Introduction

Appropriate tongue strength is essential for the oral and pharyngeal phases of swallowing and contributes to the formation, placement, and manipulation of a bolus within the oral cavity and propulsion into the pharynx [1]. Examination of tongue strength is a frequent component of the clinical assessment of swallowing by speech-language pathologists. Such assessment is usually based on subjective judgement of the force being applied by the tongue against resistance provided by the speech-language pathologist's fingers resting against the cheek or a tongue depressor. This method raises concerns regarding the reliability of tongue strength measurements due to an inability to eliminate assessor bias and the variability introduced by multiple assessors in most clinical environments. A number of tools have been designed to objectively quantify measures of tongue strength and endurance for research purposes and for routine clinical practice. Such tools have been used to study tongue strength across a range of ages [1-11], in both healthy and clinical populations, and have led to the development of a significant body of literature that documents values of tongue strength. Previous research has determined that the Iowa Oral Performance Instrument (IOPI: Figure 1) [12] is the most commonly used of these measurement devices to assess tongue strength [13]. Therefore, it is essential to establish the reliability of measurements obtained with the IOPI. In addition to isometric tongue strength, the IOPI can also be used to measure isometric tongue endurance, and the reliability of this measure should also be determined.

The reliability of a measurement is the reproducibility of the values obtained over multiple test sessions. If measures are reliable there is little error in the measurement and we can have confidence in the values obtained. There are two main types of reliability: inter-rater and test-retest (intra-rater). Inter-rater reliability assesses the degree to which values are consistent when obtained using different assessors. Test-retest reliability assesses the extent

to which the values obtained are consistent from one administration to another, and is performed by one assessor under the same test conditions on multiple occasions. These results provide an indication of the precision or variability with which these measures can be obtained. If this is known, it can be considered in determining the use of a particular measurement tool or interpretation of the values obtained. It is important to be able to differentiate between typical measurement error and real changes being assessed, such as whether a person's condition is improving or deteriorating or if a treatment is having the desired effect, and this is facilitated by a comprehensive evaluation of test-retest reliability. Reliable tests on a number of people over multiple test sessions have the following characteristics: there is little or no change in the group means of the sessions (differences may indicate learning effects); there is little or no within-subject variation over the sessions; and there is a strong test-retest correlation between the sessions [14].

A number of studies have investigated inter-rater reliability of tongue strength using the IOPI [1,3,15-24] in a range of populations. Typically, values obtained by novice users of the IOPI were compared to those of an experienced user. The measure of reliability was the correlation between values obtained from different users. The inter-rater correlation coefficients were all stronger than $r = 0.75$ with one exception; Solomon et al. (2008) reported $r = 0.535$ in a dysarthric population [21]. Youmans and Stierwalt (2006) also compared the group means between assessors and found no significant difference [1]. Only one study (Palmer 2010) reported inter-rater reliability for tongue endurance, with a perfect correlation ($r = 1$) between assessors [23].

Nine studies have reported test-retest reliability of tongue strength [3,16-21,24,25]. Robin et al. (1991) provided the first report describing the test-retest variability as low (implying reliability was high) based on the small size of an individual participant's standard deviations [7]. Subsequent studies [1,3,15-26] reported strong correlations as measures of

test-retest reliability with correlation coefficients ranging from $r = 0.76$ to $r = 0.99$. Only one study (Chang et al (2008) reported tongue endurance test-retest reliability ($r = 0.99$) [25].

In addition, Lazarus et al (2000) reported that assessors had to meet pre-established criteria of at least $r = 0.76$ for inter-observer reliability and $r = 0.90$ for test-retest reliability prior to conducting study assessments of tongue strength. Also, it should be noted that the highest tongue strength value obtained was used in all these investigations of the reliability of IOPI tongue strength measures. In summary, inter-rater and test-retest reliability of the IOPI measurement of tongue strength have been reported but there has been almost total reliance on correlation coefficients as the measure of reliability. Consequently, whether the values obtained change consistently with familiarization (identifiable by a change in the mean of a group of people) has not been identified. Further, the magnitude of any within-subject variation (typical error) that needs to be accounted for in interpreting clinical improvement has not been investigated.

As well as measuring tongue strength and endurance, the IOPI has an additional attachment that allows measurement of hand strength and endurance. Handgrip strength is an important predictor of functional decline associated with normal aging and is often used to characterize the general strength of individuals [27]. Consequently, it is appropriate to assess the reliability of the IOPI for handgrip strength and endurance at the same time as tongue strength and endurance; only Robin et al [7] indicated that these handgrip strength measures had low variability.

The primary aim of this study was to determine the test-retest reliability of the IOPI as a tool for assessments of both tongue and handgrip strength and endurance in a healthy population. A secondary aim was to identify characteristics of assessments that improve the reliability of these strength and endurance measurements for both research and clinical practice.

Methods

Study Design

Healthy adults underwent anterior and posterior tongue and handgrip strength and endurance assessment using the IOPI on four occasions separated by approximately one week. Strength assessments consisted of three attempts to exert maximal isometric force. Endurance assessments consisted of one attempt to sustain 50% of maximal isometric force. Participants were randomized to perform tongue or hand measurements first. One investigator (VA) provided all instructions to the participants and conducted all the tests. Three measures of reliability were assessed according to Hopkins [14]. Exploratory secondary analyses were also conducted to determine whether single peak or mean strength values were more reliable, and to identify other protocol strategies that influence the reliability of these strength and endurance measures.

Participants

Healthy adults were recruited from staff and students at The University of Newcastle. Each participant completed a health and medical history questionnaire to determine their eligibility. Participants were included if they ranged in age from 18 to 60 years, and were healthy with no previous or current swallowing or hand problems. Study exclusion criteria were a history of swallowing problems; abnormal oral structure and function; history of neurologic, respiratory or gastrointestinal impairment; any current or previous major injury to the tongue or hand; any tongue piercings; difficulty placing an instrument on the tongue; or a history of seizures. The University of Newcastle Human Research Ethics Committee approved the study and written informed consent was obtained from all participants prior to participation.

Instrumentation

Tongue strength and endurance assessments were collected using the current version (2.2) of the Iowa Oral Performance Instrument (IOPI) (Figure 1) [12] by placing a small, air-filled bulb longitudinally along the hard palate. The IOPI is a portable, handheld tool containing pressure-sensing circuitry, a peak-hold function, and a timer. It uses a blue air-filled PVC tongue bulb (approximately 3.5cm long and 1.2cm in diameter) which is pliable and has an approximate internal volume of 2.8ml. The bulb is connected to the IOPI via an 11.5cm PVC connecting tube with the pressure exerted against the bulb measured and displayed in kilopascals (kPa). Unlike earlier versions of the IOPI, which showed the green light as the middle light in a row of lights, the current model used in this study has the green light as the top light (100%). Handgrip strength and endurance were measured by placing a handgrip pressure bulb in the centre of the palm of the dominant hand, with the fingers wrapped around it. Participants were instructed not to press the bulb with the fingertips as this may create artificial increases in pressure. The handgrip bulb is made of soft rubber with a small air-filled bulb that is immersed in an incompressible viscous fluid in the middle. Visual feedback to participants for asessment of endurance was achieved by the light-emitting diode (LED) display on the IOPI screen. To ensure accuracy of measurement, calibration was checked once a week as recommended in the IOPI manual.

Procedure

Participants were seated in an upright position in a straight-backed chair for the duration of the testing performed at the university. Testing was conducted at various times during the day and participants were not required to fast prior to the assessment. Maximum tongue and handgrip strength and endurance measures were obtained following a previously documented procedure [6,7] with the order of tests randomized using a web-based random assignment generator. Attempts allowed in the first session included one or more non-

maximal practice trials to ensure the participant understood the task. Participants were provided with instructions for all tasks and verbal encouragement was given during each of the trials. All study participants were given verbal encouragement by the investigator saying "Push, push, push!" or "Squeeze, squeeze, squeeze". Maximum strength (P_{max}) was determined as the highest pressure recorded of the three trials [5]. The length of the endurance trial was measured in seconds using a stopwatch. Cessation of the endurance trial occurred when one of the following occurred: 1) 50% of P_{max} (represented by a green LED) could not be maintained for more than 2 s; 2) 80% of the P_{max} (represented by the second red light below the green LED) could not be maintained for more than 0.5 s; or 3) the pressure dropped sharply [28].

Tongue strength and endurance

Tongue strength and endurance data were collected in two bulb positions, the antero-median and the postero-median. To obtain antero-median measures, the IOPI bulb was placed in the centre of the tongue directly behind the front teeth (Figure 2). The postero-median position was defined by placing the straight edge of the IOPI bulb parallel to the anterior edge of the individual's back molars (Figure 3). Individual bulb placement using these landmarks allowed for a standardized placement in relation to normal structures within the oral cavity. Each participant was shown a picture of the correct bulb placement plus a standardized verbal description of the placement at the beginning of each testing session. The placement was then observed by the investigator prior to each measurement and further directions provided if necessary. While individual anatomy across participants varied (palatal shape and height of the palatal vault), standardized instruction and placement demonstrations were used to ensure the bulb location was as consistent as possible. Once the bulb was in the correct position in the oral cavity, participants were given instructions to push the bulb against the roof of their mouth with their tongue as hard as possible. Maximum tongue strength involved three

consecutive trials each of approximately two-second duration, with a short rest between trials while the investigator recorded the peak pressure measurement. No participant had a hypersensitive gag response with the bulb in the posterior position.

For endurance, the IOPI was set to 50% of the participants' maximal tongue strength and participants were required to press the bulb with the tongue against the roof of the mouth as hard as required to maintain the target force for as long as possible. Only one measurement of each anterior and posterior endurance measure was taken during each session. Timing was started when the pressure reached its target force as indicated by the appearance of a green light located on the right side of the device and participants were able to monitor their performance via the LED array.

While collecting tongue strength and endurance measures, the contributing role of the jaw has been questioned. The jaw provides structural support for the articulators, particularly the tongue and it has been suggested that tongue measures may include contributions from both the tongue and the jaw. This relationship was the basis of an investigation by Solomon and Munson (2004) [29] who examined tongue strength and endurance with 10 healthy adults where the jaw was unconstrained or constrained with a bite block. Results showed that measures of tongue function were lower when the jaw was constrained than when the jaw was unconstrained. Solomon and Munson (2004) determined that maximal measures of tongue strength and endurance were best assessed with an unconstrained jaw. Therefore, the jaw was unconstrained during all measurement tasks for this study.

Handgrip strength and endurance

The investigator (VA) ensured the correct position of the bulb within the hand (Figure 4) and participants were given instructions to squeeze the bulb as hard as possible with the whole hand for 1-2 seconds [12]. Maximum hand strength involved three consecutive trials of approximately one-second duration each, with a short rest between trials while the

investigator recorded the peak pressure measurement. For hand endurance, the IOPI was set to 50% of the participant's maximal hand strength and participants were required to squeeze the bulb as hard as required to maintain the target force for as long as possible. The timing procedure was the same as for tongue endurance.

Data management and analysis

All data were entered into Microsoft Excel (Microsoft Windows XP Professional, Version 5.1.2600) for data management and then exported into appropriate analysis programs. Participant characteristics were analyzed using a statistical software program (SPSS Statistics 20) and descriptive statistics are presented as a mean ± SD. All reliability measures were analyzed using a reliability spreadsheet developed by Hopkins and designed to assess the precision of measurement [30]. Three statistical analyses providing different indices of reliability were used. Random and systematic change outcomes through sampling error and learning effects were assessed using change in the mean between sessions. Within-subject variation was determined using typical error expressed as a coefficient of variation (%) as follows:

typical error = [(*sdiff*/√2)/mean]/100 where *sdiff* is the standard deviation of difference scores between two trials. This measure represents technical and biological sources of error in measurement within participants. Rank order repeatability of the results among trials was investigated using intraclass correlation coefficients (ICC, *r*). An acceptable level of variability in test measures is up to the researcher to determine, however values for reliability measures are a change in the mean and typical error between sessions of less than 5% (desirable) or 10% (acceptable), and ICC levels above 0.8 (desirable) and 0.6 (acceptable) [14]. The magnitude of any change was assessed by effect sizes using: large ($d > 0.8$); medium ($d = 0.5$ to 0.79); small ($d = 0.2$ to 0.49); and anything smaller than $d = 0.19$ was regarded as insubstantial or trivial [31].

When three strength trials of a measure are conducted in a session, there is uncertainty regarding which of these values should be used in evaluation, and in this study, whether this choice altered the reliability of the measures. Maximum tongue (anterior and posterior positions) and handgrip strength values were analyzed using three approaches: 1) the highest of the three trials in the session; 2) the average of the three trials in the session; and 3) the average of the two highest trials in the session. Further, when three strength trials were conducted in a session, it is possible that the values may vary substantially. There is greater confidence that a true measure of maximal strength has been obtained in a session if the variation between the two highest values is small. Therefore, additional exploratory analyses were conducted in subsets of participants where the two highest values obtained for a measure in a session varied by ≤ 5 kPa for tongue strength and ≤ 15 kPa for handgrip strength. Participants were included in this additional analysis if they met the criteria in all four sessions. If the maximal strength values obtained varied substantially between sessions, and the maximum strength obtained in the session was used to set the force target to assess endurance, then substantial variation has been introduced to the endurance assessment. Therefore a secondary analysis of the reliability of endurance assessments was conducted in a subset of participants where the maximal force used to set the 50% target force varied by ≤ 5 kPa for tongue strength and ≤ 15 kPa for handgrip strength across all four sessions.

Following the reliability analysis, the ‘minimum-raw-change required’ was determined to give an indication of the magnitude of change in a value needed for a meaningful change in the tested group mean with 95% confidence if such changes are to be used as outcome measures in intervention studies [14]. This value can also be used to determine sample sizes for future studies. This figure was calculated using the session 1 test mean multiplied by the percentage typical error between sessions 1-2 (upper 95% CI) [32]. A

second calculation was performed based on the sessions 2-3 data and compared to the first calculation.

Results

Fifty-one participants (21 males and 30 females) were recruited. All participants met the inclusion criteria and no potential participants met any exclusion criteria. Characteristics of the participants are presented in Table 1. The mean (± SD) time between assessments sessions was 12 ± 9 (range 5-21) days.

TONGUE AND HAND STRENGTH ANALYSIS

Analysis based on highest maximum strength value from three trials

Reliability statistics for the highest of three trials in each session for tongue (anterior and posterior) and hand strength are presented in Table 2. Similar reliability patterns were observed for both the tongue and hand strength measures.

Change in the mean

For anterior tongue strength, the change in the mean was largest between sessions 1-2 and substantially smaller in subsequent sessions. The mean difference between sessions 1-2 was 1.02 kPa (95% confidence interval (CI): -1.27 - 3.31). Analysis by paired *t*-test showed that the difference between sessions 1-2 was not significant ($p = 0.375$), and the magnitude of this difference was determined to be trivial (d=0.08) using effect size. The mean differences between sessions 2-3 (0.04 kPa; 95% CI: -1.49 - 1.57) and sessions 3-4 (0.17 kPa; 95% CI: -1.56 – 1.21) were also not statistically significant ($p = 0.96$ and $p = 0.80$, respectively) and were trivial in magnitude (effect sizes were $d = 0.003$ and $d = 0.07$, respectively).

For posterior tongue strength, the change in the mean was also largest between sessions 1-2 and substantially smaller between sessions 2-3 and 3-4. The mean difference between sessions 1-2 was 1.26 kPa (95% CI: -0.99 – 3.50; $p = 0.267$), and the magnitude of this difference was trivial ($d = 0.16$). The mean differences between sessions 2-3 (0.08 kPa;

95% CI: -1.55 – 1.39) and sessions 3-4 (0.14 kPa; CI: -1.68 – 1.95) were also not statistically significant ($p = 0.915$ and $p = 0.880$, respectively; effect sizes: $d = 0.003$ and $d = 0.01$, respectively).

For handgrip strength, the change in the mean was also largest between sessions 1-2 and substantially smaller in subsequent sessions, although the magnitude of variation was higher with handgrip strength compared to tongue strength. The mean difference ($p = 0.05$) between sessions 1-2 was 8.02 kPa (95% CI: 0.01 -16.03), which was small in magnitude ($d = 0.22$), whereas mean differences between sessions 2-3 (2.56 kPa; 95% CI: -3.73 – 8.87; $p = 0.416$; $d = 0.02$) and sessions 3-4 (0.52 kPa; 95% CI: -6.27 – 5.21; $p = 0.854$; $d = 0.03$) were trivial.

These results indicate good reliability for group assessments of tongue and hand strength; also one familiarization session provided improved reliability of the values obtained in healthy adults.

Typical error

In general, the typical error based on the highest value of the three trials was > 10% and therefore higher than the criterion standard for acceptable for all three strength measures between sessions 1 and 2. Typical error decreased after the first session for all strength measures however only anterior tongue strength typical error clearly met the criterion for acceptability.

Intraclass Correlation Coefficient (ICC)

The ICCs for strength ranged from acceptable to desirable levels, again with higher ICCs being achieved following session 1. All measures of tongue strength showed good reliability as indicated by correlation coefficients considered large to very large (0.77 – 0.90). Measures of handgrip strength also showed good reliability and were considered large to very large (0.69 – 0.91) [32].

Analyses using the average of two or three trials

Additional reliability analyses were conducted using the average of the three trials in each session or the average of the highest two values obtained. Little difference was observed between these two approaches, and only data from the average of the highest two trials are presented in Table 3. Similar patterns of response to those obtained using the highest of three values were observed regarding improved reliability after session 1, again supporting the benefits of familiarization. Some small reductions in typical error using an average value compared to a single maximum value were observed.

Additional criteria to reduce typical error

As described above, change in the mean and ICC indicators of reliability met the acceptable criteria but typical error levels were generally higher than acceptable. Therefore, an exploratory analysis was conducted with a subset of the data with the following additional criteria: for tongue strength, participants (anterior: $n = 28$ and posterior: $n = 25$) with the average of the two highest values within a session differing by ≤ 5 kPa, and for hand strength ($n = 28$), values differing by ≤ 15 kPa were included (Table 4). The primary impact of the additional criteria was to reduce the typical error to acceptable or even desirable levels after session 1. This has important implications for the reliability of values obtained when monitoring individuals rather than groups.

Minimum change score required

This value indicates the minimum magnitude of change in a variable required for the change to be meaningful, for example, following an intervention, and depends in part on the reliability of the measurements. As can be observed in Tables 2-4, the minimum raw change required is much higher if the first session is used for the determination compared to data from a subsequent session, again reinforcing the advantages of familiarization. The second observation is that using the additional criteria of having at least two measures within 5 kPa

(tongue) or 15 kPa (hand) provides a further reduction in the minimum raw change required to be meaningful.

TONGUE AND HAND ENDURANCE ANALYSIS

Reliability statistics for tongue (anterior and posterior) and hand endurance are presented in Table 5. Similar reliability patterns were observed for both tongue and hand endurance measures.

Analysis based on endurance values with all participants

Change in the mean

For anterior tongue endurance, the change in the mean was not necessarily improved with subsequent sesssions. The mean difference between sessions 1-2 was -2.49 s (95% CI: -4.86 – -0.12) which although small (effect size d = 0.28) was statistically significant (p = 0.04). The mean differences between sessions 2-3 (1.04 s; 95% CI: -1.33 – 3.48; p = 0.37; d = 0.13, trivial) and sessions 3-4 (1.55 s; 95% CI: -0.53 – 3.63; p = 0.140; d=0.18, trivial) were not statistically significant.

For posterior tongue endurance, the change in the mean was largest between sessions 1-2 and substantially smaller between sessions 2-3 and 3-4. The mean difference between trials 1-2 was 0.47 s (95% CI: -1.36 – 2.30; p = 0.607; d = 0.06, trivial). The mean differences between sessions 2-3 (0.59 s; 95% CI: -1.24 – 2.41) and sessions 3-4 (0.12 s; CI: -1.46 – 1.69) were also not statistically significant (p = 0.520 and p = 0.881, respectively) and were trivial in magnitude (effect sizes were d = 0.07 and d = 0.01, respectively).

For handgrip endurance, the change in the mean was also largest between sessions 1-2 and substantially smaller in subsequent sessions, although the magnitude of variation was higher with handgrip endurance compared to tongue endurance. The mean difference between sessions 1-2 was -6.77 s (95% CI: -15.45 -1.92; p = 0.12), which was small in magnitude (d = 0.26) whereas the mean differences between sessions 2-3 (1.20 s; 95% CI: -

6.11 – 8.50; $p = 0.744$; $d = 0.03$) and sessions 3-4 (-0.92 s; 95% CI: -7.83 – 5.99; $p = 0.790$; $d = 0.03$) were trivial.

Typical error

In general, the typical error based on the highest value of the three trials was > 10% and therefore higher than the criterion standard for acceptable for all three endurance measures. Typical error improved after the first trial with reduced variation most noticeable in posterior tongue endurance. Although anterior tongue endurance and handgrip endurance typical errors showed improvement following session 1, the typical errors of all endurance measures were considered unacceptable.

Intraclass Correlation Coefficient (ICC)

The ICCs ranged from unacceptable to acceptable levels, again with higher ICCs being achieved following session 1 for posterior tongue and hand endurance. All trials of tongue endurance showed moderate reliability as indicated by correlation coefficients considered small to medium (0.47 – 0.79). Trials of handgrip endurance also showed poor reliability and the correlations were considered small to medium (0.27 – 0.72) [32].

These results indicate moderate reliability for group assessments of posterior tongue and hand endurance following one familiarization session, but poor reliability of individual measurements for all endurance assessments.

Endurance analyses using values of maximal strength that were ≤ 5 kPa or 15 kPa apart

Additional reliability analyses were conducted using the participants where the maximium strength values across sessions were consistently ≤ 5 kPa apart for tongue strength and ≤ 15 kPa for hand strength (Table 6). Using this approach, change in the mean values for posterior tongue and hand endurance improved following session 1 and met either desirable or acceptable levels. Little improvement was observed using this approach for anterior tongue endurance. In general, typical error was much higher than the criterion standard of

acceptable, i.e., > 10% for all three endurance measures, ranging from 52.1% – 78.2% for anterior; 38.8% – 54.6% for posterior; and 25.7% – 45.1% for the hand.

Minimum raw change required for endurance measures

As can be observed in Table 6, the minimum raw change required was generally higher if data from the first session was used compared to the subsequent session, again reinforcing the advantages of familiarization, although the impact of this was most notable for hand endurance. The second observation is that using the additional criteria of reducing the variation of the strength value to within 5 kPa (tongue) or 15 kPa (hand) between sessions provided a further reduction in the minimum raw change required only for anterior tongue endurance.

Discussion

The key findings of this study are that tongue and hand isometric strength measurements obtained using the IOPI demonstrate excellent reliability for analysis of groups when a familiarization session is provided prior to clinical evaluation. Further, performing multiple trials within an assessment session with consistency criteria is an additional strategy to improve the reliability of these strength measurements. These strategies also improve the sensitivity of the IOPI measurements for evaluating strength improvements and the effectiveness of interventions in individuals. Unlike excellent reliability for hand and tongue strength measures, the reliability of the tongue and hand endurance measurements was generally unsatisfactory and requires further investigation.

The test-retest correlation coefficients for tongue strength observed in this study were similar to those reported previously [3,16,18-21,25,33,34] where correlation coefficients ranged between 0.75 and 0.99. Previous studies only compared the results of two sessions whereas the current study looked at the values obtained across four sessions. No previous studies reported changes in the means or indications of typical error, therefore this is the first

study to provide these important indices of IOPI measurement reliability. Chang et al (2008) are the only investigators to have previously reported the reliability of tongue endurance measurements. In contrast to the extremely high correlation (r = 0.99) value they obtained, the correlations in the current study are poor. Some possible reasons for this discrepancy are provided below. This suggests a need for further investigation of the reliability of tongue endurance values and the circumstances that contribute to more reliable values. Although Robin et al [7] stated that IOPI handgrip strength measures had low variability no quantified measure of reliability was reported, therefore the current study provides the first measures of reliability of IOPI handgrip strength and endurance measures.

The findings of this study have important applications for both researchers and clinicians. For researchers, the small to trivial changes in the means and high ICCs indicated the excellent reliability of the tongue and handgrip strength measures for group analysis. For all the IOPI strength measures, reliability was improved by one familiarization session to more desirable levels. The implications of this finding for people who are afforded only one opportunity to have their tongue strength tested before an intervention is that a greater increase in the value of the measure post-intervention is required before it can be concluded that real improvement has occurred. Other strategies to provide some familiarization within the first session such as additional trials may be preferable to the participant being required to return on another day. An important consequence of a reduction in variation is that the magnitude of change required to be regarded as meaningful (minimum raw change values) is reduced, which has additional benefits for researchers in reducing the sample size required in research studies. Familiarization typically reduced the magnitude of change in strength that would be meaningful by approximately 50%.

The question of whether the single highest strength value [1,22,26,35-40] or an average of multiple trials [41] should be used was also investigated, as both have been

reported in the literature. Differences between these approaches were small, and no approach was preferred for analysis of group data. One consideration is that there are practical reasons for not including one poor trial value in the assessment; therefore, the maximum value or the mean of the two highest values would be preferred. When consideration is given to the impact of improving within-session consistency of values (i.e., at least two values within a criterion range such as < 5kPa for tongue strength), there were small improvements in the reliability of the strength measures, particularly for handgrip strength, which importantly improves confidence that reliable maximum values have been obtained.

For clinical practice, the typical error analysis is the most important of the reliability measures as this provides an indication of the variability within an individual between sessions. Typical error was higher than the acceptable standard between the first two sessions but was reduced by familiarization and by using the average of the highest two of three values taken in a session. It was further improved by using within session consistency criteria to better establish that at least two similar near maximal force values have been obtained in a session. This suggests that in a clinical situation, more than three attempts may be required to meet the consistency criterion.

In contrast to the acceptable reliability demonstrated for tongue and hand strength, the reliability of tongue and handgrip endurance measurements was not established. Changes in the mean values were above 10% between trials 1-2, although they did decrease following subsequent trials. Typical errors were unacceptably large and ICC values were weak to moderate for both tongue and hand endurance. Therefore, further exploratory analysis was conducted to improve the reliability of endurance measures. Using the between session consistency criteria of having at least two strength measures within 5 kPa (tongue) or 15 kPa (hand) of each other resulted in acceptable reliability for posterior tongue and hand endurance

but little improvement in anterior tongue endurance. However, this approach excluded most participants from the analysis, which limits confidence in this strategy.

The method of data collection, i.e., using 50% of maximal tongue strength in each session as the endurance target, may have contributed to the unsatisfactory endurance results obtained. A previous study by Solomon, Robin, & Luschei (2000) assessed tongue strength, endurance, and stability during a sustained submaximal effort in 16 people with mild to severe Parkinson disease [17]. In the study, the authors set a definition for analyzing endurance (a steep drop in pressure; pressure signal was > 40% and < 50% of P_{max} for two seconds (s); or pressure signal was < 40% of P_{max} for 0.5s. The study also explained how changes in stability over time during a fatiguing task were measured from the endurance trials (determined by measuring five 3-second segments of each trial). The authors did not find a difference in stability between the experimental and control groups, and cannot attribute differences in endurance to a problem with stability. They did not investigate reliability for endurance. An alternative would be to set the endurance target at 50% of the maximal tongue strength achieved in session 1. Therefore, fu[42]rther investigations of the reliability of tongue and hand endurance measurements need to be undertaken, and considerations should be given to protocols and methodological strategies that could improve reliability.

The current study investigating the reliability of tongue and hand strength and endurance using the IOPI had a number of strengths. Three measures of reliability were used in the analyses providing indices of systematic and random error, with implications for both group and individual applications. An appropriate sample size was used for this analysis, and the population included healthy males and females across an age range from 18-60 years.

However, there were some limitations. Inter-rater reliability was not investigated as only one investigator provided instructions to the participants and conducted the tests. Therefore, this study should be regarded as the first step in establishing the reliability of the

IOPI. Although the validity and clinical relevance of these strength measures have yet to be established, ensuring that measures used in clinical studies are of high reliability allows for effective investigation of strength and its relationships to the functional demands of the individual. Finally, there are other devices to assess tongue strength on the market such as the MOST device [42], and the Kay Elemetrics Swallowing Workstation (Kay Elemetrics, Lincoln Park, NJ). These devices have been developed for evaluating the maximum force or pressure output at different locations on the tongue and can be used to help diagnose and strengthen weakened tongue muscles and the reliability of those devices needs to be established in a comparable manner to this investigation.

In summary, we have determined that the IOPI is reliable for the measurement of tongue and hand strength, but not endurance. A familiarization session is recommended to improve the precision of the assessment. Future studies should ensure that marking the IOPI connecting tube with tape or black marker pen once the lips were closed may be a strategy to further improve the within session placements and recording that length could possibly improve inter-session placements. Multiple attempts resulting in some consistency in the maximum values obtained should be provided to establish that a true representation of current maximal strength is obtained. Further investigation is required to determine the reliability of tongue and hand endurance measures using the IOPI.

References

1. Youmans S, Stierwalt JAG (2006) Measures of tongue function related to normal swallowing. Dysphagia 21 (2):102-111. doi:10.1007/s00455-006-9013-z
2. Crow HC, Ship JA (1996) Tongue strength and endurance in different aged adults. J Gerontol Med Sci 51A (5):M247-M250
3. Lazarus CL, Logemann JA, Pauloski BR, Rademaker AW, Larson CR, Mittal BB, Pierce M (2000) Swallowing and tongue function following treatment for oral and oropharyngeal cancer. J Speech Lang Hear R 43 (4):1011-1023
4. Nicosia MA, Hind JA, Roecker EB, Carnes M, Doyle J, Dengel GA, Robbins J (2000) Age effects on the temporal evolution of isometric and swallowing pressure. J Gerontol A Biol Sci Med Sci 55 (11):M634-M640. doi:10.1093/gerona/55.11.M634
5. Robbins J, Levine RL, Wood J, Roecker EB, Luschei ES (1995) Age effects on lingual pressure generation as a risk factor for dysphagia. J Gerontol A Biol Sci Med Sci 50A (5):M257-M262. doi:10.1093/gerona/50A.5.M257
6. Robin DA, Goel A, Somodi LB, Luschei ES (1992) Tongue strength and endurance: Relation to highly skilled movements. J Speech Hear Res 35 (6):1239-1245
7. Robin DA, Somodi LB, Luschei ES (1991) Measurement of tongue strength and endurance in normal and articulation disordered subjects. In: Moore CA, Yorkston KM, Beukelman DR (eds) Dysarthria and Apraxia of Speech: Perspectives on Management. Paul H Brookes Publishing Co, Baltimore, pp 173-184
8. Solomon NP, Robin DA, Luschei ES (2000) Strength, endurance, and stability of the tongue and hand in Parkinson disease. J Speech Lang Hear R 43 (1):256-267
9. Stierwalt JAG, Robin DA, Solomon NP, Weiss AL, Max JE (1996) Tongue Strength and Endurance: Relation to the Speaking Ability of Children and Adolescents following Traumatic Brain Injury. In: Robin DA, Yorkston K, Beukelman DR (eds) Disorders of Motor Speech: Assessment, Treatment, and Clinical Characterization. Brookes Publishing Co., Baltimore, pp 241-256
10. Stierwalt JAG, Youmans SR (2007) Tongue measures in individuals with normal and impaired swallowing. Am J Speech Lang Pathol 16 (2):148-156
11. Youmans SR, Stierwalt J, Clark HM (2002) Measures of tongue function in healthy adults. Paper presented at the American Speech-Language-Hearing Association, Atlanta,
12. IOPI Medical LLC (2008) Iowa Oral Performance Instrument: Users Manual. http://www.iopimedical.com/. Accessed Sep 17 2008
13. Steele CM, Bailey GL, Molfenter SM, Yeates EM (2009) Rationale for strength and skill goals in tongue resistance training: A review. Perspectives on Swallowing and Swallowing Disorders (Dysphagia) 18 (2):49-54. doi:10.1044/sasd18.2.49
14. Hopkins WG (2000) Measures of Reliability in Sports Medicine and Science. Sports Med 30 (1):1-15
15. Solomon NP, Lorell DM, Robin DA, Rodnitzky RL, Luschei ES (1995) Tongue strength and endurance in mild to moderate Parkinson's disease. Journal of Medical Speech-Language Pathology 3 (1):15-26
16. Solomon NP (2000) Changes in normal speech after fatiguing the tongue. Journal of Speech, Language & Hearing Research 43 (6):1416
17. Solomon NP, Robin DA, Luschei ES (2000) Strength, endurance, and stability of the tongue and hand in Parkinson Disease. Journal of Speech, Language, and Hearing Research 43 (1):256
18. Clark HM, Henson PA, Barber WD, Stierwalt JAG, Sherrill M (2003) Relationships among subjective and objective measures of tongue strength and oral phase swallowing impairments. American Journal of Speech-Language Pathology 12 (1):40

19. Lazarus CL, Logemann JA, Pauloski BR, Rademaker AW, Helenowski IB, Vonesh EF, MacCracken E, Mittal BB, Vokes EE, Haraf DJ (2007) Effects of radiotherapy with or without chemotherapy on tongue strength and swallowing in patients with oral cancer. Head & Neck 29 (7):632-637. doi:10.1002/hed.20577
20. Neel AT, Palmer PM, Gass CA (2008) Can IOPI be used to measure tongue pressure for speech sounds? Journal of Medical Speech-Language Pathology 16 (1):235-241
21. Solomon NP, Clark HM, Makashay MJ, Newman LA (2008) Assessment of orofacial strength in patients with dysarthria. Journal of Medical Speech-Language Pathology 16:251-258
22. Youmans S, Youmans G, Stierwalt JAG (2008) Differences in tongue strength across age and gender: Is there a diminished strength reserve? Dysphagia 24 (1):57-65. doi:10.1007/s00455-008-9171-2
23. Palmer P, Neel A, Sprouls G, Morrison L (2010) Swallow characteristics in patients with oculopharyngeal muscular dystrophy. Journal of Speech, Language and Hearing Research (Online) 53 (6):1567-1578
24. Neel AT, Palmer PM (2011) Is tongue strength an important influence on rate of articulation in diadochokinetic and reading tasks? J Speech Lang Hear R 55 (1):235-246. doi:10.1044/1092-4388(2011/10-0258)
25. Chang C, Chen S, Ko J-Y, Lin Y-H (2008) Early radiation effects on tongue function for patients with nasopharyngeal carcinoma: A preliminary study. Dysphagia 23 (2):193-198. doi:10.1007/s00455-007-9128-x
26. Gingrich LL, Stierwalt JAG, Hageman CF, LaPointe LL (2012) Lingual propulsive pressures across consistencies generated by the anteromedian and posteromedian tongue by healthy young adults. J Speech Lang Hear R 55 (3):960-972
27. Bohannon RW, Schaubert KL (2005) Test–retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders. J Hand Ther 18 (4):426-428. doi:10.1197/j.jht.2005.07.003
28. Solomon NP (2000) Changes in normal speech after fatiguing the tongue. J Speech Lang Hear R 43 (6):1416-1428
29. Solomon NP, Munson B (2004) The effect of jaw position on measures of tongue strength and endurance. J Speech Lang Hear R 47 (3):584-594
30. Hopkins WG (2011) A New View of Statistics: Precision of measurement. http://sportsci.org/resource/stats/precision.html. Accessed July 5 2012
31. Cohen J (1988) Statistical power analysis for the behavioral sciences. 2nd edn. Erlbaum, Mahwah, NJ
32. Hopkins WG (2006) A new view of statistics. http://www.sportsci.org/resource/stats/index.html. Accessed July 5 2012
33. Neel AT, Palmer PM (2011) Is tongue strength an important influence on rate of articulation in diadochokinetic and reading tasks? Journal of Speech, Language & Hearing Research:1092-4388_2011_1010-0258. doi:10.1044/1092-4388(2011/10-0258)
34. Solomon NP, Munson B (2004) The effect of jaw position on measures of tongue strength and endurance. Journal of Speech, Language & Hearing Research 47 (3):584-594
35. Palmer PM, Jaffe DM, McCulloch TM, Finnegan EM, Van Daele DJ, Luschei ES (2008) Quantitative contributions of the muscles of the tongue, floor-of-mouth, jaw, and velum to tongue-to-palate pressure generation. J Speech Lang Hear R 51 (4):828-835
36. Kays S, Hind J, Gangnon R, Robbins J (2010) Effects of dining on tongue endurance and swallowing-related outcomes. J Speech Lang Hear R 53 (4):898-907
37. Vitorino J (2010) Effect of age on tongue strength and endurance scores of healthy Portuguese speakers. Int J Speech Lang Path 12 (3):237-243. doi:10.3109/17549501003746160

38. Clark HM, Solomon NP (2012) Age and sex differences in orofacial strength. Dysphagia 27 (1):2-9. doi:10.1007/s00455-011-9328-2
39. Clark HM (2012) Specificity of training in the lingual musculature. J Speech Lang Hear R 55 (2):657-667. doi:10.1044/1092-4388(2011/11-0045)
40. Clark HM (2003) Neuromuscular treatments for speech and swallowing: A tutorial. Am J Speech Lang Pathol 12 (4):400
41. Clark HM, Henson PA, Barber WD, Stierwalt JAG, Sherrill M (2003) Relationships among subjective and objective measures of tongue strength and oral phase swallowing impairments. Am J Speech Lang Pathol 12 (1):40-50
42. Hewitt A, Hind J, Kays S, Nicosia M, Doyle J, Tompkins W, Gangnon R, Robbins J (2008) Standardized instrument for lingual pressure measurement. Dysphagia 23 (1):16-25. doi:10.1007/s00455-007-9089-0

Figures

Tables