

Effectiveness of Remote Virtual Assessment: The Language 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 oral language and language comprehension performance results of in-person versus remote administrations of the *Language 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 *Language 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 language 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 *Language 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

One hundred and six children, aged 5 years, 0 months, to 14 years, 0 months participated in this study. The sample consisted of forty-nine who were considered typically developing and fifty-seven with a previously diagnosed developmental language delay. Demographic characteristics are reviewed in Table 1. 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 one hundred and six 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 *Language VAT* is composed of short pre-recorded video segments. Therefore, clinicians used an electronic device during both in-person and remote administrations to access the video-based *Language Video Assessment Tool*.

During remote assessment, the examiner used the screen-sharing feature on Zoom to present and administer the *Language VAT*. After displaying a test item to the student, the examiner paused the test, stopped screen-share, and asked the test item questions per test instructions. The clinician would then listen carefully to the answers. Then, the examiner would start screen-share again and move on to the next item and continue the process until all of the *Language VAT* items were administered.

During each participant's first assessment, he/she was fully assessed using the *Language 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 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 clinical 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 *Language 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 *Language VAT* would reveal the same score and/or diagnosis as the in-person administration. The *Language VAT* was administered twice to one hundred and six participants, aged 5 years, 0 months, to 14 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 *Language VAT* are provided in Table 2. 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 14 years, 0 months) with the expected range of performance for those with a developmental language delay (ranging from 5 years, 0 months, to 14 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 0.02 to 0.16 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 *Language 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 .97 for the *Language 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 *Language VAT* remotely would result in the same findings as if it was administered in-person. One hundred and six children students participated in this study and each participant was assessed with the *Language 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 *Language 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 language comprehension and spoken language performance or the examiner's ability to adequately rate an individual's language comprehension and spoken language 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 1		
Demographics of the Equivalency Sample		
Sample Size = 106		
Demographic	<i>N</i> Normative Sample	% Normative Sample
Gender		
Male	60	57%
Female	46	43%
Total	106	100%
Race		
White	41	39%
Black	19	18%
Asian	9	8%
Hispanic	32	30%
Other	5	5%
Total	106	100%
Clinical Groups		
	57	54%

Table 2
In-Person vs. Remote Administration Equivalency of Standard Scores, Correlations and Effect Sizes

	N	In-Person		Remote		<i>r</i>	Effect Size
		Mean	SD	Mean	SD		
Typically Developing							
<i>Following Directions</i>	49	56	0.8	55	0.5	.98	0.02
<i>Restating Information</i>		55	3.4	53	3.1	.91	0.04
<i>Language Comprehension</i>	95	95	3.1	96	3.6	.96	0.10
<i>Morphology and Sentence Structure</i>	64	64	1.2	64	0.9	.93	0.07
Language Impairment							
<i>Following Directions</i>	57	39	2.1	38	3.0	.98	0.03
<i>Restating Information</i>		29	3.7	27	2.5	.93	0.07
<i>Language Comprehension</i>	54	54	3.5	53	3.1	.97	0.09
<i>Morphology and Sentence Structure</i>	39	39	2.8	37	2.4	.93	0.16

References

- American Speech-Language-Hearing Association (ASHA). (2020). *Service delivery considerations in health care during coronavirus/COVID-19* [internet] <https://www.asha.org/SLP/healthcare/SLP-Service-Delivery-Considerations-in-Health-Care-During-Coronavirus/>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Gabel, R., Grogan-Johnson, S., Alvares, R., Bechstein, L., & Taylor, J. (2013). A field study of telepractice for school intervention using the ASHA NOMS K–12 database. *Communication Disorders Quarterly*, 35(1), 44–53.
- Kaiser A.P. & Roberts M.Y. (2011) Advances in early communication and language intervention. *Journal of Early Intervention*, 33, 298–309.
- Keck C. S. & Doarn C. R. (2014). Telehealth technology applications in speech-language pathology. *Telemedicine and e-Health*, 20, 653–659.
- Nicholas J. G. & Geers A. E. (2006) Effects of early auditory experience on the spoken language of deaf children at 3 years of age. *Ear Hear*, 27, 286.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.
- Taylor O. D., Armfield N. R., Dodrill P., & Smith A. C. (2014). A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *Journal of Telemedicine and Telecare*, 20, 405–412.
- Theodoros, D. G. (2008). Telerehabilitation for service delivery in speech-language pathology. *Journal of Telemedicine and Telecare*, 14, 221–224.
- Wales D., Skinner L., & Hayman M. (2017). The efficacy of telehealth-delivered speech and language intervention for primary school-age children: a systematic review. *International Journal of Telerehabilitation*, 9-55.