Spring 5-2019

Reliability of Laryngeal Diadochokinesis Measures

Megan K. Disher

RELIABILITY OF LARYNGEAL DIADOCHOKINESIS MEASURES

A Thesis

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Master of Science

Megan Katherine Disher

Indiana University of Pennsylvania

May 2019

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Communication Disorders, Special Education, and Disability Services

We hereby approve the thesis of

Megan Katherine Disher

Candidate for the degree of Master of Science

3-25-2019 — Signature on file
Lori E. Lombard, Ph.D., CCC-SLP
Professor of Speech-Language Pathology,
Advisor

3-25-2019 — Signature on file
Lisa Hammett Price, Ph.D., CCC-SLP
Professor of Speech-Language Pathology

3-25-2019 — Signature on file
Kendra A. McAlear, CScD, CCC-SLP
Assistant Professor of Speech-Language
Pathology

ACCEPTED

Signature on file
Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: Reliability of Laryngeal Diadochokinesis Measures

Author: Megan Katherine Disher

Thesis Chair: Dr. Lori E. Lombard

Thesis Committee Members: Dr. Lisa Hammett Price
Dr. Kendra A. McAlear

Laryngeal diadochokinesis (LDDK) is an assessment of laryngeal motor function that provides information regarding neuromotor maturation and the integration of the structures needed for phonation (Modolo, Berretin-Felix, Genaro, & Brasolotto, 2011). Laryngeal diadochokinesis has clinical relevance because it offers a valid measurement of neuromotor function and vocal fold integrity. However, current research does not provide evidence regarding the degree to which speech-language pathologists can be reliable in their LDDK measurements in clinical settings. In order for LDDK to be clinically useful, however, there must be evidence that it can be a reliable measure. This study assessed the reliability of LDDK when graduate student clinicians use the pencil dotting method of data collection. Reliability was tested using both abductory and adductory stimuli to determine if a difference in interrater reliability among the two stimuli was present.

ACKNOWLEDGMENTS

Thank you to the following individuals for making this possible:

Lori Lombard

Lisa Price

Kendra McAlear

Jill Brady

Emilie Okopal

Sophia Thompson

Ryan Gross

Jessica Potash

IUP M.S. Speech-Language Pathology Class of 2019

IUP Applied Research Lab

TABLE OF CONTENTS

Chapter Page

I REVIEW OF THE LITERATURE 1

Laryngeal Diadochokinesis 1
Clinical Indications- Neuromotor Function of the Larynx 4
Clinical Assessment of Neuromotor Function 6
Benefits of Laryngeal Diadochokinesis 10
Measurement of Laryngeal Diadochokinesis 11
Lack of a Standardized Protocol 12
Oral Diadochokinesis 14
Reliability of Diadochokinesis Measurement 16
Conclusion 24

II RATIONALE FOR STUDY 26

III METHODS 27

IV DATA ANALYSIS 31

V DISCUSSION 37

Implications 38
Limitations 39

REFERENCES 41

LIST OF TABLES

Table Page

1 Audio File Demographics 29

2 Intraclass Correlation Qualities....................31

3 Intraclass Correlation for /ʌ/ Audio Files 32

4 Intraclass Correlation for /hʌ/ Audio Files 32

5 Calculated Rates (Repetitions/ sec) for /ʌ/ Audio Files 33

6 Calculated Rates (Repetitions/ sec) for /hʌ/ Audio Files 34

7 Agreement With LDDK Rate Averages 36

LIST OF FIGURES

Figure Page

1 Normative data for LDDK rate 35

CHAPTER I

REVIEW OF THE LITERATURE

Laryngeal Diadochokinesis

Diadochokinesis is defined as the "function of arresting one motor impulse and substituting one that is diametrically opposite" (Dorland's Medical Dictionary as cited in Shanks, 1966, p. 4). Laryngeal diadochokinesis (LDDK), also known as vocal fold diadochokinesis, is an assessment of laryngeal motor function (Modolo et al., 2011; Shanks, 1966). LDDK performance provides information regarding neuromotor maturation and the integration of the structures needed for phonation (Modolo et al., 2011). Clinicians use LDDK to analyze neuromotor control during the fast opening and closing of the vocal folds (Modolo et al., 2011).

The quick opening and closing of the vocal folds needed for LDDK is triggered by the initiation of quick repetitions of vowels (Modolo et al., 2011). However, the procedure used to elicit LDDK is debated among researchers. Stimuli for LDDK can consist of repetition of syllables either with or without a glottal consonant. Canter, recommended using a syllable with a glottal consonant such as /hʌ/ because repetition of a single vowel might create "pulses of air pressure acting on a fixed laryngeal valve," thereby not giving a valid measurement (as cited in Shanks, 1966, p. 7). Gratzmiller (2012) compared the differences between LDDK performance using /hʌ/ and /ʌ/ stimuli. She found that there was not a statistically significant difference (p=0.512) between performance on the two stimuli among 35 normal participants between the ages of 40 and 60 years (Gratzmiller, 2012). While this finding is important to consider when choosing stimuli, it is only representative of LDDK completed with normal individuals across a 20-year age span.

Understanding extraneous factors on LDDK performance helps clinicians and researchers make informed decisions regarding which LDDK stimuli to use. Shanks (1966) used the syllable /hʌ/ to investigate the effects of auditory feedback, somesthetic feedback, variations in pitch and intensity, and aging on LDDK. Shanks used four groups in her study. Three of the four groups consisted of 40 participants each and all were female. They were divided into three age groups: young adult (20-40 years old), early senescence (40-60 years old), and later senescence (60-80 years old). The last group consisted of 10 people (8 males and 2 females between the ages of 16 and 39) in whom the effects of the disruption of somesthetic feedback on LDDK was studied. These 10 participants had nose or ear pathology and were being seen at the Otorhinolaryngology clinic at Charity Hospital. Results from the young adult group were used to study the effects of auditory feedback and variations in pitch and intensity, in addition to being compared to the early senescence and late senescence group to examine the effects of aging. Subjects were screened and eliminated on several criteria including known neurological pathology and a history of respiratory disease (Shanks, 1966).

Shanks (1966) ensured that the investigator's judgements of LDDK were valid in two ways. The first consisted of 10 preliminary samples, including five deviant samples and five normal samples, to determine that the investigator could analyze them accurately. The investigator's assessments were also compared to assessments completed by five speech-language pathologists who were all considered to have "extensive experience" with patients with voice disorders. Differences between the investigator and the five speech-language pathologists were not considered statistically significant (chi-square of 1.62) indicating that the investigator judged the samples reliably in comparison to the voice experts. Shanks also assessed intra-judge reliability by reanalyzing some samples, which were randomized, three weeks after the initial analysis. Again, the differences between the initial analysis and the second were found to be statistically insignificant (chi-square of 1.62). These findings provide evidence of strong intra- and inter-judge reliability, for the measures of LDDK in the study (Shanks, 1966). They provide evidence that with training, LDDK judgements can be reliable. In the context of her study, they confirm that we can trust the data used to explore the effects of other variables on LDDK performance. One limitation to this evidence on reliability is that there was only one investigator who tested intra-rater reliability. Future research needs to determine whether intra-rater reliability can be high across multiple raters.

All participants in the Shanks (1966) study were asked to repeat /hʌ/ as quickly as they could, but some participants had additional directions based on what was being assessed. To assess the effect of auditory feedback on LDDK performance, participants first repeated /hʌ/ at a comfortable pitch and intensity. Afterwards, participants were again asked to complete the task at a comfortable pitch while they experienced white noise at 100dB SPL, but at a self-monitored intensity level between 72 and 74 dB SPL. Lastly, the task was repeated with no masking at a comfortable pitch, but with a controlled intensity level between 72 and 74 dB SPL to determine if controlling intensity had an independent impact on LDDK on its own.

To measure the effects of somesthetic feedback on LDDK, participants performed the task under laryngeal anesthetization at a comfortable pitch and intensity level. To measure the effect of variations in pitch and intensity on LDDK, participants were initially asked to perform that task at a comfortable pitch and intensity. Afterwards, participants performed the task at a lower intensity level at a comfortable pitch, higher intensity at a comfortable pitch, lower intensity with a higher pitch, comfortable intensity with a higher pitch, higher intensity with a higher pitch, lower intensity with a lower pitch, comfortable intensity with a lower pitch, and higher intensity with a lower pitch. Lastly, to measure the effects of aging on LDDK, participants were asked to do the task at a comfortable pitch and intensity level and then performance was compared across the age groups (Shanks, 1966).

Shanks (1966) analyzed all productions of LDDK and found that a disruption in auditory feedback caused a significant reduction in the rate of LDDK performance but no significant difference in the percentage of abducted syllables or phonation time. Participants performing the LDDK task at a designated intensity level (a comfortable intensity level, decreased intensity level from comfortable, and increased intensity level from comfortable) did not differ significantly in rate. In addition, Shanks found that disruption of laryngeal somesthetic feedback did not affect the rate, periodicity, percentage of abducted syllables, or phonation time of the LDDK task. Finally, there were no statistically significant differences across the three age groups on LDDK tasks; therefore, Shanks concluded that aging did not have an effect on LDDK performance.

Clinical Indications—Neuromotor Function of the Larynx

Because LDDK helps clinicians assess the innervation and neuromotor function of the larynx, it is important for clinicians and researchers to have a thorough understanding of neurophysiology. The larynx is innervated by the Vagus nerve (cranial nerve X), which branches into the superior laryngeal nerve, and the recurrent laryngeal nerve. The superior laryngeal nerve enters the larynx through the thyrohyoidmembranes, which connect the hyoid bone and thyroid cartilage. The superior laryngeal nerve has an internal branch and an external branch. The internal branch primarily functions as a sensory nerve, handling the sensory information in the larynx above the vocal folds. In addition, the internal branch also holds some parasympathetic secretomotor fibers which supply secretory glands above the vocal folds. The external branch of the superior laryngeal nerve enters the larynx specifically in the cricothyroid muscle and supplies motor commands to that muscle (Garret & Larson, 1991).

Normal recurrent laryngeal nerve function is a requirement for successful completion of the LDDK task, but its location within the body can leave it susceptible to damage. The recurrent laryngeal branch of the Vagus nerve branches into a left side and right side in the thorax, unlike the superior laryngeal nerve which has a direct route to the larynx. The left side of RLN starts from the Vagus nerve superficial to the aortic arch and travels up the trachea to the larynx. The right side starts from the Vagus superficial to the subclavian artery and then travels up the trachea to the larynx. Both the left and right side of the recurrent laryngeal nerve enter the larynx through the cricothyroid membrane, which connects the cricoid and thyroid cartilages. Since the recurrent laryngeal nerve travels throughout the thoracic cavity, it is more susceptible to damage during thoracic surgeries. The recurrent laryngeal nerve has both sensory and motor components. It controls the motor functions for all the intrinsic muscles of the larynx except for the cricothyroid. The recurrent laryngeal nerve also controls sensory innervation below the vocal folds (Garret & Larson, 1991). Voiced phonemes such as the vowel used in both

LDDK stimuli require closure of the vocal folds. For this closure to occur, both the neuromuscular control of the larynx and the vocal fold histology itself must be intact.

The muscle innervation of the larynx differs in comparison to other muscles throughout the body. For most muscles, the endplates of the motor neuron, where synaptic contact with the muscle is established, are located near the muscle belly. Four of the intrinsic laryngeal muscles (posterior cricoarytenoid, thyroarytenoid, interarytenoid, and cricothyroid) have motor neuron endplates that are located in the muscle fiber itself. This allows for distribution of the endplates to be relatively broad with no specific organizational pattern. Furthermore, the thyroarytenoid and lateral cricoarytenoid muscles have a faster contraction time than the posterior cricotarytenoid and the cricothyroid. These are the muscles that are important for glottal closure which allows for phonation and protects the airway during a swallow (Garret & Larson, 1991). In order for LDDK to be completed with average results, the vocal folds must be intact for phonation and muscular innervation and movement must be intact. Using one assessment, LDDK has the ability to detect neurological changes in respect to laryngeal function unlike similar assessments.

Clinical Assessment of Neuromotor Function

There are several measures other than LDDK that can be used to measure the neuromotor function of the larynx. These measures include endoscopy, electromyography, electroglottography, and several acoustic measures. Endoscopy involves using a magnifying camera lens that allows the clinician to see the vocal folds and make visual judgements regarding the vibratory pattern, movement, and appearance of the vocal folds. There are two types of endoscopy, rigid and flexible. Rigid endoscopy consists of larger fiberoptic bundles which allow for better lighting and a more magnified view of the larynx than the flexible endoscope. However, because the rigid endoscope must go through the mouth, the only available speech sample during this test is a sustained vowel. This speech sample is limited even further because it is gathered in an unnatural speaking posture, since it requires the tongue to be out and the neck to be extended. Rigid endoscopy may also require a topical anesthesia if the client has a sensitive gag reflex (Stemple, Roy, & Klaben, 2014). These issues create barriers to using rigid endoscopy for the purposes of assessing laryngeal function.

Unlike rigid endoscopy, which is inserted through the mouth, flexible endoscopy is inserted through the nasal cavity and hangs over the velum in the pharynx. Since flexible endoscopy must pass through the nasal cavity, it has a smaller fiberoptic bundle which yields a darker image than rigid endoscopy. However, passing through the nasal cavity instead of the mouth allows for more natural, connected speech tasks than the rigid endoscopy. However, flexible endoscopy requires a lubricant and topical anesthesia which must be administered by a physician. Therefore, flexible endoscopy cannot be used in an SLPs office without physician involvement, thereby creating another barrier to its use as a clinical measure of laryngeal function.

Endoscopy allows for visual-perceptual assessment that rules out lesions on the vocal folds, edema, and vocal fold cancer and can be used to evaluate vocal fold vibratory patterns (Stager & Bielamowicz, 2010; Verdolini & Palmer, 1997). In addition, clinicians can observe characteristics of spasmodic dysphonia, vocal fold paresis, and presbylaryngeus. However, endoscopy alone is not enough to provide differential diagnoses. For example, vocal fold paresis and presbylaryngeus both cause an incomplete glottal closure and look similar during endoscopy (Stager & Bielamowicz, 2010). While endoscopy provides a visualization of the movement of the structures needed for phonation, it does not provide any direct information regarding innervation. In addition, there are practical limitations to its use by SLPs.

Electromyography (EMG) measures muscle activity through needle electrodes inserted into the laryngeal muscles by a neurologist or otolaryngologist. This invasive procedure looks at electrical activity in the muscle to determine innervation. Because the laryngeal muscles cannot be seen externally, repositioning of the needle electrodes may be necessary during the evaluation to ensure correct placement (Stemple et al., 2014). Electromyography is a tool used in differential diagnoses, for example, helping to differentiate vocal fold paresis from presbylaryngeus (Stager & Bielamowicz, 2010). Electromyography can be used as a quantitative measure of muscular innervation to the larynx, as an adjunct to the qualitative measure of LDDK. Electromyography has been used to compare different LDDK protocols and determined that the best stimulus is a vowel that is paired with a glottal fricative, such as the /hʌ/ that has been used in several of the previous mentioned studies (Shanks, 1966).

Electroglottography (EGG) uses two electrodes placed on either side of the neck to measure vocal fold contact. Vibration of the vocal folds generates an electrical current in the electrodes creating a waveform display of the vocal fold vibratory pattern. Variations from person to person, such as mucous interference, electrode placement, and tissue thickness can cause errors. (Stemple et al., 2014).

Measuring the acoustic properties of the voice is another technique to assess laryngeal function. The five most common acoustic measurements used are fundamental frequency, intensity, perturbation, harmonic-to-noise ratio, and spectral analysis.

Fundamental frequency, perceived as pitch, is the rate of vocal fold vibration. Measuring the range of fundamental frequencies allows the clinician to make judgements regarding vocal fold flexibility or screen for other voice pathology such as vocal nodules. Fundamental frequency is calculated using either a sustained vowel or from a connected speech sample. Intensity, perceived as loudness, is measured in decibels. Intensity ranges and habitual intensity allow clinicians to make judgements regarding vocal fold adduction (Stemple et al., 2014). Perturbation is "the cycle-to-cycle variability in a signal," and is measured using jitter and shimmer (Stemple et al., 2014, p. 159). Perturbation, unlike fundamental frequency and intensity, is not directly correlated to one specific perception, and normal limits for perturbation are debated. Harmonic-to-noise ratio measures "the ratio of periodic or harmonic signal energy to the aperiodic or noise energy in the voice waveform" (Stemple et al., 2014, p. 161). Harmonic-to-noise ratio is used by clinicians to determine normal from dysphonic voices. Clients with a voice that has a low harmonic-to-noise ratio have high aperiodicity characterized by roughness or breathiness. Lastly, spectral analysis consists of analyzing a spectrogram, a graphic representation of the voice. Spectrograms plot intensity and frequency over time. There are two different types of spectrograms: narrowband and line spectrum. Narrowband spectrograms are a graphic representation of harmonic-to-noise ratio. Line spectrums are a graphic representation of formant patterns and show formant pattern changes (Stemple et al., 2014). While acoustic measurements provide some insight into the neuromotor function of the larynx, results are impacted by other anatomic factors such as the shape of the vocal tract and speaking techniques. In addition, acoustic measurements require several completed tasks to make conclusions regarding neuromuscular integrity.

Benefits of Laryngeal Diadochokinesis

Laryngeal diadochokinesis has several benefits over measurements such as EMG and EGG. Unlike other measures of neuromotor function of the larynx, LDDK can be completed quite quickly. Verdolini and Palmer (1997) utilized LDDK instead of other measures because it could be completed "within a few minutes" (Verdolini & Palmer, 1997, p. 218). Even when the clinician needs to spend a few minutes training the client on the LDDK task, it is faster than the preparation involved with endoscopy, more specifically when using anesthesia is necessary. Likewise, LDDK is faster than EMG, which involves invasive setup and electrode placement similar to that of EGG. Lastly, while acoustic measures can be collected as quickly as LDDK, acoustic analysis requires several different measurements to paint the picture that LDDK does on its own.

In addition to the efficiency of LDDK, Verdolini and Palmer (1997) chose LDDK for their study because it could be completed without the use of expensive equipment. Endoscopy requires the use of endoscopic equipment, anesthesia, and sterilization supplies which cost more than both the pen and paper dotting method of data collection and the waveform analysis method. Electromyography and EGG also require expensive equipment and a trained medical professional to complete the procedure. Instrumentation for acoustic analysis is negligible in that both LDDK and acoustics can use the same software for waveform analysis.

One of several reasons Verdolni and Palmer (1997) chose LDDK for their study is that it could be used in "most voice screening situations" including public places such as in malls and community centers (Verdolini & Palmer, 1997, p. 218). Endoscopy utilizes bulky equipment and specific supplies, meaning it is a procedure that must be completed in a designated setting such as an otolaryngology suite. Similarly, EMG and EGG require specific equipment and supplies that are not quickly or easily transported. Acoustic measurements need the same equipment as LDDK if you are utilizing the waveform analysis technique. However, the ability to use the pencil dotting method makes LDDK more mobile than acoustic measures. Using the pencil dotting method, clinicians can give results to the client without having to record and complete a waveform analysis on a computer later, a delay with acoustic measures that leaves the client with no immediate information following testing.

Measurement of Laryngeal Diadochokinesis

Accurate data collection procedures are essential for reliable results. The measurement of LDDK performance in research lacks often differs from what is done in the clinical setting. The most common type of LDDK data collection used by speech-language pathologists involves tallying with a pen and paper for every syllable that the client produces. Tomblin, Morris, and Spriestersbach (2000) provided a protocol in their diagnostic textbook that speech-language pathologists can use. In that protocol, Tomblin et al. recommend that the clinician “dot a pencil on paper for each production of ‘uh’ for the duration of the trial” while timing the task with a stopwatch (Tomblin, Morris, & Spreistersbach, 2000, p. 280). Since the pencil dotting method is used frequently in clinical settings, Verdolini and Palmer (1997) used the pencil dotting method in their study designed to recreate a screening environment. Gadesmann and Miller (2008) followed a similar protocol in their oral DDK study, stating “while research facilities generally employ some method of instrumentation which displays acoustic waveforms to help quantify and describe DDK performance, in the clinic DDK rate is typically measured by counting syllables—either live or from recorded samples—and reading the time with a stopwatch" (Gadesmann & Miller, 2008, p. 43).

Even though pencil dotting is the method that is commonly used in clinical settings, research studies commonly rely on a waveform analysis. Waveforms allows the researcher or clinician to see a physical representation of the LDDK performance in the shape of a waveform. The peaks of the waveform are then counted to determine the number of syllables. The visual representation of LDDK rate enhances reliability over a strictly auditory method.

Lack of a Standardized Protocol

Laryngeal diadochokinesis studies vary in the stimulus used, data collection protocol, and methods of calculating performance. Shanks (1966) reviewed three studies and the protocol that was used in each study. The first study Shanks discussed was Canter's 1961 study where the stimulus involved a rapid repetition of /hʌ/ (as cited in Shanks, 1966). The second study Shanks reported on was by Sander (1963), where /hʌ/ was also used (as cited in Shanks, 1966). The third study Shanks discussed was by Ptacek and Maloney (1963), which used a rapid repetition of /ʌ/ (as cited in Shanks, 1966). This stimulus was also used in a later study by Ptacek et al (Ptacek, Sander, Maloney, & Jackson, 1966). In Shank's study, participants were required to do a rapid repetition of /hʌ/ (Shanks, 1966).

Gratzmiller (2012) and Bassich-Zeren (2004) differed from Canter, from Sander, and from Ptacek and Maloney and utilized both /hʌ/ and /ʌ/ stimuli in their studies (Bassich-Zeren, 2004; Gratzmiller, 2012). Gratzmiller found that there was no statistically significant difference between the two stimuli in healthy adults (Gratzmiller, 2012). Bassich-Zeren found that participants with early onset Parkinson's disease demonstrated a statistically significant difference from a healthy control group on the /ʌ/ stimulus but showed no statistically significant difference on the /hʌ/ stimulus (Bassich-Zeren, 2004). There are additional studies that use LDDK, but do not report which stimuli they utilized. An example of such a study was completed by Rosen et al. where LDDK was used as an outcome measure after vocal fold augmentation (Rosen et al., 2007).

LDDK studies also vary in the method of data collection. The majority of studies reviewed used a recording device and specifically noted the participant's distance from the microphone of the recording device. This distance varied across studies. For example, Bassich-Zeren (2004) reported that the microphone was placed 4cm away from the participant, however Gratzmiller (2012) reported 6 in. and Leeper and Jones (1991) reported 6 cm. Other studies did not mention the distance away from the microphone (Ptacek et al., 1966; Rosen et al., 2007; Shanks, 1966). Present research has not indicated which of these distances were best practice, or if a difference in distance could influence results.

In addition to inconsistencies regarding data collection, LDDK studies reviewed here lacked consistency on how that data is analyzed. Data analysis across LDDK studies mostly utilized a waveform analysis approach but referred to it by several different names. Shanks referred to Ptacek and Maloney's 1963 study where power level tracings were utilized (as cited in Shanks, 1966). In Shank's (1966) study, graphic level tracings were used (Shanks, 1966). Oscillographs or osciollograms were used in Leeper and Jones's (1991) study, Gratzmiller's (2012) study, and Ptacek et al.'s (1966) study. Power level tracings, graphic level tracings, oscillographs, and oscillograms are all types of waveform analysis. Another common type of data analysis involved using spectrograms which were utilized by Shanks in addition to the waveform analysis and by Bassich-Zeren as the primary analysis method (Bassich-Zeren, 2004; Shanks, 1966). The inconsistencies in protocol regarding stimuli selection, data collection, and data analyses, in addition to the lack of research regarding the effects of these protocol differences on study outcomes, indicates a need for more extensive research.

Oral Diadochokinesis

Given the inconsistencies within LDDK methodology, consideration of oral diadochokinesis, a similar task that has a wider research base, provided insight on what additional LDDK research is needed. Oral diadochokinesis (oral DDK) is the "ability to perform rapid repetitions of relatively simple patterns of opposite muscle contractions," (Modolo et al., 2011, p. 1). Oral DDK is similar to LDDK but instead of measuring laryngeal motor function, it assesses the motor function of the lips, tongue, and velum. Research has shown that oral DDK is a valid and sensitive way to find mild neuromuscular impairment (Gadesmann & Miller, 2008). Oral DDK is often used to diagnose disorders such as dysarthria and apraxia of speech; Willams and Stackhouse even state that "poor performance on spoken DDK tasks has become the most common criterion for identifying and selecting participants for studies of developmental verbal dyspraxia" (Williams & Stackhouse, 2000, p. 268).

Oral DDK procedures are similar to LDDK, making oral DDK procedures a possible resource when developing more consistent LDDK procedures. Oral DDK tasks involve the repetition of monosyllables and polysyllables and a sequence of syllables. Similarly to LDDK, oral DDK lacks routine procedures which limits its use, even though it has been shown to provide important information (Fletcher, 1972). Crary stated that when determining which stimuli should be used for oral DDK, two questions need to be answered. They are "Which syllables should be used?" and "How will the responses be scored?" (as cited in Williams & Stackhouse, 2000, p. 269). The most commonly used monosyllables are /pə/, /tə/, and /kə/ and the most common syllable sequence is /pə tə kə/. Since non-meaningful stimuli are used, the performance relies on neuromotor ability itself, with no influence from the client's linguistic ability. However, the abstract nature of these non-meaningful stimuli made the oral DDK task difficult for children to understand. In these cases, real words such as 'patacake' and 'buttercup' have often been used but according to Williams and Stackhouse (2000), there was no research that determined that these substitutions are equal substitutions. The non-meaningful stimuli required a new motor program since it is unfamiliar to the client, but the same is not necessarily true for the substituted words (Williams & Stackhouse, 2000).

When measuring oral DDK, rate was the most commonly used measurement (Williams & Stackhouse, 2000). To measure the rate, there were several different approaches used. One such approach was the time by count measurement, which was considered the traditional measurement of oral DDK. The time by count measurement involved counting the number of syllables produced during a set time. This consisted of a two-step process: establishing time limits and establishing the number of syllables within those time limits. An alternative to this two-step process would be to watch the time and count the syllables at the same time which splits the clinician's attention to the task because the clinician has to make mental note of both the number and repetitions and the duration of the task, which makes the validity of the measurement questionable (Fletcher, 1972).

Fletcher developed and studied another way to calculate rate during oral DDK. This new technique was based on pre-determining the number of syllables that will be counted and turning a stop watch on and off at the start and end of counting, instead of pre-determining the time interval. After studying this method with 20 monosyllable samples, 15 bisyllable samples, and 10 trisyllable samples, Fletcher found that this new measurement method is a valid way of measuring oral DDK (Fletcher, 1972).

Reliability of Diadochokinesis Measurement

There is an overall lack of research regarding the reliability and validity of LDDK measurement which is needed to help clinicians to develop a consistent protocol. Only two of the studies discussed do so: Bassich-Zeren (2004) and Gratzmiller (2012). Bassich-Zeren studied the vocal functions of participants with early onset Parkinson's disease (n=12) through questionnaires, sustained vowel phonation, LDDK, and a reading of a fairy-tale passage using three different speaking situations. Results were then compared to those of a group of healthy controls (n=12). The specific conditions for the LDDK stimuli, data collection, and data analysis and rate calculation were well-defined.

The LDDK stimuli consisted of rapid repetitions of both /ʌ/ and /hʌ/ for 7 seconds. The LDDK stimuli were modeled by the experimenter and then the participants had the opportunity to practice. During practice, the experimenter provided feedback to ensure the task was being done at the participant's true ability level. The LDDK performance, along with other measurements, were recorded using a digital audiotape recorder with a "head-mounted microphone" with a "constant 4 cm speaker mouth-to-microphone distance" (Bassich-Zeren, 2004, p. 123).

The data gathered using the digital audiotape recorder was transferred to a Computerized Speech Lab (CSL Model 4300 by Kay-Elemetrics) where experimenters used “spectrograms with a frequency range of 0-7kHz” to count and calculate the rate for both tasks (Bassich-Zeren, 2004, p. 137). All of the tasks completed were measured independently by two raters. Bassich-Zeren (2004) calculated unit by unit agreement and found 100% agreement on both the /ʌ/ and /hʌ/ stimuli. With this data, Bassich-Zeren found that there was a statistically significant difference between the two groups for the /ʌ/ stimulus ($p<0.05$) but not the /hʌ/ stimulus ($p>0.05$). The early onset Parkinson’s disease participants demonstrated a difference in /ʌ/ stimulus performance when compared to the healthy participant control group but demonstrated no difference on /hʌ/ (Bassich-Zeren, 2004).

Gratzmiller’s study (2012) yielded conflicting results to Bassich-Zeren. Gratzmiller studied the differences in LDDK rates when using /ʌ/ versus /hʌ/ in 35 healthy participants that were between the ages of 40 and 60 years old. The data were recorded using a Roland CD recorder with the participants sitting 6 inches away from the microphone. The data were analyzed with an oscillogram created by the Multidimensional Voice Program. With these osciollograms, Gratzmiller calculated the LDDK rate by counting the peaks in the waveform. According to the Gratzmiller study, there is no statistically significant difference ($p=0.512$) between LDDK performance on /ʌ/ versus /hʌ/. If clinicians were to take Gratzmiller’s study without considering Bassich-Zeren’s study (2004), then it would appear that the LDDK stimuli do not matter when making decisions regarding LDDK protocol. However, if a clinician had compared those with early onset Parkinson’s to healthy individuals using just the /hʌ/ stimuli, detection of possible early onset Parkinson’s would have gone unnoticed.

Clinicians turn to other studies that use LDDK as either a predictive or outcome measurement because there is a lack of research specifically regarding the reliability and validity of LDDK. An example of such a study was by Rosen et al. (2007). These researchers used LDDK, maximum phonation time, and s/z ratio (ratio of /s/ and /z/ production as a measure of vocal fold efficiency) as outcome measurements after vocal fold augmentation using calcium hydroxylapatite in an observational-descriptive single group study with a pretest and post-test. Speech-language pathologists completed these quantitative measures pre-augmentation, 1-month post, 3-months post, and 6-months post on 68 patients across 12 different facilities. Rosen et al. did not comment on specifics regarding LDDK performance such as the stimulus used, data collection method, or calculation method. However, the improvement in LDDK performance between each of the testing periods was statistically significant ($p<0.001$).

Instead of using LDDK as an outcome measure, Verdolini and Palmer (1997) studied the validity of LDDK as a diagnostic tool. In an observational-analytic case-control study, they assessed the validity of using LDDK and s/z ratio, as a combined diagnostic protocol, to predict a diagnosis of structural abnormalities (specifically, nodules and vocal process granulomas), dynamic abnormalities (unilateral abducted vocal fold paralysis, Parkinson's disease), or a normal larynx (which includes functional voice disorders). Considering each diagnosis, predictions were made on how the client would perform on the LDDK and s/z ratio tasks. For example, those with nodules were expected to have poor performance on the s/z ratio task but LDDK performance that was within normal limits. It is important to note that performance for vocal process granulomas was predicted to be within normal limits for both the s/z ratio and LDDK because it does not affect glottal closure nor the dynamic movement of the vocal folds (Verdolini & Palmer, 1997).

Verdolini and Palmer (1997) developed a protocol that was consistent with voice screening scenarios. The S/Z ratio and LDDK stimuli were presented and elicited by a speech-language pathologist with 11 years of experience with voice disorders. The LDDK measure was completed using a ‘pencil dotting’ method, tallying each syllable production in real time. This is represented in typical voice screening situations. The data for 12 of the participants (27%) was randomly chosen to be re-evaluated for interjudge, intrajudge, and oscilloscopic reliability. Verdolini and Palmer found that the speech-language pathologist who completed the data collection was reliable when compared to herself, to a doctoral level graduate student, and to oscillographic calculations (reliability between 0.972 and 0.986). Once the s/z ratio and LDDK tasks were completed and calculated, the speech-language pathologist made inferences as to what an endoscopic evaluation would show. An endoscopic evaluation was then completed to confirm or deny the speech-language pathologist’s predictions.

Verdolini and Palmer found that the results of s/z ratio and LDDK performance predicted organic laryngeal pathology in 80% of participants who did indeed have organic laryngeal pathology. Of the 20% of participants with organic pathology that was not detected, 11% included those with granulomas, because their results were within normal limits (as expected). Verdolini and Palmer concluded that while this is sufficient in a screening situation, they would like to see a higher validity for these measures to correctly find normal larynx. The results of this study were that the s/z ratio and LDDK performance found a false positive detection for pathology in 30% of normal participants.

More research using these measures on participants with less-specific diagnoses is needed (Verdolini & Palmer, 1997).

A thorough understanding of common factors that could influence LDDK performance compared to normative data is important for making clinically appropriate judgements. Ptacek, Sander, Maloney, and Jackson (1966) studied changes in performance on a series of tasks in two age groups, under age 40 and over the age of 65, in a cross-sectional study. There were 31 participants in the under-40 female group, 28 participants in the under-40 male group, 36 participants in the over-65 female group, and 27 participants in the over 65-male group. Ptacek et al. analyzed pitch range, oral DDK and LDDK, maximum vowel intensity, maximum vowel duration, maximum intraoral (breath) pressure, and vital capacity. Ptacek et al. used /ʌ/ as the LDDK stimulus in this study and had the participant repeat it for 7 seconds. To gather the data, Ptacek et al. used a microphone as an amplifier and rectifier-integrator. This data was then generated into an oscillogram using a Brush oscillographic recorder. There was a statistically significant difference ($p<0.01$) between the rate on the LDDK task between the males of both age groups and females of both age groups. Therefore, Ptacek et al. argued that comparing the LDDK performance, rate specifically, to published data of younger participants creates the appearance of disorder instead of accounting for the natural aging process (Ptacek et al., 1966). Ptacek et al.'s findings conflicted with those found by Shanks (1966) stating that aging had no effect on LDDK performance. Given that Shanks used only /hʌ/ and that Ptacek et al. only used /ʌ/, it begs the question, do the effects of the natural aging process only effect performance on one stimulus, but not the other?

Another way to interpret information regarding the reliability and validity of LDDK was by examining research on oral DDK. A study by Modolo et al. (2011) examined the effects of age and gender on oral and laryngeal DDK of 150 Brazilian-Portuguese-speaking children, divided by age into 8, 9, and 10-year-old groups which were also divided by gender in a cross-sectional study. Oral DDK stimuli for this study included /pə/, /tə/, /kə/, and /pətəkə/. Laryngeal diadochokinetic stimuli included sustained phonation of /a/ and /i/. Both the LDDK task and oral DDK tasks were completed and recorded for 6 seconds. Once the samples were collected and downloaded into professional audio software (Sound Forge 7.0), Modolo et al. trimmed the sample to 3 seconds by trimming the initial 2 seconds and the final 1 second (Modolo et al., 2011).

Using the trimmed 3 second sample, Modolo et al. (2011) used the Motor Speech Profile Advanced (MSP) software for analysis of the monosyllable stimuli for oral DDK and the LDDK sustained vowels. The MSP software calculated values for average DDK rate per second, number of vocalizations per second (DDK velocity), standard deviation, coefficient of variation of the DDK period, and coefficient of variation of DDK peak intensity. The trisyllabic measurement of oral DDK was analyzed using the Multi Speech Main Program software and the number of sequences per second was calculated (Modolo et al., 2011).

After analysis of the samples, Modolo et. al. found that the older children (10-year-old group and 9-year-old group when compared to the 8 year olds) had a higher rate for the monosyllabic stimuli and a decrease in the amount of time needed to produce the trisyllabic sequence. When analyzing differences in genders, the rate of repetition of /tə/ was higher in females than in male students. However, males produced a higher number of utterances as compared to females on the trisyllabic stimuli. Modolo et al. found similar results regarding age and gender for the LDDK stimulus /i/ but no differences in gender and age for the LDDK stimulus /a/ (Modolo et al., 2011).

Gadesmann and Miller's 2008 study took a direct look at the reliability of oral DDK using inter and intra-reliability. A tally system was used rather than waveform analysis since tallying the number of productions is what is used in most clinical settings. Gadesmann and Miller used two groups of raters in their study. The first group included ten experienced speech-language pathologists with mean working experience of 9.2 years. The second group consisted of 10 participants "with no connection to or experience of speech and language therapy" (Gadesmann & Miller, 2008, p. 40).

Gadesmann and Miller used speech samples from Darly et al.'s study for the raters to analyze. Each participant analyzed 12 audio-clip samples that were chosen to represent a variety of disorders and severities, ages, and genders. With these 12 samples, the raters were asked to complete 6 analyses which included the rate of syllable repetition during the first 5 seconds, the rate of syllable production during the entire sample, the regularity of rhythm, the regularity of distinctness, the regularity of loudness, and the rater's overall impression (Gadesmann & Miller, 2008).

Inter-rater reliability between the speech-language pathologists in group one yielded intraclass correlation rates ranging from 0.080 to 0.640 on the 6 analyses. Given that an intraclass correlation of 0.70 is what is considered acceptable reliability, it is concerning that the group of speech-language pathologists did not yield reliable results. The non-speech-language pathologist group also yielded unreliable results with intraclass correlations ranging from 0.084 to 0.544 for the 6 analyses. In addition, intra-rater correlation was also deemed poor. Gadesmann and Miller (2008) compared 40 files that were completed by the rater a second time and found that only 11 were deemed significantly correlated. Gadesmann and Miller's study helps clinicians to understand the true reliability of a task commonly used in diagnostic protocols.

Another example of an oral DDK study that clinicians could use to help make conclusions regarding LDDK was by Ackerman et al. (1995). Ackermann et al. studied oral DDK patterns in individuals with Parkinson's disease, Huntington's disease, Friedreich's ataxia, and those with a purely cerebellar syndrome to determine its efficacy as a diagnostic tool for these diseases as compared to a control group of 15 subjects who had central nervous system disease/trauma in a case-control study (Ackermann, Hertrich, & Hehr, 1995). Participants were asked to repeat each syllable (/pa/, /ta/, and /ka/) as quickly as possible on a single breath. Analysis of the recorded task was completed using a computerized speech lab, allowing the segmentation of the speech signal to be displayed on the screen for visual (and auditory) data for counting. Ackermann et al. measured four parameters: "(1) the mean number of syllable repetitions per train; (2) the median syllable duration; (3) the intratrain variation coefficient of syllable durations; (4) the percentage of incomplete stop consonant occlusions during the three syllable trains" (Ackermann et al., 1995, p. 17). Results found there to be a significant difference ($p<0.00001$) between the control and disease-specific groups on these four parameters. In addition, a significant difference was found between all of the disease groups with the exception of Huntington's disease and cerebellar syndrome. More specifically, Parkinson's disease and Friedreich's ataxia presented with specific, unique characteristics during task performance than other groups. Participants with Parkinson's disease "produced rather long repetition trains at a normal syllabic rate, comprising mostly incomplete closures" (Ackermann et al., 1995, p. 20). Participants with Freidreich's ataxia presented with decreased repetitions and a slower rate with complete closure (Ackermann et al., 1995).

After data analysis, Ackermann et al. came to more broad conclusions than the previous unique characteristics about the differences between these disease groups and the control. The Huntington's disease, Friedreich's ataxia, and cerebellar syndrome groups demonstrated a lower syllabic rate concomitantly with increased syllable duration as compared to the control group. These findings determined the four parameters of oral DDK that are used in this study to be a sensitive measurement of orofacial motor impairment in general and to have group specificity for Parkinson's disease and Friedreich's ataxia (Ackermann et al., 1995).

Conclusion

Laryngeal diadochokinesis is an ideal screening tool for neuromotor function of the larynx because it provides a simple assessment of both neuromotor innervation and structural integrity. When compared to other popular measures such as endoscopy, EMG, EGG, and acoustic measures, it is more efficient, cheaper, more mobile, and less invasive. Despite the potential benefits of LDDK, there is a lack of evidence pertaining to the standardization of an LDDK protocol and establishing validity and reliability. When considering oral DDK, which is a similar measurement tool, inter- and intra-reliability is not as high as it should be for clinical use according to a study by Gadesmann and Miller (Gadesmann & Miller, 2008). If this is the case for a very similar and more popular measurement, then it is reasonable to believe that the same may also be true for LDDK. In conclusion, the LDDK is not well studied in the current evidence base and more studies are needed to determine reliability, validity, and to develop a standardized protocol.

CHAPTER II

RATIONALE FOR STUDY

Assessment tools that are reliable across speech-language pathologists are essential for correctly identifying voice disorders. Tools that lack reliability risk both under- and over-diagnosing clients. Given the lack of reliability as shown in oral DDK, and the lack of research to establish reliability in LDDK, this study looks to establish the inter-rater reliability between a group of speech-language pathologists. Reliability is a measure of how a measure can be duplicated (Koo & Li, 2016). Interrater reliability is the similarity of data between more than one rater. Establishing interrater reliability among speech-language pathologists for LDDK will determine if the pencil dotting method of LDDK data collection should be utilized and trusted in a clinical setting.

In addition to establishing reliability for LDDK, this study also aimed to help clinicians make informed decisions regarding which LDDK stimulus to use. In the present research, there are contradictions regarding whether an adductory vs. abductory stimulus is best (/ʌ/ vs. /hʌ/). Gratzmiller (2012) stated that there is no statistically significant difference between rates calculated during the two stimuli, while Canter recommended using a syllable with a glottal consonant such as /hʌ/ because repetition of a single vowel might create “pulses of air pressure acting on a fixed laryngeal valve,” thereby not giving a valid measurement (as cited in Shanks, 1966, p. 7). If there was a difference between the reliability of human raters on one stimulus versus the other, it would suggest that clinicians should be using the stimulus in which human raters are more reliable.

CHAPTER III

METHODS

This study utilized an observational-analytic within-subjects design. To determine the reliability of laryngeal diadochokinesis, participants were presented with recordings of laryngeal diadochokinesis and asked to calculate a rate using a pencil dotting method. Rates obtained by participants were compared to other participants and rates obtained utilizing a waveform analysis.

Eligible participants were recruited for participation in this study based on their program of study and level of completion in the program. Participants were graduate students at Indiana University of Pennsylvania who were completing their third semester of graduate work in the speech-language pathology Master's program. All participants in the study had completed the necessary coursework to have an understanding of the LDDK task during their second semester of the program by completing SPLP 618, a course on voice disorders. Eligible participants received hearing screenings as a requirement of the program and passed. Eligible participants were contacted via email with a copy of the *Informed Consent Form* which stated the intentions of the study and possible risks and rewards of participation. Eligible participants were asked to contact the researcher regarding their participation and any questions or concerns. Participants were asked to sign a copy of the *Informed Consent Form* prior to participation in the study. The *Informed Consent Form* and study protocol were approved by the Indiana University of Pennsylvania Institutional Review Board. Of the 21 eligible participants, 12 volunteered to participate. Therefore, the final sample consisted of 12 participants who were all 3rd semester graduate students in the speech-language pathology Master of

Science program at Indiana University of Pennsylvania. All participants were female and between the ages of 22-27.

Participants were presented with audio files through classroom presentation speakers equipped in the classrooms at Indiana University of Pennsylvania. Audio files were played for all 12 participants at the same time. Participants were given the opportunity to listen to audio files as many times as necessary. Participants indicated that they needed to hear the audio file again by raising their hands after the recording finished playing. Of 26 recordings, four recordings were repeated one time and one recording was repeated twice. Participants were presented with one additional audio file to practice calculating rate and were able to ask questions regarding procedure once this practice trail was completed. No questions were asked.

Participants in this study used a method of tallying the number of syllable repetitions for each audio file that was a simulation of the pencil dotting method often used in clinical settings. Participants listened to the audio files and tallied on paper using a pencil dot system how many times the syllable was repeated. Participants then divided the number of repetitions by the total time of the audio file to get a rate using a calculator. Participants recorded these calculated rates onto deidentified recording paper which was submitted to the researcher.

Data collected from the participants were compared to waveform analyses conducted by the researcher in order to determine the reliability of the pencil dot method. The researcher calculated a rate for each of the audio files using a waveform analysis to compare accuracy of the rate calculations. Waveform analysis was completed using an oscillogram in the Multidimensional Voice Program software (MDVP). Using the waveform analysis, the researcher counted the number of waveform oscillations and divided it by the time of the recording in seconds.

Audio files of LDDK performance were used from a database of LDDK recordings used in the umbrella research project (IRB 11-131 Laryngeal Diadochokinesis: Clinical Measurement and Age Related Issues). All audio files had previously been trimmed to be 5 seconds in length due to the needs of the study the audio files were obtained for. Audio files varied across age and gender. Demographic information pertaining to the audio files was listed in Table 1. Audio files are listed in the order they were presented to the participants.

Table 1

Audio File Demographics

File Number	Deidentified LDDK File Code	Gender	Age Range (years)	LDDK Stimulus Used
1	12M	Male	20-29	/ʌ/
2	39F	Female	20-29	/ʌ/
3	24M	Male	30-39	/ʌ/
4	33F	Female	30-39	/ʌ/
5	38M	Male	40-49	/ʌ/
6**	3F	Female	40-49	/ʌ/
7	2M	Male	50-59	/ʌ/
8	20F	Female	50-59	/ʌ/
9	3M	Male	60-69	/ʌ/
10	34F	Female	60-69	/ʌ/
11*	60M	Male	70-79	/ʌ/
12	342M	Male	80-89	/ʌ/
13*	46F	Female	80-89	/ʌ/

14	12M	Male	20-29	/hʌ/
15	39F	Female	20-29	/hʌ/
16	24M	Male	30-39	/hʌ/
17*	33F	Female	30-39	/hʌ/
18	38M	Male	40-49	/hʌ/
19*	3F	Female	40-49	/hʌ/
20	2M	Male	50-59	/hʌ/
21	20F	Female	50-59	/hʌ/
22	3M	Male	60-69	/hʌ/
23	34F	Female	60-69	/hʌ/
24	60M	Male	70-79	/hʌ/
25	342M	Male	80-89	/hʌ/
26	46F	Female	80-89	/hʌ/

Note
* indicates the audio file was repeated, the number of * indicates the number of times the audio file was replayed

CHAPTER IV

DATA ANALYSIS

The IBM SPSS Statistics Data Editor software was utilized for all statistical analyses of this study. The intraclass correlation measurement was used to determine interrater reliability among the 12 participants. Intraclass correlation (ICC) is a measurement of reliability that provides information regarding the degree of correlation between measurements and the agreement between two measurements. ICC for this study was calculated with a mean rating (k=12), absolute agreement, two-way mixed effects model. Koo & Li (2016) define ICC estimates and their quality. Their estimates and quality ranges were used to evaluate the reliability data in the current study (see Table 2).

Table 2

Intraclass Correlation Qualities

ICC Estimate	**Quality**
Less than 0.5	Poor
0.5-0.75	Moderate
0.75-0.9	Good
0.9+	Excellent

The first statistical analysis was the interrater reliability of the 12 participants, specifically on audio files with /ʌ/ as the LDDK stimulus. The ICC estimate for these rates indicated the interrater reliability was 0.838, which was considered moderate to excellent based on the 95% confidence interval (see Table 3).

Table 3

Intraclass Correlation for /ʌ/ Audio Files

Intraclass Correlation	**95% Confidence Interval**	
	Lower Bound	Upper Bound
.838	.698	.937

The second statistical analysis was the interrater reliability of the 12 participants, specifically on audio files with /hʌ/ as the LDDK stimulus. The ICC estimate for these rates was 0.678, which indicated the interrater reliability is poor to good based on the 95% confidence interval (see Table 4).

Table 4

Intraclass Correlation for /hʌ/ Audio Files

Intraclass Correlation	**95% Confidence Interval**	
	Lower Bound	Upper Bound
.678	.468	.862

A paired t-test analysis was completed to determine if there is a statistically significant difference between the average of the 12 human raters on all audio files and the waveform analysis. When the human perceptual raters were compared to the researcher waveform analysis using the /ʌ/ LDDK stimulus, p<0.01, indicating a statistically significant difference between the average of rates and waveform analysis. The difference between the average of the human raters (4.4795 repetitions/sec) and the waveform analysis (4.3692 repetitions/ sec) was 0.1103 repetitions/sec. Differences between the average of human raters to the waveform calculation for each audio file were calculated (see Table 5). These differences were representative of the average across all audio files, with a range of 0.06 repetitions/sec to 0.55 repetitions/sec with the majority of differences falling between 0.10 repetitions/sec to 0.20 repetitions/sec.

Table 5

Calculated Rates (Repetitions/ sec) for /ʌ/ Audio Files

Audio File	**Wave-form Analysis**	**Participants Mean**	**Participants**											
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
12M	4.8	4.95	5.0	5.2	4.2	5.8	4.6	5.4	4.8	5.6	4.6	48	5.2	5.2
39F	5.0	5.27	5.4	5.4	5.0	5.2	4.8	5.4	5.2	5.6	5.2	5.0	5.4	5.6
24M	4.4	4.58	4.4	4.6	4.4	4.6	4.2	4.4	4.4	5.0	4.6	4.4	4.6	5.4
33F	4.6	4.82	4.8	5.2	4.6	4.6	4.6	4.8	4.8	5.0	4.8	4.6	4.6	5.4
38M	3.6	3.62	3.8	3.8	3.6	3.4	3.8	3.6	3.6	3.8	3.2	3.6	3.6	3.6
3F	5.4	5.19	5.0	5.6	5.2	4.8	4.8	5.6	5.6	5.2	4.8	5.2	5.0	5.4
2M	4.6	4.57	4.4	4.8	4.4	4.6	4.4	5.0	4.8	4.4	4.4	4.4	4.2	5.0
20F	4.8	4.97	4.8	5.2	4.6	4.8	4.4	5.0	5.0	5.2	5.2	4.8	5.0	5.6
3M	4.4	4.56	4.2	4.6	4.6	4.6	4.4	5.0	4.4	4.8	4.6	4.2	4.4	5.0
34F	4.6	4.78	4.6	5.0	4.8	4.6	4.6	5.2	4.6	4.8	4.4	4.6	4.8	5.4
60M	4.4	4.56	4.2	4.8	4.2	4.6	4.4	4.8	4.2	5.0	4.2	4.4	4.2	5.0
342M	4.2	4.08	4.2	4.6	4.4	4.6	4.2	4.6	4.2	4.8	4.4	4.2	4.5	5.0
46F	4.0	4.15	3.8	4.2	3.8	4.0	4.0	4.0	3.8	4.6	4.8	4.0	4.0	4.8

A paired t-test analysis was also completed, comparing human raters and the waveform analysis using the /hʌ/ LDDK stimulus, also indicating a statistically significant difference ($p<0.01$) between the rater average and the waveform analysis. The difference between the average of the human raters (4.6841 repetitions/sec) and the waveform analysis (4.5231 repetitions/sec) using the /ʌ/ stimulus was 0.125 repetitions/sec. Differences between the average of human raters to the waveform calculation for each audio file were calculated (see Table 6). These differences were representative of the average across all audio files, with a range of 0.02 repetitions/sec to 0.27 repetitions/sec.

Table 6

Calculated Rates (Repetitions/ sec) for /hʌ/ Audio Files

Audio File	Wave-form Analysis	Participants Mean	Participants 1	2	3	4	5	6	7	8	9	10	11	12
12M	4.4	4.46	4.4	4.8	4.4	4.2	4.0	4.8	4.4	4.6	3.8	4.2	5.0	5.0
39F	5.2	5.30	5.4	5.4	4.6	5.2	4.8	5.8	5.4	5.8	5.2	5.2	5.4	5.4
24M	4.4	4.55	4.4	4.4	4.2	4.6	4.2	4.6	4.6	5.0	4.2	4.6	4.6	5.2
33F	4.8	4.95	5.0	5.2	4.8	4.8	4.6	4.8	5.0	5.0	5.0	4.6	5.2	5.4
38M	3.0	3.08	3.2	3.0	3.0	2.8	3.0	3.0	3.2	3.4	3.2	3.0	3.0	3.2
3F	5.4	5.50	5.8	5.4	5.2	5.2	5.2	6.2	5.8	5.4	5.2	5.6	5.0	6.0
2M	4.0	4.20	4.0	4.6	4.2	4.2	4.2	4.6	4.2	4.4	4.0	3.8	3.8	4.4
20F	4.8	4.97	5.0	5.0	4.6	5.0	4.4	5.0	5.2	5.2	5.0	5.0	4.6	5.6
3M	3.8	4.00	3.8	4.0	3.8	3.8	3.8	4.2	4.0	4.4	4.2	3.8	3.8	4.4
34F	5.0	5.12	4.8	5.0	5.0	5.2	5.0	5.6	5.4	5.0	4.6	5.0	5.0	5.8
60M	4.8	4.65	4.0	5.0	4.6	4.6	4.2	5.4	4.8	4.8	4.6	4.6	4.6	5.6
342M	3.4	3.45	3.4	3.6	3.4	3.6	3.4	3.6	3.6	2.4	3.6	3.4	3.4	4.0
46F	3.8	3.25	3.8	4.0	3.6	4.0	3.8	4.2	3.8	4.2	3.8	3.8	4.0	4.0

To determine the impact of these statistically significant differences, the calculations were compared to average diadochokinetic rates. This data is represented in Figure 1 (Lombard, Weible, VanHorn, & Solomon, 2017). This data places LDDK rate performance into three categories, average, above average, and below average rate productions. Using these categories, both the waveform analysis and the rates calculated by human raters are compared to determine if the human raters concluded the same clinical findings as the waveform analysis (see Table 7). Of the calculated rates for /ʌ/ audio files, 92% agreed with the rate performance category determined by the waveform analysis. Of the calculated rates for /hʌ/ audio files, 97% agreed with the rate performance category determined by the waveform analysis. When considering disordered vs. normal, new clinicians are able to calculate rates that agree with the waveform analysis determination with 92-97% agreement when counting with the pencil dotting method.

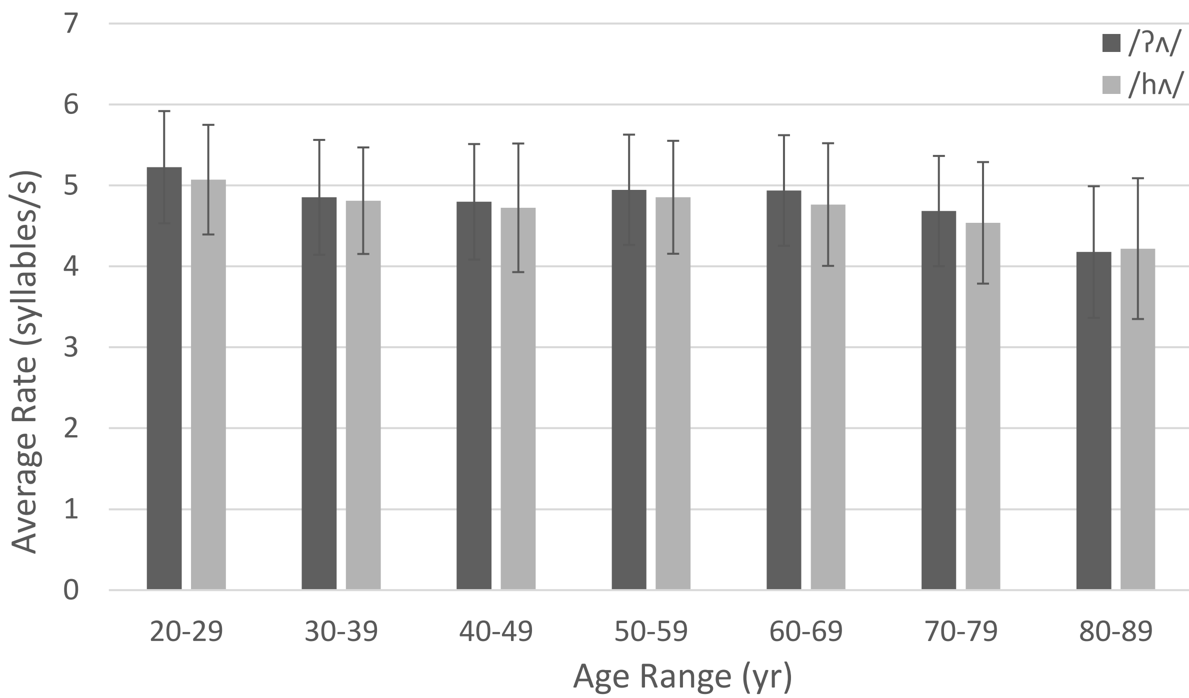

Figure 1. Normative data for LDDK rate. Adapted from "Laryngeal diadochokiensis across the adult lifespan," by Lombard, Weible, VanHorn, and Solomon, 2019. Paper presented at the Voice Foundation International Symposium, Philadelphia, PA.

Table 7

Agreement with LDDK Rate Averages

Audio File	**Waveform Analysis** (repetitions/ sec)	**Waveform Rate LDDK Performance Category**	**Participants Mean** (repetitions/ sec)	**Participants Mean LDDK Performance Category**	**Number of Participants with Agreement to Waveform Analysis**
/ʌ/ Stimulus					
12M	4.4	Below average	4.46	Below Average	10/12
39F	5.2	Average	5.30	Average	12/12
24M	4.4	Average	4.55	Average	12/12
33F	4.8	Average	4.95	Average	12/12
38M	3.0	Below Average	3.08	Below Average	12/12
3F	5.4	Average	5.50	Average	9/12
2M	4.0	Below Average	4.20	Below Average	9/12
20F	4.8	Average	4.97	Average	12/12
3M	3.8	Below Average	4.00	Below Average	10/12
34F	5.0	Average	5.12	Average	11/12
60M	4.8	Average	4.65	Average	11/12
342M	3.4	Average	3.45	Average	11/12
46F	3.8	Average	3.25	Average	12/12
/hʌ/ Stimulus					
12M	4.8	Average	4.95	Average	11/12
39F	5.0	Average	5.27	Average	12/12
24M	4.4	Average	4.58	Average	12/12
33F	4.6	Average	4.82	Average	12/12
38M	3.6	Below Average	3.62	Below Average	12/12
3F	5.4	Average	5.19	Average	9/12
2M	4.6	Average	4.57	Average	11/12
20F	4.8	Average	4.97	Average	12/12
3M	4.4	Average	4.56	Average	12/12
34F	4.6	Average	4.78	Average	12/12
60M	4.4	Average	4.50	Average	12/12
342M	4.2	Average	4.08	Average	12/12
46F	4.0	Average	4.15	Average	12/12

Waveform analysis rates for /ʌ/ audio files (see Table 5) ranged from 3.0 to 5.2 repetitions/ second, with a range of 2.2 repetitions/ sec. The distribution within the range was almost equally dispersed with four audio files between 3.0 and 3.9 repetitions/ sec, six audio files between 4.0 and 4.9 repetitions/sec, and three audio files between 5.0- 5.2

repetitions/ sec. Waveform analysis rates for /hʌ/ audio files (see Table 6) ranged from 3.6 repetitions/sec to 5.4 repetitions/ sec, with a range of 1.8 repetitions/ sec. The distribution within the range was not as equally distributed as during the /ʌ/ stimulus, with one audio file between 3.6-3.9 repetitions/ sec, eight audio files between 4.0-4.9 repetitions/sec, and two audio files between 5.0-5.4 repetitions/ sec.

CHAPTER V

DISCUSSION

While the validity of LDDK as a measurement of laryngeal motor function has been reported in previous studies, none of those studies identified the reliability of speech-language pathologists when measuring LDDK using the method that is the most feasible in clinical settings, the pencil dotting method. In addition, research regarding LDDK varied in use of stimuli, varying between /ʌ/ and /hʌ/. This study aimed to determine if speech-language pathologists are reliable when utilizing the pencil dotting method and if speech-language pathologists are more reliable when listening to one LDDK stimulus than another.

The results of the current study revealed that graduate student speech-language pathologists demonstrate moderate to excellent intraclass correlation (interrater reliability) between participants when calculating rates for /ʌ/ audio files. However, when compared to the objective measurement of LDDK rates (waveform analysis), the average of the participant rates was statistically different. The difference between the rater average and waveform analysis was 0.1103 repetitions /sec. When using LDDK in a clinical setting, a difference of 0.1103 repetitions/ sec warranted a difference in diagnostic outcomes in 8% of rate calculations.

When graduate student speech-language pathologists calculated rates for /hʌ/ audio files, the intraclass correlation was considered to be poor to good. Similarly, to the /ʌ./ stimulus, a statistically significant difference between human raters and waveform was found. The difference between the rater average and waveform analysis was 0.125 repetitions /sec. When using LDDK in a clinical setting, a difference of 0.125

repetitions/ sec warranted a difference in the diagnostic outcomes in 3% of rate calculations.

Implications

When the pencil dotting method is used to count LDDK syllable repetitions for recordings across age and gender, graduate student speech-language pathologists at Indiana University of Pennsylvania calculated rates that were not reliable when compared to rates calculated using a waveform analysis. The intraclass correlations between human raters and researcher completed waveform analysis indicated a statistically significant difference when rating both stimuli. Despite these findings, the consideration of a 0.1103-0.125 average difference between the average of human perceptual raters and the waveform analysis still appears clinically insignificant based on diagnostic outcomes. New speech-language pathologists were found to agree with the diagnostic determination of the waveform analysis on 92% to 97% of calculated rates. This information should inform clinicians that the pencil dotting method of counting LDDK repetitions is not reliable until further research indicates otherwise to provide the most accurate diagnostic results. It is recommended that clinicians utilize a waveform analysis. Waveform analysis software is becoming widely accessible through smartphone and tablet applications.

When considering which stimulus should be utilized in a clinical setting, more research should be conducted regarding reliability before a conclusion is made. When rating the /ʌ/ stimulus, graduate student speech-language pathologists demonstrated moderate to excellent intraclass correlations, but also demonstrated a statistically significant on both stimuli. When rating the /hʌ/ stimulus, graduate student speech-language pathologists demonstrated poor to good intraclass correlations. Therefore, additional studies should be completed to determine if either of these stimuli are reliable to use in the clinical setting.

Limitations

This study should be replicated with both the same and different conditions to control for limitations of this original study to confirm findings. The first limitation was in regard to participant demographics. Participants in this study were all graduate student clinicians from the same university. Graduate students lack experience in the field and have limited to zero experience working directly with voice clients. In addition, all student participants were female and within the ages of 21-27 years. Future participants should vary, including speech-language pathologists who use LDDK often in their clinical work and received their graduate education from institutions other than Indiana University of Pennsylvania.

This study was also limited in both the number of human participants and the number of audio files presented. Future research should aim to use more human raters and should present raters with more audio files. With more human raters, more statistically trustworthy effect sizes could be calculated. In addition, this study did not address the raters ability to listen to the same audio file and produce same or similar results. Future research should also choose audio files with rates that are matched to ascertain if there is a difference between the interrater reliability and human rater accuracy. All audio files in this study were also deemed within normal limits by a previous study. All audio files also varied in both gender and age range. Possible impacts of the gender and age group of the person completing the LDDK were not explored in the present study.

In addition, the math computations to determine a rate should be checked by the researcher in future studies. In this study, participant math computations were not checked for correctness and therefore could have impacted the study. However, all audio files were trimmed to 5 seconds in length and the participant was responsible for dividing their counted number by 5. All participants had access to calculators during the study and it is unlikely that this had an impact on the data of the present study.

This study should also be replicated under different environmental circumstances. The classroom environment of the present study is not representative of a clinical setting. Future studies should replicate the clinical setting. Examples of this would include calculating the rate live, not with a recording in a one person recording, and one person completing the LDDK task. It would also include using the initial performance of the LDDK task during counting, and not allowing repetitions to be completed/utilized by the rater.

References

Ackermann, H., Hertrich, I., & Hehr, T. (1995). Oral Diadochokinesis in neurological dysarthrias. *Folia Phoniatrica et Logopaedica*, *47*(1), 15-23. doi: 10.1159/000266338

Bassich-Zeren, C. J. (2004). *Vocal dysfunction in young-onset Parkinson's Disease.* (Doctoral dissertation), University of Maryland, College Park, MD. Retrieved from http://drum.lib.umd.edu/handle/1903/1823

Fletcher, S. G. (1972). Time-by-count measurement of diadochokinetic syllable rate. *Journal Of Speech And Hearing Research, 15*(4), 763-770. doi: 10.1044/jshr.1504.736

Gadesmann, M., & Miller, N. (2008). Reliability of speech diadochokinetic test measurement. *International Journal of Language & Communication Disorders/ Royal College of Speech & Language Therapists, 43*(1), 41-54. doi:10.1080/13682820701234444

Garret, J., & Larson, C. (1991). Neurology of the laryngeal system. In *Phonosurgery: Assessment and surgical management of voice disorders* (Ed.), (pp. 43-76): New York, NY: Raven.

Gratzmiller, H. L. (2012). *A comparison of laryngeal diadochokinetic tasks.* (Master's Thesis), Retrieved from Proquest. (1508297).

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012

Leeper, H. A., & Jones, E. (1991). Frequecy and intensity effects upon temporal and aerodynamic aspects of vocal fold diadochokinesis. *Perceptual and Motor Skills, 73*(3 pt. 1), 880-882.

Lombard, L., Weible, K., VanHorn, M., & Solomon, N.P. (2017, June). *Laryngeal diadochokinesis across the adult lifespan.* Paper presented at the Voice Foundation International Symposium, Philadelphia, PA.

Modolo, D. J., Berretin-Felix, G., Genaro, K. F., & Brasolotto, A. G. (2011). Oral and vocal fold diadochokinesis in children. *Folia Phoniatrica et Logopaedica*, *63*(1), 1-8. doi:10.1159/000319728

Ptacek, P. H., Sander, E. K., Maloney, W. H., & Jackson, C. C. R. (1966). Phonatory and related changes with advanced age. *Journal of Speech, Language, and Hearing Research, 9*(3), 353-360. doi: 10.1044/jshr.0903.353

Rosen, C. A., Gartner-Schmidt, J., Casiano, R., Anderson, T. D., Johnson, F., Reussner, L., & McWhorter, A. (2007). Vocal fold augmentation with Calcium Hydroxylapatite (CaHA). *Otolaryngology-Head & Neck Surgery*, *136*(2), 198-204. doi: 10.1016/j.otohns.2006.07.014

Shanks, S. J. (1966). *An investigation of the nature of vocal fold diadochokinesis of adult subjects and of the effect of pitch, intensity and aging upon the performance of this phonatory task.* (Doctoral disseratation). Lousiana State University, Baton Rouge, LA. Retreived from http://worldcat.org/z-wcorg/database.

Stager, S. V., & Bielamowicz, S. A. (2010). Using laryngeal electromyography to differentiate presbylarynges From paresis. *Journal of Speech, Language, and Hearing Research*, *53*(1), 100-113. doi:10.1044/1092-4388(2009/08-0244)

Stemple, J., Roy, N., & Klaben, B. K. (2014). *Clinical voice pathology: Theory and management* (5th ed.). San Diego, CA: Plural Publisher.

Tomblin, J. B., Morris, H. L., & Spreistersbach, D. C. (2000). *Diagnosis in speech language pathology* (2nd ed.). San Diego, CA: Singular Publishing Group.

Verdolini, K., & Palmer, P. M. (1997). Assessment of a "profiles" approach to voice screening. *Journal of Medical Speech-Language Pathology, 5*(4), 217-232.

Williams, P., & Stackhouse, J. (2000). Rate, accuracy and consistency: Diadochokinetic performance of young, normally developing children. *Clinical Linguistics & Phonetics*, *14*(4), 267-293. Retreived from Ebscohost. (107143109).