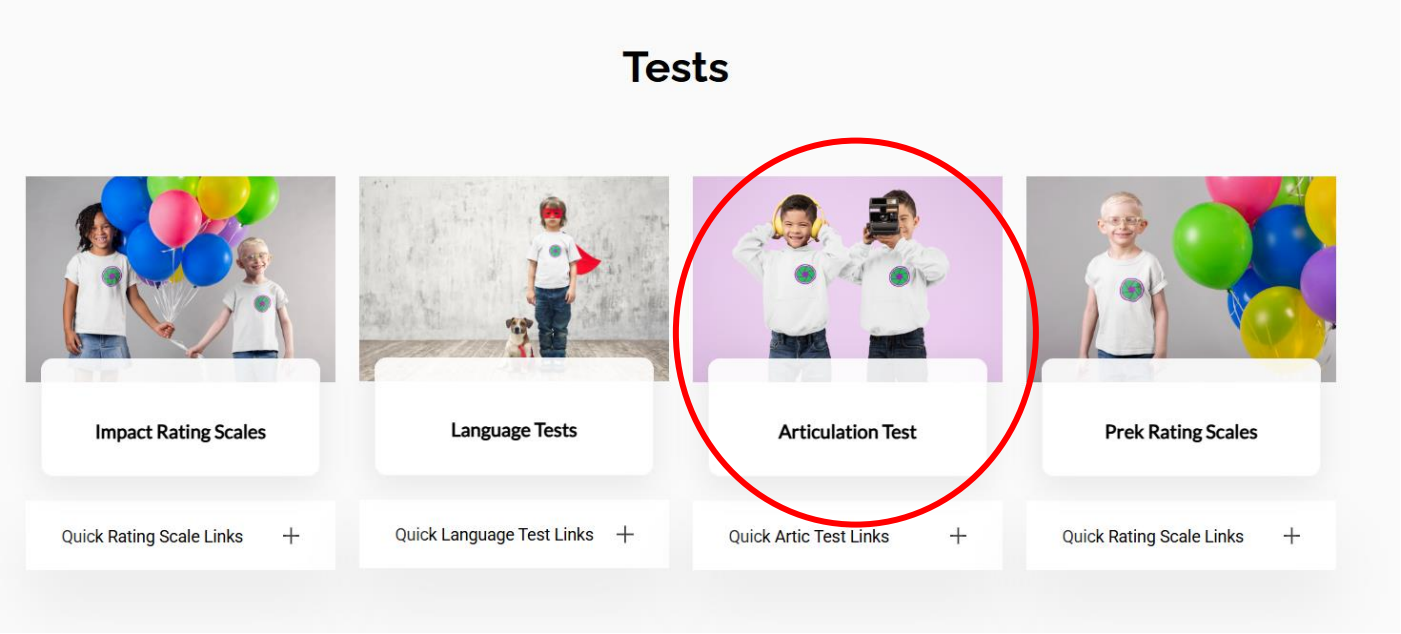
Test Materials

The *Articulation and Phonology VAT* is broken down into three sections based on age. A separate video test is used for age groups 2:9-5:11, 6:0-9:11 and 10:0-21 years old. The *Articulation and Phonology VAT* is accessible online (www.SLPplatform.com)

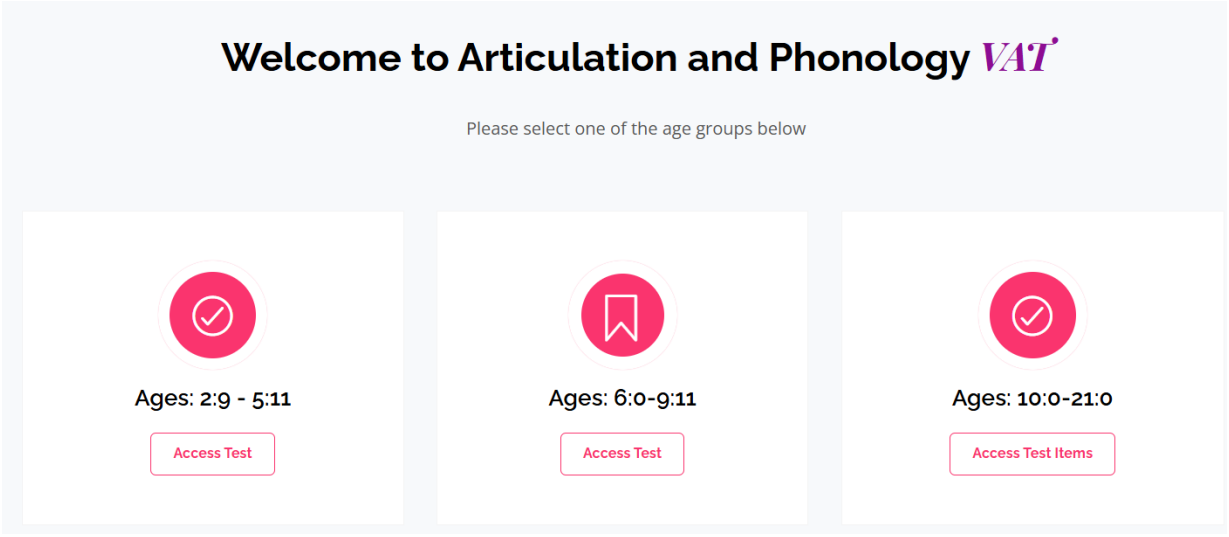
Accessing Articulation and Phonology VAT online

Begin by logging onto your account online at www.SLPplatform.com.

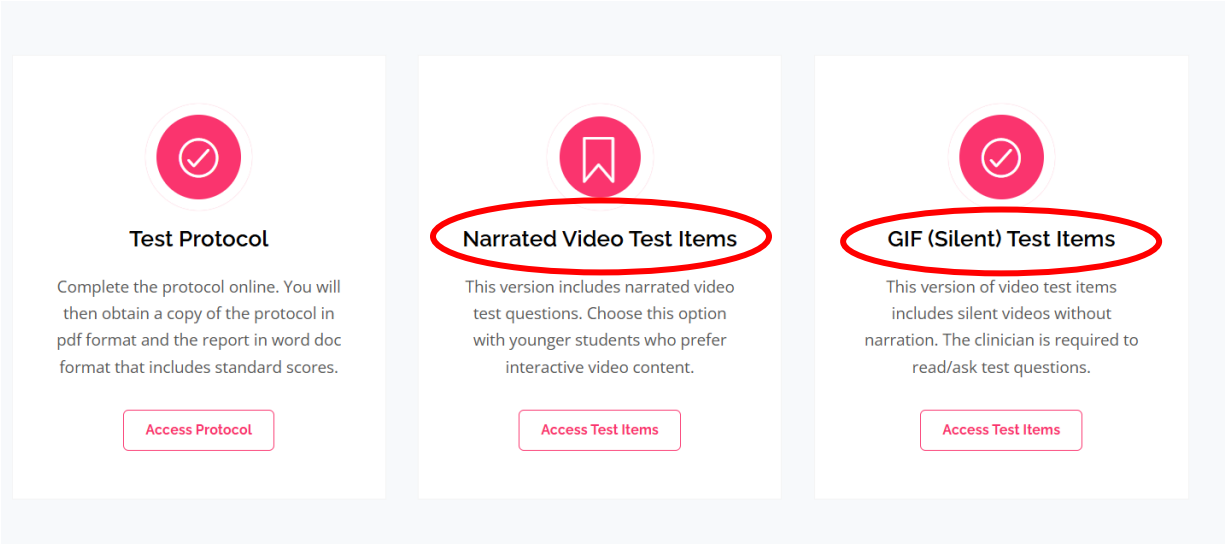
Next, select the "Administer Tests" tab. Select the *Articulation and Phonology VAT* by clicking on the picture.



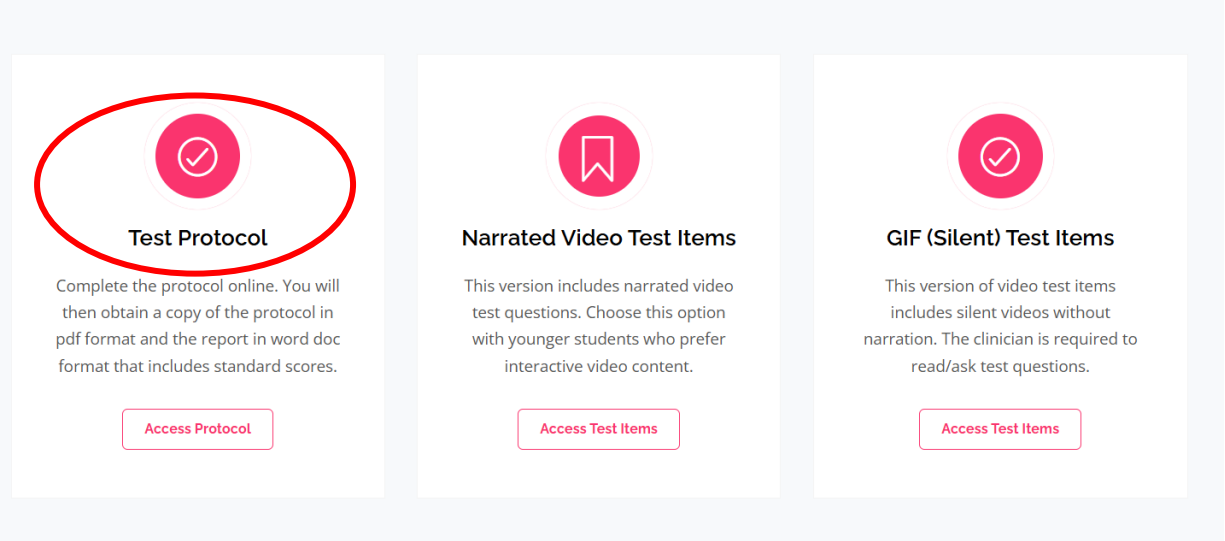
Step 1: After you have selected the *Articulation and Phonology VAT*, select the test that targets your student's age range.



Step 2: Once the assessment loads, there will be two viewing options available. *Option A* allows you to use narrated videos. *Option B* allows you to administer the test using silent videos. This option requires that the clinician read the test item questions. Both options contain the same target words, choose the option you prefer. You will need to use two devices to administer the test. One device is required to record the responses online. The second device is required to play the video-based test items.



Select the “Test Protocol” tab to access the protocol:



Step 3: Begin administering the test. Tell the examinee that, “***You are going to see some videos. Please look at the videos and answer the questions. Are you ready?***” Start with the practice item. On the protocol, follow the written instructions to record the examinee’s production errors such as sound distortions, substitutions, omissions, etc. and make note of phonological processes. When you are finished with the administration of the test along with the completion of the protocol, click on the “Submit” button. The system will generate a scored protocol that contains standard scores and percentile ranks. A complete articulation/phonological report will be generated as well. Enter your own (the examiner’s) email address to receive a copy of the protocol and report by email.

Chapter

4

Development, Standardization and Normative Information

This section describes the procedures followed in developing test items, implementing the pilot and normative study, and selecting the items for the final version of the test. This section also details the normative samples obtained to standardize and validate the *Articulation and Phonology VAT*. All test development and standardization project procedures were reviewed and approved by IntegReview IRB (now known as Advarra), a fully AAHRPP-accredited independent review board that provides ethical review for all phases of industry-sponsored and federally funded research in the U.S. Additionally, all test development and standardization methodology was based on best practices in research, and conducted in compliance with complex regulatory requirements, frameworks, and guidelines set forth by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014).

Test Item Development

The words chosen for the *Articulation and Phonology VAT* represent concepts that are unambiguous and easily recognizable in video format, that are used to elicit and measure target sounds. Selection of the target sound inventory began with a review of research and theory related to the development of articulation error patterns as well as an analysis of which articulation errors are most predictive of an articulation impairment for specific age groups. Similar process was employed during the phonology item selection. This analysis resulted in identification of three separate sets of words for ages 3:0 – 5:11 (57 items), 6:0 – 9:11 (62 items) and 10:0-12:11 (64 items) containing target consonants, consonant blends, vowels, diphthongs and vocalic /r/ in the initial, medial and/or final position of words, as well as 8 phonological error patterns. The selected test items were designed to be developmentally appropriate for use with young children and older students. The interactive format of videos minimizes the reluctance of many young children to speak to an unfamiliar examiner. Additionally, the test items and their videos were chosen to be clear and easy to recognize for young children. The New General Service List (Browne et. al, 2013) and the Reading Teacher's Book of Lists, Fifth Edition (Fre et. al, 2006) were used to ensure that only high-frequency and commonly used words were included in the test item inventory.

The test was developed in three phases: pilot study, normative study, and national standardization. The procedures for each phase are detailed below.

Pilot Study

The pilot study was conducted to determine the appropriateness of all video test stimuli and to review all test instructions. The purpose of the study was to determine how readily children of varying ages could name the target words portrayed in the presented videos.

The pilot study included 94 children from the ages of 3:0 to 12:11. The sample was 21% Hispanic, 9% African American, 53% White, 4% Asian and 13% other ethnicities (60% males and 40% females). The pilot study included 75% typically developing children and 25% children with identified articulation impairment and/or phonological disorders.

The examinee responses were coded as correct, prompted (correct with verbal prompts), imitated (correct after imitation) or incorrect. At that time, the articulation and phonological errors were not tracked. The results of the pilot study were found to be effective for test item selection and for the interactive aspect of the video-based format.

Item analyses were implemented based on Classical Test Theory (CTT) and Item Response Theory (IRT) which are psychometric tools used to refine and evaluate test items to ensure that they effectively elicit target words. Both CTT and IRT are foundational frameworks in psychometrics that emphasize the importance of test standardization. The CTT analyses reviewed the proportion of children answering the items correctly as well as the correlation of item performance with raw score. Additionally, all test items were analyzed using the Rasch one-parameter model. While CTT assumes that all items in an assessment instrument make an equal contribution to the performance of examinees, IRT takes into the consideration the fact that some items are more difficult than others. This means that the probability of success on items is due both to student ability and also to item difficulty. As a result of these analyses, some items were eliminated and items that were identified to be difficult to name for certain age groups, were assigned to the older examinee group (10:0-12:11).

Normative Study

Following the pilot study, a normative study was conducted to establish norms for the Articulation and Phonology VAT by testing typically developing children representative of the general U.S. population. A clinical group was included for validation purposes. Additional goals of the normative study included investigation of optimal weighted scoring system/criteria as well as optimal test administration time. The study reviewed administrative and scoring procedures preliminary to national standardization. The test content was evaluated both qualitatively and quantitatively for bias.

The normative study included 168 children from the ages of 3:0 to 15:11. The sample was 12% Hispanic, 8% African American, 57% White, 7% Asian and 16% other ethnicities (60% males and 40% females). The pilot study included 88% typically developing children and 12% children with identified articulation impairment and/or phonological disorders.

The goals of the normative study were achieved. Since an articulation and phonology test is designed to identify those examinees with articulation and phonological disorders, it would be expected that

individuals identified as likely to exhibit articulation/phonological deficits would score lower than those who are typically developing. The mean for the outcome variables were compared between the clinical and the typically developing groups of examinees using Kruskal Wallis analysis of variance (ANOVA). Further comparisons in mean scores between the groups were examined using Mann-Whitney U test. The level of significance was set at $p \leq 0.05$. Further comparisons using Mann- Whitney U test showed that there was a significant difference among all the study groups ($p < 0.001$).

Based on the responses of all examinees, some test items were modified, while others were removed altogether. The test directions and scoring procedures were fine-tuned. Suggestions of the field test examiners were thoroughly reviewed prior to the national standardization.

Scoring

The Articulation and Phonology VAT uses a weighted scoring system. Each target sound of the test is assigned a value based on factors that are most predictive of an articulation impairment and/or phonological disorder: frequency of occurrence, syllable position as well as age of acquisition (Edwards et at., 2009; Eisenberg et al., 2010; McLeod, 2018; Bankson et al., 2019; Crowe et al., 2020). For example, misarticulation of /s/ is assigned a greater (more significant) value than /θ/ because of its higher frequency of occurrence and earlier sound mastery age. Similarly, in the phonology domain, sound omission is assigned a greater (more significant) value than sound substitution because this phonological error pattern is eliminated as a phonological process at a significantly earlier age compared to sound substitution and has a greater impact on speech intelligibility.

National Standardization

One of the ways we can tell if an assessment is a strong test, is if it includes adequate norms. Norm-referenced testing is a method of evaluation where an individual's scores on a specific test are compared to scores of a group of test-takers (e.g., age norms) (AERA, APA, and NCME, 2014). Previous research has suggested that utilizing a normative sample can be beneficial in the identification of a disability. Additionally, research has suggested that the inclusion of children with disabilities in the normative sample may negatively impact the test's ability to differentiate between children with disorders and children who are typically developing (Peña, Spaulding, & Plante, 2006). When reviewing a test's normative sample, it is important to consider size, gender, race and ethnicity, age, geographic location, and whether individuals with disabilities were included in the normative sample.

The national standardization consisted of 2 phases. The first phase of the normative data collection for the *Articulation and Phonology VAT* was based on the test performance of 1405 examinees across 11 age groups (shown in Table 4.1) in 17 states (Arizona, California, Colorado, Nevada, Idaho, Illinois, Iowa, Kansas, Ohio, Minnesota, Florida, New York, Pennsylvania, Florida, South Carolina, Texas, Washington).

Table 4.1: Representation of the Sample, by Age Group

Age Group	Age	*N*	%
1	3-0 to 5-11	312	22
2	6-0 to 6-11	128	9
3	7-0 to 7-11	133	9
4	8-0 to 8-11	121	9
5	9-0 to 9-11	102	7
6	10-0 to 10-11	98	7
7	11-0 to 11-11	96	7
8	12-0 to 12-11	84	6
9	13-0 to 13-11	94	7
10	14-0 to 14-11	96	7
11	15-0 to 21-0	141	10
Total		1405	100%

The second phase of the normative data collection for the *Articulation and Phonology VAT* was based on the test performance of additional 94 examinees ages 2:9 through 2:11 years old (shown in Table 4.3) in 5 states (California, Ohio, Illinois, New York, Florida).

The data was collected throughout the 2016-2023 school years by 34 state licensed speech-language pathologists (SLPs). The SLPs were recruited through Go2Consult Speech and Language Services, a speech-language pathology services and nonpublic agency certified by the CA Department of Education in conjunction with the Lavi Institute, an ASHA approved CE provider. All standardization project procedures were reviewed and approved by IntegReview IRB (now known as Advarra), a fully AAHRPP-accredited independent review board that provides ethical review for all phases of industry-sponsored and federally funded research in the U.S. To ensure representation of the national population, the *Articulation and Phonology VAT* standardization sample was selected to match the US Census data reported in the ProQuest Statistical Abstract of the United States, 2017 (ProQuest, 2017). The sample was stratified within each age group by the following criteria: gender, race or ethnic group and geographic region. The demographic table below (Table 4.2) specifies the distributions of these characteristics and shows that on the whole, the sample is nationally representative.

Table 4.2: Demographics of the Normative Sample vs. US Population

Normative Sample Size = 1405

Demographic	N Normative Sample	% Normative Sample	% US Population
Gender			
Male	704	50.1%	49%
Female	701	49.9%	51%
Total	1405	100%	100%
Race			
White	857	61%	77%
Black	197	14%	13%
Asian	70	5%	4%
Other	98	7%	6%
Hispanic	183	13%	12%
Total	1405	100%	100%
Clinical Groups			
	none	none	none
US Regions			
Northeast	235	17%	16%
Midwest	296	21%	22%
South	508	36%	38%
West	366	26%	24%
Total	1405	100%	100%
Parents' Educational Level			
Four years of college or more	393	28%	31%
Some college	393	28%	27%
High school graduate	436	31%	30%
Less than high school graduate	183	13%	12%
Total	1405	100%	100%

Table 4.3: Demographics of the Normative Sample (age group 2:9-2:11) vs. US Population

Normative Sample Size = 94

Demographic	N Normative Sample	% Normative Sample	% US Population
Gender			
Male	56	60%	49%
Female	38	40%	51%
Total	94	100%	100%
Race			
White	59	63%	77%
Black	11	12%	13%
Asian	5	5%	4%
Other	8	8%	6%
Hispanic	11	12%	12%
Total	94	100%	100%
Clinical Groups			
	none	none	none
US Regions			
Northeast	10	11%	16%
Midwest	22	23%	22%
South	34	36%	38%
West	28	30%	24%
Total	94	100%	100%
Parents' Educational Level			
Four years of college or more	28	30%	31%
Some college	28	30%	27%
High school graduate	27	29%	30%
Less than high school graduate	11	11%	12%
Total	94	100%	100%

Criteria for inclusion in the normative sample

A good assessment is one that yields results that will benefit the individual being tested or society as a whole (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014). One way we can tell if an assessment is a good test, is if it includes adequate norms. Previous research has suggested that utilizing a normative sample can be beneficial in the identification of a disability and that the inclusion of children with disabilities may negatively impact the test's ability to differentiate between children with disorders and children who are typically developing (Peña, Spaulding, & Plante, 2006). Since the purpose of the *Articulation and Phonology VAT* is to help to identify students who present with articulation and phonological deficits, it was critical to exclude students from the normative sample who have diagnoses that are known to influence articulation and

phonology (Peña, Spaulding, & Plante, 2006). Thus, students who had previously been diagnosed with articulation, phonological impairments, or motor planning deficits were not included in the normative sample. In order for students to be included in the normative sample for this assessment tool, students must have met criteria of having typical articulation and phonological development, and show no evidence of speech intelligibility difficulties. Students used in the present normative sample had no other diagnosed disabilities and were not receiving speech and language support or any other services. Thus, the normative sample for the *Articulation and Phonology VAT* provides an appropriate comparison group (i.e., a group without any known disorders that might affect articulation/phonology) against which to compare students with suspected disorders. The *Articulation and Phonology VAT* is designed for students who are native speakers of English and/or are English language learners (ELL) who have demonstrated a proficiency in English based on state testing scores and school district language evaluations. Students who were native English speakers and also spoke a second language were also included in this sample.

Norm-referenced testing is a commonly used method of evaluation that compares an individual's scores on a specific test to the scores of a group of test-takers (e.g., age norms) (AERA, APA, and NCME, 2014). Clinicians can compare students' performance on the *Articulation and Phonology VAT* to this normative sample to determine whether a student is scoring within normal limits or, if their scores are indicative of an articulation and/or phonological impairment. Administration, scoring, and interpretation of the *Articulation and Phonology VAT* must be followed in order to make comparisons to normative data.

Item Analysis and Final Item Selection

After the completion of phase II of the national standardization project, further item analyses were implemented based on Classical Test Theory (CTT) and Item Response Theory (IRT). The CTT analyses reviewed the proportion of children answering the item correctly as well as the correlation of item performance with raw score. Additionally, all test items were analyzed using the Rasch one-parameter model. As a result of these analyses, some items were eliminated and/or replaced. Appendix A contains a final summary of all test words and sounds assessed in each age-group.

Chapter

5

Validity and Reliability

This section of the *Articulation and Phonology VAT* manual provides information on the psychometric characteristics of validity and reliability. Validity helps establish how well a test measures what it is supposed to measure and reliability represents the consistency with which an assessment tool measures certain ability or skill. The first half of the chapter evaluates content, construct, criterion, and clinical validity of the *Articulation and Phonology VAT*. The latter half of the chapter reviews the consistency and stability of the *Articulation and Phonology VAT* scores, in addition to test retest and inter-rater reliability.

Validity

When considering the strength of a test, one of the most important aspects to consider is validity. Content validity refers to whether the test provides the clinician with accurate information on the ability being tested. Specifically, content validity measures whether or not the test actually assesses what it says it's supposed to assess. According to McCauley and Strand (2008), there should be a justification of the methods used to choose content, expert evaluation of the test's content, and an item analysis. Content-oriented evidence of validation addresses the relationship between a student's learning standards and the test content. Specifically, content-sampling issues take a look at whether cognitive demands of a test are reflective of the student's learning standard level. Additionally, content sampling may address whether the test avoids inclusion of features irrelevant to what the test item is intended to target.

Discriminant Analysis and Cut Scores

In the past, it was believed that when the purpose of an assessment was to *identify an individual with a disorder*, we would expect the student to score on the lower end of the distribution. However, this is not always the case and often times, students do not score on the lower half of the normal distribution. When we consider the normal distribution, it is skill level that exists on a continuum, not whether or not a disorder is present. If the purpose of the assessment is to identify whether a disorder is present, we must look at discriminant analysis – which is a distribution of scores for "typically developing" and distribution for "impaired" individuals. At some point along these two distributions there should be a point that maximally discriminates across the two groups and this is known as the ***cut score***. Above the cut score, individuals are classified as typically developing, and below the cut score, individuals are classified as impaired. Depending on the test, the two distributions may differ on how much or how little they overlap, but there will always be a point where maximal discrimination takes place. Cut-scores are test-specific and a specific test score will only work with that specific test.

Single-cut Scores

It is often common practice to use single cut scores (e.g., -1.5 standard deviations) to identify disorders, however, this is not evidence-based and there is actually evidence that advises against using this practice (Spauling, Plante, & Farinella, 2006). When using single cut scores (e.g., -1.5 SD, -2.5 SD, etc.) we may under identify students with impairments on tests for which the best-cut score is higher and over identify students impairments on tests for which the best-cut score is lower. Additionally, using single cut scores may go against IDEA's (2004) mandate, which states assessments must be valid for the purpose for which they are used.

Inclusion/Exclusion Criteria for the Discriminant Analysis and the Group Differences Study

Typically developing participants were selected based on the following criteria: 1) exhibited hearing sensitivity within normal limits; 2) presented with age-appropriate speech and language skills; 3) successfully completed each school year with no academic failures; and 4) attended public school and placed in general education classrooms.

Inclusion criteria for the articulation impairment group was: 1) having a current diagnosis of articulation impairment or delay (based on medical records and/or school-based special education eligibility criteria); 2) currently attending a local public school, and enrolled in the general education classroom; and 3) exhibited hearing sensitivity within normal limits.

Inclusion criteria for the articulation impairment secondary to hearing loss group was: 1) having a current diagnosis of articulation impairment or delay (based on medical records and/or school-based special education eligibility criteria); 2) currently attending a local public school, and enrolled in the general education classroom; and 3) exhibited hearing loss based on medical records and audiologist reports.

Finally, the inclusion criteria for the phonological group was: 1) having a current diagnosis of speech impairment (based on medical records and/or school-based special education eligibility criteria, and exhibiting at least two documented phonological processes that impact speech intelligibility); 2) being enrolled in the general education classroom based on medical records;

Sensitivity and Specificity

As a result of cut scores, we are provided with information on sensitivity and specificity. *Sensitivity* refers to the ability of a test to identify impaired individuals as impaired and it is calculated by the number of individuals the test identifies as impaired divided by the number of truly impaired individuals. *Specificity* refers to the ability of a test to identify normal individuals as normal and is calculated by the number of people the test identifies as not-impaired divided by the number of truly not-impaired individuals.

Table 5.1 shows the cut scores needed to identify articulation and phonological disorders within each age range. Additionally, this table demonstrates the sensitivity and specificity information that indicates the accuracy of identification at these cut scores. Sensitivity and specificity are diagnostic validity statistics that explain how well a test performs. Vance and Plante (1994) set forth the standard that for a language assessment to be considered clinically beneficial, it should reach at least 80% sensitivity and specificity. Thus, strong sensitivity and specificity (i.e., 80% or stronger) is needed to support the use of a test in its identification of the presence of a disorder or impairment. Sensitivity measures how well the assessment will accurately identify those who truly have an articulation or phonological disorder (Dollaghan, 2007). If sensitivity is high, this indicates that the test is highly likely to identify the articulation or phonological disorder, or, there is a low chance of "false positives." Specificity measures the degree to which the assessment will accurately identify those who do not have an articulation or phonological disorder, or how well the test will identify those who are "typically developing" (Dollaghan, 2007).

Table 5.1 *Articulation and Phonology VAT* sensitivity, specificity and likelihood ratios

Age group	Cut score	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
3:0-3:11	77	88	87	4.29	.09
4:0-4:11	77	87	86	6.12	.14
5:0-5:11	77	88	87	6.12	.13
6:0-6:11	78	89	88	4.29	.12
7:0-7:11	77	91	88	4.87	.08
8:0-8:11	77	92	90	4.29	.09
9:0-9:11	77	91	89	6.07	.18
10:0-10:11	77	89	90	6.12	.13
11:0-11:11	78	88	89	6.09	.11
12:0-12:11	77	92	89	6.07	.18
13:0-13:11	77	92	90	6.07	.18
14:0-14:11	78	92	91	6.07	.18
15:0-15:11	77	92	89	6.48	.12
16:0-21:0	77	94	96	7.27	.16

Note: Total N=2733; typically developing group n=1405; clinical group=1328

Content Validity

The validity of a test determines how well the test measures what it purports to measure. Validity can take various forms, both theoretical and empirical. This allows us to compare an instrument with other measures or criteria, which are known to be valid (Zumbo, 2014). For the content validity of the test, expert opinion was solicited. Twenty-six speech language pathologists (SLPs) were asked to review the *Articulation and Phonology VAT*. All SLPs were licensed in the state of California, held the Clinical Certificate of Competence from the American Speech-Language-Hearing Association, and had at least 5 years of experience in assessment of children with articulation and phonological impairments. Each of these experts was presented with a comprehensive overview of test description, as well as rules for standardized administration and scoring. They all reviewed 8 full-length administrations. Following

this, they were asked 30 questions related to the content of the video assessment tool and whether they believed the assessment to be an adequate measure of articulation and phonology skills. For instance, their opinion was solicited regarding whether the questions and the raters' responses properly evaluated production of consonant clusters. The reviewers rated test items on a decimal scale. All reviewers agreed that the *Articulation and Phonology VAT* is a valid measure of articulation and phonology, in students who are ages 3 to 21 years. The mean ratings were 28.1±1.3.

Construct Validity

Developmental Progression of Scores

Articulation and phonology is developmental in nature and skills change with age. Mean raw scores for examinees should increase with chronological age, demonstrating age differentiation. Mean raw scores and standard deviations for the *Articulation and Phonology VAT* are divided into thirteen age intervals displayed in Table 5.2
below.

Table 5.2: Normative Sample's mean raw scores and standard deviations on the *Articulation and Phonology VAT*

Age Group	Mean Raw Scores and Standard Deviations
3:0-3:11	32 (2.7)
4:0-4:11	28 (2.8)
5:0-5:11	19 (2.4)
6:0-6:11	4 (1.7)
7:0-7:11	4 (1.4)
8:0-8:11	3 (1.2)
9:0-9:11	3 (1.0)
10:0-10:11	2 (0.9)
11:0-11:11	1 (0.8)
12:0-12:11	1 (0.8)
13:0-13:11	1(0.7)
14:0-14:11	0 (0.2)
15:0-15:11	0 (0.2)
16:0-21:0	0 (0.2)

Progressions of speech sound development

The table below displays the average age that 90% of all children in the normative sample produced each consonant, vowel and consonant cluster sound correctly. Note: A child's speech may be considered delayed if he/she has not developed the sound by the end of each given age. Intervention should be considered when speech sound production has a negative impact on academics, social interactions with peers, and/or draws negative attention to a child's speech.

Table 5.7
90% Mastery Ages for Articulation of Target Sounds

Age	Consonants	R and R Vowels	Vowels	Consonant Clusters
3:0	b,p,m,n,w d,t,h		ɒ, ɪ, ʊ, æ, aʊ,l,u, oʊ	
3:6-3:11	k,g,f, ŋ			tw
4:6-4:11	v,ʤ,j			kw
5:0-5:6	s,z			sw, st
5:6-5:11	tʃ, ʃ, ð			
6:0-6:6	l			kl,bl,pl,gl,sl
6:6-6:11		ɚ, ru		kr,br,tr,dr,fl
7:0-7:6		r, ra, ɛr		
8:0-8:11	θ			

Criterion Validity

In assessing criterion validity, the *Articulation and Phonology VAT* was correlated to other measures of articulation and phonology: *Arizona Articulation and Phonology Scale - Fourth Edition* (Arizona-4; Fudala & Stegall, 2017) and the *Diagnostic Evaluation of Articulation and Phonology* (DEAP; Dodd, Holm, Crosbie, & Ozanne, 2003). Time between test administrations ranged from the same day to 5 days.

The concurrent validity was assessed using Pearson's correlation among all measures. Correlation coefficients of ≥0.7 are recommended for same-construct instruments while moderate correlations of ≥ 0.4 to ≤0.70 are acceptable. The level of significance was set at $p \leq 0.05$. When assessing validity, the *Articulation and Phonology VAT* was substantially correlated with the *DEAP* and the *Arizona-4*: 0.87, and 0.83 respectively, $p < 0.001$.

Table 5.3: Pearson's Correlations between three measures of articulation and phonology (N = 21)

Articulation and Phonology Tests	DEAP	Arizona-4
Articulation and Phonology VAT †	.87	.83

Abbreviations: DEAP, Diagnostic Evaluation of Articulation and Phonology (2006); Arizona-4, Arizona Articulation and Phonology Scale - Fourth Edition (2017).
† Significant at an alpha of 0.001 level of significance.

Group Differences

Since an articulation and phonology assessment tool is designed to identify those examinees with articulation and/or phonological impairments, it would be expected that individuals identified as likely to exhibit articulation/phonological impairments would score lower than those who are typically developing. The mean standard scores for three clinical groups of examinees (articulation impairment, articulation impairment secondary to hearing loss, phonological [reduced speech intelligibility due to multiple phonological processes]) were administered the *Articulation and Phonology VAT* and are listed in Table 5.4. The mean for the outcome variables were compared among the three clinical groups and the typically developing group of examinees using Kruskal Wallis analysis of variance (ANOVA). The level of significance was set at $p \leq 0.05$. Table 5.4 reviews the ANOVA, which reveals a significant difference between all three groups.

Table 5.4: Scaled Score Means (and Standard Deviations) of Subtests for Three Clinical Groups and a Demographically Matched Typically Developing Group, (N = 193)

	AI (n=52)	HL (n=36)	SIP (n=49)	TD group (n=56)	p –value*
Age Range: 3-5 years [a,b,c]	42 (2.9)	54 (3.4)	41 (3.6)	29 (3.1)	<.001
Age Range: 5-6 years [a,b,c]	39 (3.1)	51 (3.1)	43 (3.9)	19 (2.9)	<.001
Age Range: 6-8 years [a,b,c]	20 (3.1)	48 (3.9)	39 (3.4)	4 (1.5)	<.001
Age Range: 8-10 years [a,b,c]	9 (3.1)	14 (2.7)	11 (3.8)	3 (1.7)	<.001
Age Range: 10-14 years [a,b,c]	6 (2.9)	10 (2.6)	8 (1.8)	1 (0.6)	<.001

Abbreviation: AI, articulation impairment; HL, hearing loss; SIP, speech impairment secondary to phonological processes. TD, typically developing.
* Kruskal-Wallis Analysis of Variance test
[a] significant difference between AI and TD groups
[b] significant difference between HL and TD groups
[c] significant difference between SIP and TD groups

Standards for fairness

Standards of fairness are crucial to the validity and comparability of the interpretation of test scores (AERA, APA, and NCME, 2014). The identification and removal of construct-irrelevant barriers maximizes each test-taker's performance, allowing for skills to be compared to the normative sample for a valid interpretation. Test constructs and individuals or subgroups of those who the test is intended for must be clearly defined. In doing so, the test will be free of construct-irrelevant barriers as much as possible for the individuals and/or subgroups the test is intended for. It is also important that simple and clear instructions are provided.

Inter-rater Reliability

Inter-rater reliability measures the extent to which consistency is demonstrated between different raters with regard to their scoring of examinees on the same instrument (Osborne, 2008). For the *Articulation and Phonology VAT*, inter-rater reliability was evaluated by examining the consistency with which the examiners are able to follow the test scoring procedures. Two clinicians simultaneously rated students. The results of the scorings were correlated. The coefficients were averaged using the z-transformation method. The resulting correlations for the subtests are listed in Table 5.5.

Table 5.5: Inter-rater Reliability Coefficients, Articulation and Phonology VAT

Age Group	Reliability
Age Groups: 1, 2, & 3	.89
Age Groups: 4, 5, & 6	.91
Age Groups: 7, 8, 9, 10, & 11	.93

Test-Retest Reliability

This is a factor determined by the variation between scores or different evaluative measurements of the same subject taking the same test during a given period of time. If the test is a strong instrument, this variation would be expected to be low (Osborne, 2008). The *Articulation and Phonology VAT* was completed with 59 randomly selected examinees, ages 3-0 through 21-0 over two testing periods. The interval between the two periods ranged from 16 to 20 days. To reduce recall bias, the examiners did not inform the examiners at the time of the first testing session that they would be testing again. All subsequent testing sessions were completed by the same examiners who administered the test the first time. The results are listed in Table 5.6. The test-retest coefficients for the various age groups were all greater than .80 indicating strong test-retest reliability for the *Articulation and Phonology VAT*.

Table 5.6: Test - Retest Reliability, Mean Standard Scores and Standard Deviations

		1st Test		2nd Test		Correlation Coefficient
Age Groups	N	Mean	SD	Mean	SD	
1,2, & 3	21					
Clinician		99	2	100	2	0.88
4,5, & 6	20					
Clinician		101	1	101	1	0.92
7, 8, 9, 10 & 11	18					
Clinician		100	1	100	1	0.91

Effectiveness of Remote Virtual Assessment: The Articulation and Phonology Video Assessment Tool (VAT)

Over the past few years, the need for valid and reliable remote assessments has become more evident. In March 2020, we saw many schools and clinics around the world close their doors and turn to virtual speech and language services due to the COVID-19 pandemic. Now, as we are moving our way out of the pandemic, we are continuing to see virtual speech and language services. The reason, possibly, is because virtual speech and language services work (Gabel, Grogan-Johnson, Alvares, Bechstein, & Taylor, 2013) and can be more convenient for some families and individuals.

When we consider the individuals who are receiving speech and language services, the majority are in a critical period of speech and language development (Nicholas & Geers, 2006), and thus, it is crucial that services continue on in order to avoid negative effects on academic performance, peer relationships, and overall quality of life (Wales, Skinner, & Hayman, 2017; Taylor, Armfield, Dodrill, & Smith, 2014; Kaiser & Roberts, 2011). Previous research has suggested that tele-practice can be an effective model for assessment and treatment (Wales, Skinner, & Hayman, 2017; Keck & Doarn, 2014; Theodoros, 2008; Gabel, Grogan-Johnson, Alvares, Bechstein, & Taylor, 2013). Additionally, the American Speech-Language-Hearing Association (2020) has approved tele-practice as an appropriate method for the assessment and treatment of speech and language disorders. In order to feel confident in the accuracy, reliability, and validity of remote assessments, clinicians can evaluate how scores obtained during remote assessment compare to those scores obtained from in-person administration.

The present study compares speech sound performance results of in-person versus remote administrations of the *Articulation and Phonology Video Assessment Tool* (VAT). In order to examine the equivalency between in-person and remote assessments, a test-retest design was used for this study. Each individual who participated in this study was tested twice with the *Articulation and Phonology Video Assessment Tool* (VAT), once in-person and once remotely. The same clinician administered both the in-person and remote assessment for each participant. Additionally, the order of which assessment format (in person vs. remote) occurred was counterbalanced. The purpose of the present study is to determine if there are any significant differences in speech sound performance results when testing in-person compared to testing remotely. The present study will also evaluate rater-reliability by evaluating if there are any differences in the clinician's ratings of performance when testing occurs in-person vs. remotely.

The *Lavi Institute* provides a technical manual for the administration and scoring of the *Articulation and Phonology VAT.* It is a requirement that the clinician administering the test read and become familiar with the administration, recording, and scoring procedures before using this, or any, assessment tool.

METHOD

Participants

Eighty-eight children, aged 5 years, 0 months, to 13 years, 0 months participated in this study. The sample consisted of forty-four who were considered typically developing and forty-four with a previously diagnosed articulation/phonological disorder. Demographic characteristics are reviewed in Table 6. The study's sample was balanced for age, gender, and race or ethnic group.

Four examiners participated and administered the assessment used in this study. All examiners were state licensed, ASHA-certified speech-language pathologists (SLPs). The SLPs collected data from September 2020 to December 2022. The SLPs were recruited through *The Lavi Institute*, a research and professional development company. All examiners received compensation for their participation in the study. The eighty-eight participants were also recruited through *the Lavi Institute* and received compensation (e.g., gift card) for their participation.

Materials and Procedures

Prior to all in-person and remote assessments, parent consent was provided to assess each child. Parents also provided consent to have their child's data included for the purpose of this study. Examiners confirmed with parents the day before the remote assessment took place that each child had access to an electronic device, such as a laptop or tablet, with headphones and a built-in microphone. Remote administration was completed securely over the online Zoom platform. Individual meeting links with passwords were provided for each participant and additional licensing was provided for the examiner to secure HIPAA compliance.

The *Articulation and Phonology VAT* is composed of short pre-recorded video segments, which contain 45-55 target words. Therefore, clinicians used an electronic device during both in-person and remote administrations to access the video-based *Articulation and Phonology Video Assessment Tool.*

During remote assessment, the examiner used the screen-sharing feature on Zoom to present and administer the *Articulation and Phonology VAT*. After displaying a test item to the student, the examiner paused the test, stopped screen-share, and asked the student to name and/or label the item that was displayed. The clinician would then listen carefully to the production of each word and record any distortion, substitution, omission, or lisp of the targeted sounds. The clinician also made note of any phonological process, such as stopping, fronting, initial consonant deletion, or gliding. Then, the examiner would start screen-share again and move on to the next item and continue the process until all of the *Articulation and Phonology VAT* items were administered.

During each participant's first assessment, he/she was fully assessed using the *Articulation and Phonology VAT.* Each participant was then scheduled for his/her follow-up assessment at least three weeks later. A student's speech sound production skills are not expected to change significantly during this time period. Thus, the test-retest method is beneficial in comparing the results of a student's in-person versus remote speech sound performance. Additionally, due to this research design, the present study counterbalanced the order of the test format. For example, half of the participants in the typically developing group and half of the participants in the speech

sound disorder group received an in-person assessment the first time they were assessed and then received remote assessment the second time. The remaining participants received the remote administration the first time they were assessed and an in-person assessment on the second test date.

During both in-person and remote assessments, examiners recorded each participant's responses on the online digital protocol. The results of each assessment were then calculated on the test's website page. The *Articulation and Phonology VAT* yields a raw score, standard score, and percentile rank. Participants' standard scores from both testing formats were compared to obtain test-retest reliability. Raw scores from both testing conditions were used to obtain rater-reliability.

RESULTS

Test-retest reliability is the ability for a test to reveal the same score and/or diagnosis when given more than once over a short interval of time. This method was used to determine if the remote administration of the *Articulation and Phonology VAT* would reveal the same score and/or diagnosis as the in-person administration. The *Articulation and Phonology VAT* was administered twice to eighty-eight participants, aged 5 years, 0 months, to 13 years, 0 months, once in-person and once remotely. The interval between the two testing dates ranged from 20 to 25 days. Participants had the same examiner during the first and second administration. The results are displayed below in Table 1. All participants were grouped initially for primary analysis. The test-retest coefficients for the in-person and remote formats were greater than .80 indicating strong test-retest reliability.

Mean raw scores and standard deviations for in-person and remote standard scores of the *Articulation and Phonology VAT* are provided in Table 7. The variance in means across groups is composed of the expected range of performance for typically developing participants (ranging from 5 years, 0 months, to 13 years, 0 months) with the expected range of performance for those with an articulation/phonological disorder (ranging from 5 years, 0 months, to 13 years, 0 months). To calculate the effect size, the difference between the mean standard scores of the two testing instances was divided by the pooled standard deviation. An effect size range from 0.01 to 0.09 was realized for the entire sample. An effect size of 0.2 is considered small, 0.5 is considered medium, and 0.8 is considered large (Cohen, 1992). As such, the observed effect sizes were considered small meaning there is insignificant change between the two test conditions (i.e., in-person and remote). Additionally, there were no statistically significant differences found between in-person and remote administrations for the Articulation and Phonology VAT.

In order to investigate the reliability of the examiner's ratings, raw scores from in-person and remote testing were compared for each participant. To calculate rater reliability, the intraclass correlation coefficient was used, following the method outlined by Shrout and Fleiss (1979). The intraclass correlation coefficients were .99 for the Articulation and Phonology VAT indicating a very high level of agreement across the test administration conditions (i.e., in-person and remote) for the same participant.

DISCUSSION

The purpose of this study was to determine if administering the *Articulation and Phonology VAT* remotely would result in the same findings as if it was administered in-person. Eighty-eight students participated in this study and each participant was assessed with the *Articulation and Phonology VAT* remotely and in-person. There was an average three-week gap between each test session. Additionally, test order was counterbalanced so that some students received the remote administration first and some received the in-person administration first. Each student's remote and in-person assessment results were compared, and there were no significant differences found between the two formats of assessment. Additionally, remote and in-person assessment resulted in strong reliability of raw and standard scores.

The results of this study demonstrate that in addition to successful in-person administration, the *Articulation and Phonology VAT* can also be successfully administered remotely via a secure online platform such as Zoom. Remote assessment does not appear to impact an individual's speech sound performance or the examiner's ability to adequately rate an individual's speech production. Additionally, the results of the present study provide evidence that assessment tools can be successfully adapted for remote use and continue to yield valid and reliable results.

In the future, studies can continue to investigate the use of in-person assessment tools adapted for remote administration. Additionally, larger sample sizes with more diverse clinical populations should be used to determine the equivalency of normative assessments via remote administration. In doing so, the findings of future studies can establish whether remote administration of assessments is appropriate. Future studies should also investigate the use of other virtual online platforms and examine if there are any extraneous factors that may impact remote vs. in-person assessment administration. By continuing to investigate the use of remote assessments, clinicians can feel more confident using remote assessments and also guide researchers and test developers in the future.

Table 6

Demographics of the Equivalency Sample

Sample Size = 88

Demographic	*N* Normative Sample	% Normative Sample
Gender		
Male	49	56%
Female	39	44%
Total	88	100%
Race		
White	27	31%
Black	9	10%
Asian	7	8%
Hispanic	39	44%
Other	6	7%
Total	88	100%
Clinical Groups		
	44	50%

Table 7

In-Person vs. Remote Administration Equivalency of Standard Scores, Correlations and Effect Sizes

		In-Person		Remote			
	N	Mean	SD	Mean	SD	*r*	Effect Size
All participants	88	7	2.3	7	2.7	.98	0.01
Typically Developing	44	4	1.6	4	1.4	.95	0.06
Articulation Impairment	44	12	2.7	11	3.1	.91	0.09

References:

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). Standards for educational and psychological testing. Washington, DC: AERA.

American Speech-Language-Hearing Association. (2016). *Speech sound disorders: Articulation and phonology* [Practice Portal]. Retrieved from https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology/

Aram, D. M. & Nation, J. E. (1980). Preschool language disorders and subsequent language and academic difficulties. *Journal of Communication Disorders, 13,* 159–170.

Bauman-Waengler, J. (2004). *Articulatory and phonological impairments: A clinical focus. 2.* Needham Heights, MA: Allyn and Bacon.

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research, 38,* 446–462.

Black, L. I., Vahratian, A., & Hoffman, H. J. (2015). *Communication disorders and use of intervention services among children aged 3–17 years; United States, 2012* (NHS Data Brief No. 205). Hyattsville, MD: National Center for Health Statistics.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2006). *Diagnostic Evaluation of Articulation and Phonology* (U.S. ed.). San Antonio, TX: Pearson.

Dollaghan, C. A. (2007). The Handbook for Evidence-Based Practice in Communication Disorders. Baltimore, MD: MD Brookes.

Eadie, P., Morgan, A., Ukoumunne, O. C., Eecen, K. T., Wake, M., & Reilly, S. (2015). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine & Child Neurology, 57,* 578–584.

Fudala, J. B. & Stegall, S. (2017). *Arizona Articulation and Phonology Scale - Fourth Edition (Arizona-4).* Torrance, CA: Western Psychological Services.

Hall, P. K. & Tomblin, J. B. (1978). A follow-up study of children with articulation and language disorders. *Journal of Speech and Hearing Disorders, 43,* 227–241.

Individuals with Disabilities Education Act. (2004). Section 300.8 child with a disability. Retrieved from https://sites.ed.gov/idea/regs/b/a/300.8

King, R. R., Jones, C., & Lasky, E. (1982). In retrospect: A fifteen-year follow-up report of speech-language-disordered children. *Language, Speech, and Hearing Services in Schools, 13,* 24–32.

Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004) The development of early literacy skills among children with speech difficulties: A test of the "critical age hypothesis." *Journal of Speech, Language, and Hearing Research, 47*, 377–391.

Osborne, J. W. (Ed.). (2008). Best practices in quantitative methods. Los Angeles: Sage Publications.

Overby, M.S., Trainin, G., Smit, A. B., Bernthal, J. E., & Nelson, R. (2012). Preliteracy speech sound production skill and later literacy outcomes: A study using the Templin Archive. *Language, Speech, and Hearing Services in Schools, 43,* 97–115.

Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research, 52,* 1175-1188.

Plante, E. and Vance, R. (1994) Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools 25,* 15 24.

Roulstone, S., Miller, L. L., Wren, Y., & Peters, T. J. (2009). The natural history of speech impairment of 8-year-old children in the Avon Longitudinal Study of Parents and Children: Error rates at 2 and 5 years. *International Journal of Speech-Language Pathology, 11*(5), 381–391.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1461–1481.

Spaulding, T.J., Plante, E., & Farinella, K.A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37*(1), 61-72.

Zumbo, B. D. & Chan, E. K. (2014). *Validity and Validation in Social, Behavioral, and Health Sciences*. Springer.

Appendix A

Summary of Test Words and Sounds Assessed, ages 2:9-5:11

Target Sound in Words	Initial Position of Words	Medial Position of Words	Final Position of Words
/b/ boat, basketball, baby, butterfly, bath	**boat** **basketball** **baby** **butterfly** **bath**	**baby** **robot**	
/p/ peach, pajamas, octopus, ship	**peach** **pajamas**	**octopus** **zipper**	**ship**
/m/ mouse, muffin, mouth, swimming, tomato, drum	**mouse** **muffin** **mouth**	**swimming** **tomato**	**drum**
/n/ nose, plane, chicken, submarine, sun, lion, muffin	**nose**		**plane** **chicken** **submarine** **sun** **lion** **muffin**
/ŋ/ swimming, walking			**swimming** **walking**
/w/ water, walking	**water** **walking**		
/h/ head	**head**		
/d/ duck, dog, spider, head	**duck** **dog**	**spider**	**head**
/t/ tomato, water, butterfly, cat, goat, boat	**tomato**	**water** **butterfly**	**cat** **boat** **goat** **robot**
/g/ goat, yogurt, dog	**goat**	**yogurt**	**dog**
/k/ cat, chicken, duck, quack, book	**cat**	**chicken**	**duck** **quack** **book**
/f/ fish, farm, muffin, giraffe, leaf	**farm** **fish**	**muffin**	**giraffe** **leaf**

/s/ submarine, sunglasses, sun, juice, mouse, octopus	**submarine** **sunglasses** **sun**	**sunglasses**	**juice** **mouse** **octopus**
/z/ zoo, sunglasses, nose, zipper	**zoo** **zipper**		**sunglasses** **nose**
/l/ lion, leaf, yellow, basketball, snail, school	**lion** **leaf**	**yellow**	**basketball** **snail** **school**
/ʃ/ ship, fish	**ship**		**fish**
/ʤ/ juice, giraffe, pajamas	**juice** **giraffe**	**pajamas**	
/tʃ/ chicken, peach	**chicken**		**peach**
/v/ volcano	**volcano**		
/j/ yogurt, yellow	**yogurt** **yellow**		
/θ/ mouth**,** bath			**bath** **mouth**
/r/ robot**,** submarine	**robot**	**submarine**	
/ar/ star			**star**
/ə/ spider, water			**spider** **water** **zipper**

Target Cluster in Words	Initial Position of Words	Medial Position of Words	Final Position of Words
/kw/ quack	**quack**		
/pl/ plane	**plane**		
/bl/ blue	**blue**		
/st/ star	**star**		
/sp/ spider	**spider**		
/sw/ swimming	**swimming**		
/sn/ snail	**snail**		
/sk/ school	**school**		
/dr/ drum	**drum**		

Summary of Test Words and Sounds Assessed, ages 6:0-21:0

Target Sound in Words	Initial Position of Words	Medial Position of Words	Final Position of Words
/b/ bathtub, buzzing, bed, robot, cabbage, strawberry	**bathtub** **buzzing** **bed**	**robot** **cabbage** **strawberry**	**bathtub**
/p/ potato, sleeping, caterpillar, zipper, ship, sheep	**potato**	**sleeping** **caterpillar** **zipper**	**ship** **sheep**
/m/ muffin, mouth, emoji, watermelon, drum, thumb	**muffin** **mouth**	**emoji** **watermelon**	**drum** **thumb**
/n/ nose, chicken, sun, lion, muffin, watermelon	**nose**		**chicken** **sun** **lion** **muffin** **watermelon**
/ŋ/ buzzing, sleeping, crying, playing, ring			**buzzing** **sleeping** **crying** **playing** **ring**
/w/ watermelon	**watermelon**		
/h/ house	**house**		
/d/ duck, dog, spider, bed	**duck** **dog**	**spider**	**bed**
/t/ teacher, tissue, tiger, potato, alligator, caterpillar, watermelon, robot, goat, chocolate	**teacher** **tissue** **tiger**	**potato** **alligator** **caterpillar** **watermelon**	**robot** **goat** **chocolate**
/g/ goat, tiger, alligator, yogurt, dog	**goat**	**tiger** **alligator** **yogurt**	**dog**
/k/ kangaroo, couch, cabbage, cucumber, caterpillar, chicken, chocolate, duck, clock	**kangaroo** **couch** **cabbage** **cucumber** **caterpillar**	**chicken** **cucumber** **chocolate**	**duck** **clock**
/f/ fish, elephant, muffin, giraffe	**fish**	**elephant**	**giraffe**
/s/ sun, glasses, house	**sun**	**glasses**	**house**

/z/ zebra, buzzing, glasses, nose, leaves	**zebra** **zipper**	**buzzing**	**leaves** **nose** **glasses**
/l/ leaves, lion, elephant, chocolate, alligator, caterpillar, watermelon, snail	**leaves** **lion**	**elephant** **chocolate** **alligator** **caterpillar** **watermelon**	**snail**
/ʃ/ ship, sheep, tissue, fish	**ship** **sheep**	**tissue**	**fish**
/ʤ/ giraffe, jumping, emoji, cabbage	**giraffe** **jumping**	**emoji**	**cabbage**
/tʃ/ chicken, chocolate, teacher, couch	**chicken** **chocolate**	**teacher**	**couch**
/v/ volcano, leaves	**volcano**	**leaves**	
/j/ yogurt	**yogurt**		
/θ/ thumb, bathtub, mouth	**thumb**	**bathtub**	**mouth**
/ð/ that, breathe	**that**		**breathe**
/r/ robot, ring, giraffe, kangaroo, strawberry	**robot** **ring**	**giraffe** **kangaroo** **strawberry**	
/ar/ star			**star**
Target Cluster in Words	Initial Position of Words	Medial Position of Words	Final Position of Words
/kl/ clock	**clock**		
/pl/ playing	**playing**		
/gl/ glasses	**glasses**		
/fl/ fly	**fly**		
/nt/ elephant			**elephant**
/nk/ skunk			**skunk**
/mp/		**jumping**	

jumping			
/st/ star	**star**		
/sp/ spider	**spider**		
/sl/ sleeping	**sleeping**		
/sn/ snail	**snail**		
/sk/ skunk	**skunk**		
/kr/ crying	**crying**		
/tr/ tree	**tree**		
/dr/ drum	**drum**		
/br/ breathe	**breathe**	**zebra**	
/str/ strawberry	**strawberry**		