Evaluation of Assessment Scales of Hypernasality: The Temple St. Scale of Nasality (TSS) vs. the Visual Analogue Scale (VAS) in terms of Inter-Rater Reliability in the Trained and Untrained Listener

Student Name: Martina Dwyer

Student Number: 12010189

Project Supervisor: Adjunct Professor Triona Sweeney

Journal for Submission: The Cleft Palate-Craniofacial Journal

Module Leader: Hilary O’ Neill

Word Count: 6,794

Thesis submitted in partial fulfilment of the requirements for MSc in Speech and Language Therapy

Friday 16th May 2014
Acknowledgements

The first author greatly appreciates the help and expert guidance received from Dr. Triona Sweeney, her supervisor for this project. The first author would like to express her sincere gratitude for this support. Many thanks also go to Professor Sue Franklin for her invaluable statistical advice. Sincere thanks to Trish Montgomery, Siobhán Coleman and Jean Brown, Senior Technical Officers and also to Aileen Wright for their technical assistance and generosity with the use of facilities during data collection. Sincere thanks are also due to the project participants for their kind permission to use their data and to give so generously of their time. In addition, many thanks to Clinic Administrators Geraldine Morrissey and Jacinta Shanahan for their assistance in the postal collection of data. Finally, sincere gratitude is expressed to the first author’s partner in data collection, Sarah Jane Osborne.
# Table of Contents

ABSTRACT ................................................................................................................................. 2  
INTRODUCTION .......................................................................................................................... 3  
METHODS .................................................................................................................................... 9  
RESULTS ...................................................................................................................................... 17  
DISCUSSION ............................................................................................................................... 19  
REFERENCES .............................................................................................................................. 24  
APPENDICES .............................................................................................................................. 32  
Appendix 1: The Temple St. Scale of Nasality and Nasal Airflow Errors (TSS) ......................... 33  
Appendix 2: Speech Sample ......................................................................................................... 38  
Appendix 3: Information Letter and Attached Consent Form for Trained Listeners .................. 40  
Appendix 4: Demographic Questionnaire for Trained Listeners .............................................. 44  
Appendix 5: Information Letter and Attached Consent Form for Untrained Listeners ............. 45  
Appendix 6: Definitions of Terms Provided at Training Course ............................................... 49  
Appendix 7: Instructions for Opening Encrypted USB Keys ....................................................... 50  
Appendix 8: Instructions for Completion of Rating Procedures ............................................... 52  
Appendix 9: TSS Survey Monkey Weblink and Screenshot ....................................................... 54  
Appendix 10: Visual Analogue Scale Form ............................................................................... 55  
Appendix 11: Temple St. Scale – Hypernasality (Paper Version for Students) ........................... 56  
Appendix 12: Debriefing Report ............................................................................................... 59
ABSTRACT

Background: Hypernasality is a distinctive speech problem often associated with cleft palate and velopharyngeal dysfunction (VPD; Sweeney 2011). Assessment of hypernasal speech is essential; speech has long been considered as one of the primary outcome measures of palatal surgery (Grunwell et al 1993; McWilliams et al 1990). The TSS provides a standardised speech sample and five point descriptive ordinal scale suitable for rating hypernasality. Clinicians wishing to utilize this tool may avail of a training course. It has been suggested that the standardised speech sample provided by the TSS might be utilized in conjunction with a visual analogue scale (VAS), and that the role in training in the use of both scales might be explored.

Objectives:

Aim 1: To compare the inter-rater reliability of the descriptive ordinal scale provided within the TSS with that of a VAS.

Aim 2: To compare inter-rater reliability for both scales for trained and untrained listeners.

Methods: Trained and untrained listeners rated the speech of 10 individuals with cleft palate and VPD, once according to the TSS and again 4 weeks later via a VAS.

Results: Intra-class correlation coefficients revealed very good levels of inter-rater reliability for both tools, with slightly higher levels of inter-rater reliability demonstrated by the VAS (Altman 1991). Trained listeners demonstrated higher levels of inter-rater reliability on the TSS than on the VAS, where they had been trained on the TSS and not the VAS. Untrained listeners displayed higher levels of inter-rater reliability than trained listeners on both scales. Untrained listeners were more reliable utilizing the VAS than when using the TSS.

Conclusions: Both tools can be used to rate hypernasality with confidence in terms of inter-rater reliability. The results of this pilot study give support to Cheng’s (2006) unpublished study demonstrating the success of a VAS in rating hypernasality. The role of training in enhancing the reliability of the rating of hypernasal speech remains unclear and requires further examination.

Keywords: cleft palate, hypernasality, perceptual assessment, descriptive scale, visual analogue scale
INTRODUCTION

In normal speech, the majority of consonants and vowels are produced when oral and nasal cavities are separate (Sell and Ma 1996). Such consonants and vowels are often referred to as oral consonants and vowels. Separation of the oral and nasal cavities results from upward and backward movements of the soft palate along with simultaneous medial movements of the lateral pharyngeal walls (Sell and Ma 1996). Additionally, occasionally, the posterior pharyngeal wall will move anteriorly, assisting in the closure of both cavities (Sell and Ma 1996). All such movements will allow oral consonants and vowels to be produced with oral airflow and oral tone, with appropriate intraoral pressure for consonant production (Sell and Ma 1996). On the other hand, nasality has been described as ‘the balance between oral and nasal resonance’ (Sweeney 2011, p. 199). Normal nasal resonance displays a range of acceptability and is perceived along a continuum (Peterson-Falzone et al 2009).

However, velopharyngeal dysfunction (VPD) can result in hypernasality, which might be defined as ‘the occurrence of excessive nasal resonance perceived during speech production’ (Sweeney 2011, p. 201). VPD occurs when the soft palate and lateral pharyngeal walls fail to shut off the nasal cavity from the oral cavity during the production of oral consonants and vowels (Sell and Ma 1996). VPD might occur as a result of velopharyngeal insufficiency, where anatomical or structural defects prevent adequate velopharyngeal closure e.g. cleft palate (Sweeney 2011; Trost-Cardamone 1989). VPD might also be caused by velopharyngeal incompetency, when a disorder of neurological or physiological aetiology results in poor movement of velopharyngeal structures e.g. dysarthria (Sweeney 2011; Trost-Cardamone 1989). VPD might also arise as a result of velopharyngeal mislearning of articulation patterns (Sweeney 2011; Trost-Cardamone 1989).

Assessment of hypernasal speech is vital. Speech has long been recognised as one of the primary outcome measures of palatal surgery in VPD (Grunwell et al 1993; McWilliams et al 1990). For example, in cleft lip and palate care, when speech difficulties are not addressed appropriately and in a timely manner, communication problems are likely to occur. This can lead to learning difficulties (Jocelyn et al 1996), social exclusion (Broder et al 1992) and adverse psychosocial adjustment and well-being (John et al 2006). In order to make any conclusions relating to modification of oral structures, the appropriateness of
speech intervention, or the planning and execution of therapeutic procedures, a close examination of speech production is indispensable (McWilliams et al. 1990). The identification and rating of hypernasality is therefore essential.

Investigations of hypernasal speech fall into two major categories; perceptual assessments and instrumental examinations. Perceptual assessments depend on the listener’s ability to perceive nasality during speech (Sweeney 2011). The perceptual assessment of hypernasal speech is considered the ‘gold’ standard, representing the most valid method of rating nasality (Smith and Kuehn 2007). Instrumental examinations of acoustic and aerodynamic aspects of hypernasal speech indirectly measure nasality, quantifying phenomena related to hypernasality (Sweeney 2011). Two widely used clinical tools for the instrumental assessment of hypernasality include the PERCI Speech Aerodynamic Research System (SARS; MicroTronics Inc. 1994) and the Nasometer (Kay Elemetrics Corporation 2003). The PERCI SARS is a computer software/hardware interface for assessment of speech aeromechanics, measuring oral pressure, nasal pressure, oral flow and nasal flow (Sweeney 2011). The Nasometer supplies an instrumental measure of nasality by measuring oral and nasal acoustic energy during speech production, thereby calculating a nasalance score (Sweeney et al. 2004). However, questions remain concerning the validity of such techniques in terms of measuring hypernasality. Indeed, Kuehn (1982, p. 518) highlights that, ‘in a sense, a speech disorder does not exist until it is perceived by a listener’. Therefore, currently, many key researchers in this area conclude that instrumental techniques might be considered indices of nasality, arguing that the final indicator of nasality will always come from the ear of the listener (Kuehn and Moller 2000; Moll 1964; Kanter 1948).

Therefore, the reliability of the perceptual judgement of hypernasality would appear highly significant. However, research to date has indicated considerable variation in the reliability of listener judgements of nasality (Sweeney 2000; Counihan and Cullinan 1970). The perceptual rating of nasality appears unreliable for a number of reasons.

Perceptual ratings of nasality are often unreliable as a result of the very nature of nasality itself. Hypernasality is often considered a multidimensional parameter of speech; listeners base ratings on three dimensions; including nasality, voice and pitch (Zraick et al.
Furthermore, it is relatively rare to observe hypernasality in isolation; hypernasality typically presents with co-existing disorders including hyponasality, nasal emission, articulatory/phonological disorder and voice disorder (Kent 1996; McWilliams et al. 1990; Whitehill 2010). Hypernasal speech is often judged as more nasal when observed alongside co-occurring articulation problems (McWilliams 1954). In addition, raters tend to confuse breathiness with hypernasality (Imatomi 2005). Therefore, perceptual overlap between hypernasality and other voice and speech categories is one likely cause of poor judgement reliability (Kent 1996).

However, in spite of its nature, many other factors influencing the reliability of perceptual ratings of hypernasality can be shaped. The reliability of the perceptual judgement of hypernasality is also often dependent on listening conditions (Sell et al. 2009). Listening conditions represent an important factor to address. The quality of the recording and listening environments is highly significant, as is the need to ensure consistency of the amplitude of speech samples (Gooch et al. 2001).

Perceptual ratings of nasality frequently vary according to the type of speech sample being analysed. Sweeney (2011) proposes the use of standardised speech samples for perceptual assessment of nasality. Such samples should include single word utterances, standardised sentences, and automatic and conversational speech, in order to maximise reliability (Sweeney 2011). Listeners demonstrate significantly poorer reliability for nasality judgements based on isolated vowels or syllables when compared with sentences (Counihan and Cullinan 1970; Spiestersbach and Powers 1959). Van Demark (1964) recorded a high correlation between a task of sentence repetition and spontaneous speech. In addition, an automatic speech sample, including counting or rote speech such as nursery rhymes, can be advantageous when obtaining a speech sample from younger children (Peterson-Falzone et al. 2009). Automatic speech is also valuable for obtaining ratings of vulnerable high-pressure targets and high and low vowels in connected speech (Sweeney 2011). Additionally, rating of conversational speech is crucial; this produces the most representative sample of a client’s speech performance (Kuehn and Moller 2000).

The reliability of the perceptual rating of hypernasality can depend upon the type of rating scale used. Descriptive scales and equal appearing interval (EAI) scales are often
preferred approaches to the perceptual assessment of nasality (Sell 2005). The reliability of a descriptive scale often depends on the extent to which such a scale is standardised and whether concise definitions of the scalar points are provided (Sweeney 2011). EAI requires listeners to assign each stimulus a number so as to place that stimulus along a linear partition of hypernasality (Schiavetti 1992). However, Whitehill et al (2002) highlight that EAI ratings of nasality display poor validity and reliability. To address the limitations of EAI scaling direct magnitude estimation (DME) has been proposed (Stevens and Galanter 1957). DME is a scale typically used with a standard, or reference stimulus, chosen as a good exemplar of “midrange” of the parameter being measured (Weismer and Laures 2002). Direct magnitude estimation (DME) has been shown to demonstrate good reliability in the assessment of hypernasality (Whitehill et al 2002; Zraick and Liss 2000). Whitehill et al (2002) therefore concluded that equal interval scaling may not be a valid method for the evaluation of hypernasality, and recommended the use of direct magnitude estimation in research studies. However, Folkins and Moon (1990) highlight the amount of work required to complete DME procedures in the clinic. Therefore, controversy continues regarding the most suitable rating scale for assessment of hypernasality.

Another factor influencing the reliability of perceptual ratings is that of listener variability. Listener factors include the listener’s particular internal standard for the dimension being judged, individual perceptual strategies or biases, overall sensitivity to the dimension being rated, listener fatigue, attention lapses and transcription errors (Lohmander et al 2009; Eadie and Baylor 2006; Kreiman et al 1993). Although the latter three are considered random, the first three factors can be shaped by training (Lee et al 2009). Training has proven to reduce variability among listeners’ perceptual judgements across speech parameters (Moller and Starr 1984; Bassich and Ludlow 1986).

The Temple St. Scale of Nasality and Nasal Airflow Errors (TSS) has been tested for validity, and acceptability and presents with moderate reliability (Sweeney 2000; Sweeney and Sell 2008; Sweeney and Fennell 2009; Appendix 1). The TSS has been developed as a clinical assessment of speech characteristics associated with cleft palate and VPD; it includes a perceptual speech assessment examining hypernasality, among other speech parameters (Fennell 2009). The TSS attempts to address factors influencing reliability threefold. It provides an appropriate speech sample, concise descriptive scales, and a training
programme for those who wish to utilize it. The TSS provides a standardised speech sample suitable for rating hypernasality; where spontaneous conversational speech, automatic rote speech and sentence repetition tasks are perceptually examined (Sweeney 2011; Sweeney 2000; see Appendix 2). Test sentences included in the sample are adapted from the Great Ormond St. Speech Assessment (GOS.SP.ASS; Sell et al 1999); an assessment for speech disorders associated with cleft palate and/or velopharyngeal dysfunction. However, unlike the TSS, the GOS.SP.ASS does not assess nasality and nasal airflow in detail. Sentences included are considered imageable, meaningful, relevant and easy to imitate (Sell et al 2009). Sentences have been selected to include phonemes vulnerable to velopharyngeal dysfunction, and contain approximately 11% nasal consonants, which is illustrative of the distribution of nasal consonants in normal conversational speech (Sweeney et al 2004). The TSS also provides a descriptive ordinal scale for assessing hypernasality, utilizing detailed definitions of terms and descriptive definitions of scalar points (Sweeney 2011; Sweeney 2000). Clinicians wishing to utilize the tool may avail of a training course, where definitions of terms are explained and audio demonstrations are provided (Sweeney and Fennell 2009).

As controversy regarding the use of rating scales in this area perseveres, it has been suggested that the standardised speech sample provided by the TSS might be utilized in conjunction with a visual analogue scale (VAS). Visual analogue (VA) scaling refers to the method in which observers place a mark, in proportion to the perceived magnitude of each stimulus, along a straight line (usually 10cm long) with fixed and predefined extremes of the quality being measured (Wewers and Lowe 1990). The ends of the scale are defined as extreme limits of the parameter being measured (Paul-Dauphin et al 1999). The scale is orientated from left to right, with the left end generally describing the worst aspects of the parameter being measured, and the right generally defining the best aspects of the parameter being measured (Paul-Dauphin et al 1999). Two studies have investigated the use of VAS in the rating of voice parameters. Wuyts et al (1999) compared a four-point ordinal scale with a VAS in the perceptual evaluation of dysphonia; the VAS demonstrated lower inter-rater reliability. When Yiu and Ng (2004) compared a VAS with EAI scaling in the perceptual rating of roughness and breathiness, similar inter-rater reliability values were discovered. An unpublished dissertation conducted by Cheng (2006) suggested that a VAS might display good reliability for the rating of nasality and nasal airflow errors. Raters
participating in Cheng’s (2006) study also reported that the VAS was ‘easier to use’ when compared with other rating scales. A dearth of published studies evaluating the use of VAS in the rating of hypernasality therefore promoted further investigation as paramount. The current study aimed to capitalise on this gap in the current literature. Specifically, this study presented with two aims.

Aim 1: To compare the inter-rater reliability of the descriptive scale provided within the TSS with that of a VAS.

Aim 2: To compare inter-rater reliability for both scales for trained and untrained listeners.
METHODS

Participants:

Thirty-one participants were recruited for the purpose of the study. Participants included eleven trained listeners and twenty untrained listeners.

Trained Listeners

Trained listeners were qualified SLTs who were recruited via the Dublin Cleft Centre and the Irish Association of Speech and Language Therapists (IASLT) Cleft Palate Special Interest Group (SIG). Trained listeners were advised that their participation in the project could be included in their yearly Continuing Professional Development (CPD) profile for the IASLT. For inclusion in the study, trained listeners were required to have completed a TSS training course. Trained participants received an information letter and attached consent form via email and were required to return this form via post to the University of Limerick, or as a scanned document via email to the first author (see Appendix 3). Four trained listeners ceased participation in the study following confirmation of consent, citing time pressure and a lack of appropriate listening equipment for reasons for drop out. As a result, data was collected for seven trained listeners only. Detailed information on trained listeners is presented in Table 1 (also see Appendix 4).
Table 1. Demographic information for trained listeners.

<table>
<thead>
<tr>
<th>Number of Years Qualified</th>
<th>Number of Years Since Completion of TSS Training</th>
<th>Number of Years Working with Clients with Hypernasality</th>
<th>Number of Clients with Hypernasality in the Year Prior to Listening Task</th>
<th>Setting*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>0 - 2</td>
<td>0 - 2</td>
<td>0 - 10</td>
<td>Dublin Cleft Centre</td>
</tr>
<tr>
<td>5 - 10</td>
<td>2 - 5</td>
<td>2 - 5</td>
<td>10 - 30</td>
<td>Community Care</td>
</tr>
<tr>
<td>Over</td>
<td>Over</td>
<td>Over</td>
<td>Over 50 - 100</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>N = 0</td>
<td>N = 2</td>
<td>N = 1</td>
<td>N = 0</td>
<td>N = 5</td>
</tr>
<tr>
<td>N = 3</td>
<td>N = 3</td>
<td>N = 2</td>
<td>N = 0</td>
<td>N = 1</td>
</tr>
<tr>
<td>N = 4</td>
<td>N = 2</td>
<td>N = 2</td>
<td>N = 2</td>
<td></td>
</tr>
<tr>
<td>*Data unavailable for one participant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Untrained Listeners

Untrained listeners were recruited through the MSc in SLT programme at University of Limerick. Untrained listeners were student SLTs who had not completed any TSS training. For inclusion in the study, untrained listeners were required to have no or little prior experience in the perceptual evaluation of hypernasality. Previous experience of hypernasality was limited to exposure to a few speech samples of hypernasality during a 1 hour phonetics lecture on resonance disorders and a 2 hour lecture on cleft lip and palate and associated craniofacial conditions. Naïve listeners were preferred since they had no formal exposure to pathological voices and lacked specific internal standards for judging hypernasality (Kreiman et al 1992). Students received an information letter and attached consent via email and were required to return this form to the first author (see Appendix 5).

Listener Training:

Several 1.5 day training courses took place between 2001 and 2013. Training courses were conducted by the project supervisor. Training outlined the notion of perceptual assessment and provided information on reliability and validity. Training also detailed common factors affecting the reliability and validity of perceptual assessment. SLTs were provided with definitions of speech parameters with audio demonstrations, helping them to distinguish hypernasality from other cleft-type speech parameters (see Appendix 6). During training, SLTs also became familiarized with the TSS assessment booklet, which comprises several descriptive scales for rating cleft-related speech parameters including the TSS hypernasality scale (see Appendix 1). The training course also provided clinicians with the opportunity to practice rating up to 12 cases in a quiet room with limited background noise. The speakers’ profiles were kept as similar as possible across groups. Raters listened to each sample twice and classified each level of speech according to the perceptual framework of the TSS assessment booklet.
Speech Samples:

Combined audio and video recordings of ten speakers with cleft lip and palate and velopharyngeal dysfunction were selected for the rating procedures. All speakers were Irish, attended Children’s University Hospital (CUH), and were unknown to the listeners. Speakers had been recorded for annual review and cleft care audit purposes, and informed consent had been obtained for all recordings to be utilized for research purposes. An equal balance of male and female speakers were included as in Cheng’s (2006) study. Speakers were excluded in cases of severe dyspraxia/dysarthria; learning disability (greater than mild); bilateral hearing loss above 45 dB; upper respiratory infection; moderate-severe dysphonia; mixed nasal resonance; and inability to complete assessment protocol due to limited co-operation. Speakers represented the full spectrum of severity in terms of hypernasality and presented with all forms of speech presentation associated with cleft lip and palate. The project supervisor assigned hypernasality ratings via the TSS hypernasality rating scale to each of the speakers (see Appendix 1). Such ratings are detailed in Table 2, alongside other demographic information for speakers.
Table 2. Demographic information for speakers.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Assigned TSS Hypernasality Rating</th>
<th>Gender</th>
<th>Age (Years : Months)</th>
<th>County of Origin</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>d</td>
<td>Female</td>
<td>5:03</td>
<td>Kerry</td>
<td>Unilateral Cleft Palate</td>
</tr>
<tr>
<td>2</td>
<td>Absent</td>
<td>Female</td>
<td>5:00</td>
<td>Louth</td>
<td>Cleft Palate</td>
</tr>
<tr>
<td>3</td>
<td>c</td>
<td>Male</td>
<td>5:09</td>
<td>Westmeath</td>
<td>Cleft Palate</td>
</tr>
<tr>
<td>4</td>
<td>a</td>
<td>Male</td>
<td>7:11</td>
<td>Kerry</td>
<td>Cleft Palate</td>
</tr>
<tr>
<td>5</td>
<td>d/e</td>
<td>Female</td>
<td>5:08</td>
<td>Dublin</td>
<td>Cleft Palate</td>
</tr>
<tr>
<td>6</td>
<td>e</td>
<td>Male</td>
<td>19:08</td>
<td>Limerick</td>
<td>Velopharyngeal Dysfunction</td>
</tr>
<tr>
<td>7</td>
<td>c</td>
<td>Male</td>
<td>5:09</td>
<td>Dublin</td>
<td>Unilateral Cleft Palate</td>
</tr>
<tr>
<td>8</td>
<td>Absent</td>
<td>Male</td>
<td>10:03</td>
<td>Cork</td>
<td>Bilateral Cleft Palate</td>
</tr>
<tr>
<td>9</td>
<td>b</td>
<td>Female</td>
<td>9:08</td>
<td>Westmeath</td>
<td>Submucous Cleft Palate</td>
</tr>
<tr>
<td>10</td>
<td>b</td>
<td>Female</td>
<td>11:05</td>
<td>Dublin</td>
<td>Submucous Cleft Palate</td>
</tr>
</tbody>
</table>

Samples were audio and visually recorded in a quiet room with the speaker facing natural light and the face and upper neck of the speaker framed within a Zoom Q3 HD camera. The rationale for the combined audio and video presentation evolved in the development of the Cleft Audit Protocol for Speech-Augmented (John et al 2006). Two of the recordings were made with the addition of an external microphone and amplifier; such recordings were made for cleft care audit purposes and therefore necessitated higher quality audio recording. Each of the speakers’ samples were anonymised, and renamed as Video 1, Video 2 and so on. Samples were then transferred onto encrypted USB keys, in random order.

The standardised speech sample provided by the TSS was utilized. This sample consisted of spontaneous speech, automatic rote speech and sentence repetition as detailed in Appendix 2.
Listening Tasks and Rating Procedures:

The listening tasks and rating procedures were split into two one hour listening sessions. The first listening session required listeners to rate each of the ten speakers according to the TSS hypernasality scale. During the second listening session, listeners were obliged to rate the same ten speakers, presented in the same order, via a VAS. A four week gap was left between the first and the second listening sessions in order to prevent the possibility of recall (Sell et al 2009).

Trained Ratings:

Encrypted USB keys were sent to trained listeners via registered post, with attached instructions for unlocking and opening such keys (see Appendix 7). Trained listeners were required to watch the samples on their own computers. Six of the seven trained listeners used Sennheiser 90 headphones and/or Genelec speakers to listening to the samples during both listening sessions. One participant did not describe the type of listening equipment utilized. For the first listening session, trained listeners received an email with an attached Microsoft Word file outlining the manner in which they were obliged to complete the rating procedures (see Appendix 8). This file included the password required to open the encrypted USB keys, a questionnaire relating to demographic information, and a Survey Monkey web link where clinicians were required to rate each video (or speaker) according to the TSS hypernasality scale (see Appendix 9). Four weeks later, for the second listening session, participants were required to rate the same samples on an A4 sized page with 10 visual analogue scales (see Appendix 10). For each video (or speaker), participants were presented with a 10cm line and were required to rate the sample along this line by making a mark with a pen or pencil to any position on the horizontal line according to its perceived magnitude of hypernasality. Trained listeners were obliged to return their VAS ratings to the University of Limerick via post, or as a scanned document via email to the first author.

Untrained Ratings:

Untrained listeners attended two listening sessions at a lecture hall at the University of Limerick, where they listened to and rated the same samples as had previously been rated by trained
listeners. At both sessions, an encrypted USB key containing speech samples was unlocked by the first author. Samples were then presented to students via an NEC NP-P401W Entry Level Installation 4000-Lumens LCD Projector with 16W Speaker. The lecture hall was quiet and background noise was limited by the first author. At the first session, students were required to rate samples according to the TSS via a Survey Monkey web link (see Appendix 9; which had been emailed to students prior to the first session) or via a paper version of the TSS (see Appendix 11). At the second session, four weeks later, students were required to rate the same samples according to the same VAS rating measure as presented to trained listeners. Paper versions of scales were collected by the first author at the end of both sessions.

**Data Management:**

All hard copy ratings were collected by the first author. Hard copies were anonymised and their respective data was transferred to SPSS Version 22 on a password-protected computer for statistical analysis. It is planned that such hard copies will be destroyed 3 years following completion of the study. Anonymised Survey Monkey data was also transferred to SPSS Version 22. All anonymised SPSS data will be deleted 5 years following the completion of the study. Trained listeners were requested to return encrypted USB keys to the University of Limerick via registered post, and where this was not possible, hand collection of such keys was arranged. Clinicians were requested to delete all video files from encrypted keys once they had completed the ratings. Where this was not completed, files were deleted by the first author. Encrypted keys were then stored in a locked filing cabinet at CUH.

**Data Analysis:**

Data was considered normal according to the Shapiro-Wilk test of normality for small sample sizes (Field 2009). Inter-rater reliability was considered twofold: first via intra-class correlations (ICCs) and secondly via measures of percentage agreement.
Inter-rater reliability was first calculated using intra-class correlations (ICC), which produce results comparable to weighted Kappa, thereby demonstrating the proportion of agreement corrected for chance (Fleiss and Cohen 1973; Streiner and Norman 2008). Inter-rater reliability was assessed using ICC coefficients for consistency and absolute agreement, constructed according to the methods outlined in Streiner and Norman (2008). The ICC for consistency among raters assessed whether raters were consistent in the order in which they placed individuals, and the ICC for absolute agreement determined whether raters agreed with each other with respect to the actual values they assigned individuals (Sell et al 2009). ICCs for consistency and absolute agreement were computed for both scales for all participants. Participants were then divided into trained (n = 7) and untrained (n = 20) groups and ICCs for consistency and absolute agreement were calculated for both scales for both groups. Inferring the mathematical equivalence between ICCs and the Kappa statistic (Fleiss and Cohen 1973), interpretation of the tabulated ICC values was based on the semantic categories adapted by Altman (1991) from Landis and Koch (1977), as shown in Table 3.

Table 3. Interpretation of intra-class correlation coefficients (ICCs).

<table>
<thead>
<tr>
<th>ICC</th>
<th>Strength of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.20</td>
<td>Poor</td>
</tr>
<tr>
<td>0.21–0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41–0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61–0.80</td>
<td>Good</td>
</tr>
<tr>
<td>0.81–1.00</td>
<td>Very good</td>
</tr>
</tbody>
</table>

Mean percentage agreement scores across all rater pairs for both scales were calculated to accompany ICCs. Again, participants were then split into trained (n = 7) and untrained (n = 20) groups and mean percentage agreement scores were calculated for both scales for both groups. TSS
ratings were considered to agree when such ratings were equal. VAS ratings were considered to agree when such ratings fell within one cm (100mm) of each other.

**Debriefing of Results**

Appropriate debriefing information sheets were provided to participants upon conclusion of the study (see Appendix 12).
RESULTS

ICCs for consistency and absolute agreement, along with measures of mean percentage agreement are presented for both aims below. 95% confidence intervals are reported for ICCs.

**Aim 1:** To compare the inter-rater reliability of the descriptive hypernasality scale provided within the TSS with that of a VAS

Table 4 demonstrates the inter-rater reliability values of both the TSS and the VAS for all 27 participants. ICC values were similar across the two different scales, and indicate very good agreement for both scales. The ICC for consistency among raters was very good for both scales, indicating very good agreement among raters in terms of the order in which they placed individuals. The ICC for absolute agreement among raters was also very good for both scales, indicating good reliability in terms of agreement with each other with respect to the actual values they assigned individuals. The VAS displayed slightly higher levels of agreement than the TSS for both ICCs. Mean percentage agreement scores indicated lower levels of agreement than ICC values for both scales. However, a similar trend was observed in that the VAS demonstrated a slightly higher level of mean percentage agreement than the TSS.

**Table 4. Inter-rater reliability scores for all participants.**

<table>
<thead>
<tr>
<th>Scaling Method</th>
<th>TSS</th>
<th>VAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC Consistency (95% Confidence Interval)</td>
<td>.985 (.967 - .995)</td>
<td>.986 (.969 - .996)</td>
</tr>
<tr>
<td>ICC Absolute Agreement (95% Confidence Interval)</td>
<td>.982 (.961 - .995)</td>
<td>.985 (.969 - .996)</td>
</tr>
<tr>
<td>Mean Percentage Agreement</td>
<td>38.2%</td>
<td>40.83%</td>
</tr>
</tbody>
</table>
**Aim 2:** To compare inter-rater reliability for both the TSS and VAS for trained and untrained listeners

Table 5 displays inter-rater reliability values of both the TSS and the VAS for both trained and untrained listeners. Again, ICC values were similar for both trained and untrained participants utilizing both scales, indicating very good agreement among raters on both scales for both groups. The ICC for consistency among raters was very good for both trained and untrained listeners on both scales, demonstrating very good agreement among raters in both groups in terms of the order in which they placed individuals. The ICC for absolute agreement among raters was also very good for both trained and untrained listeners on both scales, specifying good reliability in terms of agreement among raters in both groups with respect to the actual values they assigned individuals. Untrained listeners demonstrated higher ICC values than trained listeners on both scaling methods. Trained listeners demonstrated higher ICC values on the TSS than on the VAS. Untrained listeners displayed higher ICC values on the VAS than on the TSS. Mean percentage agreement scores demonstrated lower levels of agreement when compared with ICC values for both groups on both scales. Nevertheless, again, when compared with ICC values, similar trends were observed for mean percentage agreement scores. Untrained listeners demonstrated higher levels of mean percentage agreement than trained listeners on both scales.

**Table 5. Inter-rater reliability scores for trained and untrained participants.**

<table>
<thead>
<tr>
<th>Scaling Method</th>
<th>TSS</th>
<th>VAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trained</td>
<td>Untrained</td>
</tr>
<tr>
<td>ICC Consistency (95% Confidence Interval)</td>
<td>.950 (.881 - .985)</td>
<td>.983 (.962 - .995)</td>
</tr>
<tr>
<td>ICC Absolute Agreement (95% Confidence Interval)</td>
<td>.950 (.884 - .996)</td>
<td>.977 (.950 - .993)</td>
</tr>
<tr>
<td>Mean Percentage Agreement</td>
<td>36.19%</td>
<td>39.79%</td>
</tr>
</tbody>
</table>
DISCUSSION

Intra-class correlation coefficients (ICCs) from the current pilot study indicate that both the descriptive ordinal scale provided by the TSS and a VAS are highly reliable tools for the perceptual rating of hypernasality in terms of inter-rater reliability. Although the TSS clearly appeared advantaged in its ability to provide high levels of inter-rater reliability, the VAS proved slightly more reliable. While no previously published research has focused on the use of a VAS in hypernasality assessment, the current study lends support to Cheng’s (2006) unpublished study demonstrating the success of a VAS in rating hypernasality. It also supports a study by Yiu and Ng (2004) comparing EAI and VA scaling methods in the perceptual rating of roughness and breathiness, where comparable inter-rater reliability values were found. Nevertheless, it is also worth noting that averaging data across clinicians can weaken important aspects of individuals’ perceptions (Kreiman et al 1993) and therefore the results of this study should only be generalised to individual clinical judgments with caution (Metz et al 1990).

ICCs were similar for both trained and untrained participants utilizing both the TSS and the VAS, indicating very good agreement among raters in both groups on both scales. However, the role of training in enhancing the reliability of the rating of hypernasal speech remains unclear. Predictably, trained listeners demonstrated higher ICC values on the TSS than on the VAS, where they had been trained on the TSS and not the VAS. Brunnegard and Lohmander (2007) expressed concern when their study examining the speech of children with cleft palate displayed lower levels of inter-rater reliability for hypernasality when compared with other speech parameters. They argued the significance of clearly defined parameters and scalar points in the development of assessment tools for hypernasality. Like other associated tools such as the CAPS-A audit tool (John et al 2006), the TSS meets this recommendation, where listeners are trained in the application of its definitions and scalar points. This may explain why trained listeners were more reliable on the TSS when compared with the VAS, where no definitions of the scalar points on the VAS had been provided. This might also account for another finding of note, where a higher ICC value was demonstrated by untrained listeners on the VAS when compared with the ICC value illustrated by
the same listeners on the TSS. Untrained listeners had not been provided with definitions of the scalar points on the TSS and had not been trained in the application of their use. Furthermore, untrained listeners may have been more reliable when utilizing the VAS as a result of its ease to use. Cheng (2006) promoted VA scaling as a relatively simple scaling method to use, and therefore proposed VA scaling as a highly suitable means of evaluating hypernasality for clinical and research purposes.

However, unexpectedly, untrained listeners demonstrated higher ICC values than trained listeners on both scaling methods. This finding is clearly inconsistent with previous research underlining the significance of training and experience in perceptual judgment (Bassich and Ludlow 1986; Lewis et al 2003; Hayden and Klimacka 2000). This discrepancy may be explained by the fact that in many of these studies, experienced listeners had previously worked together (e.g. Lewis et al 2003), and they may have developed a ‘common ear’ over time (Hartelius et al 2003). While some of the trained listeners in our study worked together, others worked at isolated clinical sites, and this may have deflated levels of inter-rater reliability. The difference in inter-rater reliability scores between the two groups might also be due to the lack of standardisation of listening conditions for trained raters in the present study. Trained raters were required to watch and listen to the samples on their own computers, alone, in settings where listening conditions could not be controlled for. On the other hand, untrained raters listened to the samples in a quiet lecture hall where background noise was limited. In addition, untrained raters completed their judgements in close proximity to each other, and this may have inflated levels of inter-rater reliability. The difference in inter-rater reliability between trained and untrained listeners might also be accounted for when one examines the length of time elapsed since completion of TSS training for some trained listeners. Three of the trained listeners had completed their TSS training between 2 and 5 years ago, and two others had completed their training over 5 years ago. The necessity of refresher training in perceptual judgement is highlighted by John et al (2006, p. 279), who states that listeners need to be “re-calibrated at regular intervals”.

20
Nevertheless, in spite of the slight differences demonstrated between levels of inter-rater reliability levels between the trained and untrained groups, one cannot ignore the high ICC values displayed by both groups on both scales. It is essential to emphasise the worthy outcomes observed in the present study in terms of inter-rater reliability, and to stress the strengths of the methodology employed. Speech samples were obtained from numerous levels of production, and the speech material was constructed to limit the influence of surrounding consonants on target sounds. Video recordings were limited to ten to prevent listener fatigue (Stoel-Gammon 2001). Raters were blind to the speakers’ diagnoses and good quality recordings were ensured. Strict inclusionary and exclusionary criteria underlined the differences between trained and untrained listeners; trained listeners were only included having completed a TSS training course, and untrained listeners were ensured to lack formal exposure to pathological voices and in turn, specific internal standards for judging hypernasality (Kreiman et al 1992). All such factors are likely to have positively influenced levels of inter-rater reliability in the current study.

While mentioning the high ICC values obtained in this study, one cannot disregard the discrepancy between such values and those values obtained for mean percentage agreement. Although trends demonstrated by mean percentage agreement values were concurrent with trends illustrated by ICC values, the manner in which percentage agreement was calculated in this study might be considered as a limitation of this study. Strict definitions of agreement were considered in the present study. TSS ratings were considered to agree when only such ratings were equal. VAS ratings were considered to agree only when such ratings fell within one cm (100mm) of each other. One popular modification of percentage agreement calculations found in the testing literature involves extending the definition of agreement by including the adjacent scoring categories on a rating scale, and if employed in the current study, this technique might have resulted in mean percentage agreement values equivalent to those values found via ICC statistical analyses (Stemler 2004).

The manner in which mean percentage agreement values were computed is not the only limitation to the current study. As with all research, a number of limitations to this study exist, and
as a result, all findings presented in this analysis should be treated with caution. The nature of the speech samples selected may have influenced inter-rater reliability values. Selection of speech samples for the current study was based mainly on the degree of hypernasality and co-existing speech and voice disorders were not controlled for. Therefore, concomitant speech and voice disorders and perceptual overlap may have complicated inter-rater reliability values (Imatomi 2005; Kent 1996). In Cheng’s (2006) study, while co-morbid voice disorders existed within hypernasality samples, hyponasality and articulation difficulties were controlled for. In addition, Cheng (2006) proposed the use of synthetic speech samples in research studies examining the perceptual assessment of hypernasality. However, again, it is also important to note that it is relatively rare to observe hypernasality in isolation (Kent 1996; McWilliams et al 1990; Whitehill 2010), and therefore synthetic speech samples might not be representative of real-life clinical cases.

The length of spontaneous speech samples was also considered a limitation to the current study. In many instances, particularly for younger speakers, listeners commented that spontaneous speech samples were too short to make decisive ratings. While it is possible to minimize listener and speaker variables, it is not always possible to ensure elicitation of comparable quantities of fluent, sentence based, normal volume speech during conversation with an unfamiliar adult in front of a camera at age 5 (Sell et al 2009), and younger speakers made up the majority of the speech samples in the current study. Sell et al (2009) propose that clinician responses are neutral in this task when making recordings, so that ratings are not coloured by a knowledge of the intended meaning. This is, however, hard to attain considering that conversational flow is enhanced by confirmation of understanding by the clinician (Sell et al 2009). Sell et al (2009) suggest that clinicians recording samples use open-ended questions through a progression of questions on a particular topic such as, ‘What did you do for your fifth birthday’, ‘Have you been to any good birthday parties’ ‘What did you do there?’ Alternatively, questions about school, after school activities and story retelling may be helpful, e.g. ‘Have you seen the new Toy Story film?’ ‘What happens to Woody?"

Further methodological criticisms of the current study might also be considered when drawing conclusions from the analysis. The design of the demographics questionnaire given to
trained listeners demonstrated some methodological flaws, where overlap existed between response options. Furthermore, some trained listeners commented that the design of the VAS form had impacted on their ratings. The presentation of all ten VA scales on one A4 page was found to complicate ratings where one rating was influenced by its proximity to another rating. Moreover, no audiological screening of participants was carried out as in Cheng (2006) or Lee et al (2009) where all listeners had normal hearing as defined by passing a pure-tone audiological screening at 20 dB HL at the octave frequencies of 250–4000 Hz for the better ear. Whilst the possibility of recall between listening sessions was reduced by leaving a four week gap between both sessions in the current study, speech samples were presented in the same order at both sessions, and therefore the possibility of recall may not have been completely eliminated.

In terms of further research, there are a number of avenues that might be explored in the future. As discussed, the issue of eliciting adequate spontaneous speech samples, especially from younger speakers, clearly necessitates further study. The comparison of intra-rater reliability between the TSS hypernasality scale and a VAS also represents a gap in the current literature that might be filled. All ratings in this study were anonymized, which meant that intra-rater reliability analyses could not be carried out. However, the most significant avenue that might be further examined lies in the role of training in the use of both scales, particularly via larger sample sizes. The sample in the current study was small, and the sample of trained listeners was considered particularly small. Recruitment of trained listeners for this study was limited by strict inclusionary and exclusionary criteria for trained listeners, as well as with a high dropout rate. Therefore, interpretation of the discrepancy between trained and untrained listeners in the current study should be tentative in nature. The main conclusion to be drawn from the present study suggests that both the TSS hypernasality scale and the VAS utilized in this study can be used to rate hypernasality with confidence in terms of inter-rater reliability. This finding, along with Fennell’s (2009) recommendation to develop the TSS as a nation-wide tool, should promote training in the tool, which in turn should provide larger pools of trained listeners for future research studies.
REFERENCES


APPENDICES
Appendix 1: The Temple St. Scale of Nasality and Nasal Airflow Errors (TSS)
TEMPLE STREET SCALE OF NASALITY AND NASAL AIRFLOW ERRORS

NAME

AGE DOA

CLEFT TYPE

TESTER

<table>
<thead>
<tr>
<th>Audio tape</th>
<th>Video tape</th>
<th>Nasometry</th>
<th>PERCI SARS</th>
</tr>
</thead>
</table>

AV CONSENT OBTAINED:
<table>
<thead>
<tr>
<th>Test sentences (Adapted from GOS.SP.ASS 1996)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/m/ Mary came home early</td>
</tr>
<tr>
<td>/p/ The puppy is playing with a rope</td>
</tr>
<tr>
<td>/b/ Bob is a baby boy</td>
</tr>
<tr>
<td>/l/ The phone fell off the shelf</td>
</tr>
<tr>
<td>/v/ Dave is driving a van</td>
</tr>
<tr>
<td>/b/ This hand is cleaner than the other</td>
</tr>
<tr>
<td>/n/ Neil saw a robin in a nest</td>
</tr>
<tr>
<td>/b/ A ball is like a balloon</td>
</tr>
<tr>
<td>/l/ Tim is putting on a hat</td>
</tr>
<tr>
<td>/d/ Daddy mended the door</td>
</tr>
<tr>
<td>/i/ I saw Sam sitting on a bus</td>
</tr>
<tr>
<td>/j/ The zebra lives at the zoo</td>
</tr>
<tr>
<td>/sh/ Sean is washing a dirty dish</td>
</tr>
<tr>
<td>/ch/ Charlie’s watching a football match</td>
</tr>
<tr>
<td>/y/ John’s got a magic badge</td>
</tr>
<tr>
<td>/l/ The young chicks are yellow</td>
</tr>
<tr>
<td>/ng/ The bell’s ringing</td>
</tr>
<tr>
<td>/d/ Karen is making a cake</td>
</tr>
<tr>
<td>/g/ Gary’s got a bag of lego</td>
</tr>
<tr>
<td>/u/ Hannah hurt her hand</td>
</tr>
<tr>
<td>/s/ Stuarts hamster scrambled up his sleeve</td>
</tr>
<tr>
<td>LP We were away all year</td>
</tr>
<tr>
<td>LP Laura will wear a yellow welly</td>
</tr>
</tbody>
</table>
### Automatic Speech

| _Jack & Jill/Humpty Dumpty_ | Counting: 1 - 20, 60 - 70 |

### Conversational speech

#### Temple Street Scale - Nasality

<table>
<thead>
<tr>
<th>Hypernasality</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) mild, evident but acceptable.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) mild/moderate, unacceptable distortion, evident on close vowels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) moderate, evident on close and open vowels.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) moderate/severe, evident on all vowels and some consonants.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) severe, evident on all vowels and most voiced consonants.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Hypernasality: Consistent | Inconsistent

<table>
<thead>
<tr>
<th>Hyponasality</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) mild - evident, but acceptable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) moderate - all vowels reduced nasality.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) severe - total denasal production of nasal consonants.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Hypernasality: Consistent | Inconsistent

<table>
<thead>
<tr>
<th>Cul de Sac</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
</table>

#### Temple Street Scale - Nasal Airflow

<table>
<thead>
<tr>
<th>Nasal Emission</th>
<th>Frequent</th>
<th>Infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>Inconsistent</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nasal Turbulence</th>
<th>Frequent</th>
<th>Infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>Inconsistent</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nasal/Velopharyngeal Fricative</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>Inconsistent</td>
<td></td>
</tr>
</tbody>
</table>

#### Mirror Test: R /p t k/ L /p t k/


### INSTRUMENTAL SCORES

<table>
<thead>
<tr>
<th>Pressure/Flow:</th>
<th>pa</th>
<th>mp</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF</td>
<td>VPA</td>
<td>NF</td>
</tr>
<tr>
<td>HF</td>
<td>LP</td>
<td>Nasal</td>
</tr>
</tbody>
</table>

### CLEFT TYPE SPEECH CHARACTERISTICS:

(Adapted from GOS.SPASS, Sell, Harding & Grunwell 1997)

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Dentalisation.................................</td>
</tr>
<tr>
<td>1</td>
<td>Lateralization.................................</td>
</tr>
<tr>
<td>2</td>
<td>Palatalization.................................</td>
</tr>
<tr>
<td>3</td>
<td>Double articulation............................</td>
</tr>
<tr>
<td>4</td>
<td>Backing to velar................................</td>
</tr>
<tr>
<td>5</td>
<td>Backing to uvular...............................</td>
</tr>
<tr>
<td>6</td>
<td>Pharyngeal articulation..........................</td>
</tr>
<tr>
<td>7</td>
<td>Glottal articulation............................</td>
</tr>
<tr>
<td>8</td>
<td>Active Nasal fricative............................</td>
</tr>
<tr>
<td>9</td>
<td>Weak/nasalization..............................</td>
</tr>
<tr>
<td>10</td>
<td>Nasal realisations of plosive/fricatives.......</td>
</tr>
<tr>
<td>11</td>
<td>Gliding of fricatives/affricatives............</td>
</tr>
</tbody>
</table>
## Consonant Production

| m | p | b | f | v | n | t | d | s | ʈ | t̪ | d̪ | k | g | h | w | l | r | j |
| SI |
| SF |

## Imitation/Stimulability

| m | p | b | f | v | n | t | d | s | ʈ | t̪ | d̪ | k | g | h | w | l | r | j |

Developmental Errors: .................................................................

- **Voice**
  - Normal
  - Dysphonic
  - Reduced Volume

- **Hearing**
  - Normal
  - Loss

- **Language**
  - Normal
  - Delayed
  - Disordered

### Orofacial

(Adapted from GOS. SPASS, Sell, Harding & Grunwell '97)

1. **nose**: NAD deviated septum obstructed alar abn
2. **lips**: NAD restricted movement open posture
3. **occlusion**: I II III open bite
4. **dentition**: NAD supermuneracy missing teeth malaligned
5. **tongue**: mobility posture humping tie
6. **palatal fistula**: present absent
7. **fistula size**: 1 < 2mm 2 bet. 2.5mm 3 bet 5-8mm 4 > 8mm 5 complete breakdown
8. **palatal structure**: .................................................................
9. **nasopharynx**
   - tonsils .................................................................
   - deep pharynx ..............................................................
   - pharyngeal wall movement ..............................................
   - pharyngoplasty .............................................................

### Surgical Status:

### Speech & Language Therapy:

- 0 Not indicated
- 1 therapy ongoing
- 2 regular review
- 3 group
- 4 unavailable
- 5 short term
- 6 no uptake
- 7 discharge

### Influencing Factors

### Recommendations:

© Triona Sweeney Phd., The Children's University Hospital, Temple Street, Dublin.
Appendix 2: Speech Sample

1. Spontaneous Speech:

- Answers to questions exploring school, summer camp, birthday, video, cinema, an outing etc.

2. Automatic Rote Speech:

- Counting from 1 to 20 and from 60 to 70.
- Counting from 1 to 20 and from 60 to 70.
- Repeating days of the week, months of the year.

3. GOS.SP.ASS Sentence Repetition Task (Sell et al 1999):

Sentences were spoken in a natural way by the clinicians conducting the review, without exaggeration and not as a string of individually enunciated words. Effort was also made to retain the word order to ensure that the release of word final consonants was heard.

Sentences:

/m/ Mary came home early.
/p/ The puppy is playing with a rope.
/b/ Bob is a baby boy.
/t/ The phone fell off the shelf.
/v/ Dave is driving a van.
/ð/ This hand is cleaner than the other.
/n/ Neil saw a robin in a nest.
/l/ A ball is like a balloon.
/t/ Tim’s putting a hat on.
/d/ Daddy mended a door.
/s/ I saw Sam sitting on a bus.
<table>
<thead>
<tr>
<th>Sound</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>/z/</td>
<td>The zebra was at the zoo.</td>
</tr>
<tr>
<td>/s/</td>
<td>Sean is washing a dirty dish.</td>
</tr>
<tr>
<td>/ts/</td>
<td>Charlie’s watching a football match.</td>
</tr>
<tr>
<td>/dz/</td>
<td>John’s got a magic badge.</td>
</tr>
<tr>
<td>/n/</td>
<td>The bell’s ringing.</td>
</tr>
<tr>
<td>/k/</td>
<td>Karen is making a cake.</td>
</tr>
<tr>
<td>/g/</td>
<td>Gary’s got a bag of Lego.</td>
</tr>
<tr>
<td>/h/</td>
<td>Hannah hurt her hand.</td>
</tr>
</tbody>
</table>

/s/ clusters
Stuart’s hamster scrambled up his sleeve.

Low-pressure Sentence 1: We were away all year.
Low-pressure Sentence 2: Laura will wear a yellow welly.
Appendix 3: Information Letter and Attached Consent Form for Trained Listeners

To Whom It May Concern:

We are MSc in Speech and Language Therapy students from the University of Limerick conducting our final year research project on the reliability of hypernasality ratings in general cleft assessments. This is a question which has been raised as a result of Cleft Audit Protocol for Speech – Augmented (CAPS-A) and Americleft training. We are now eagerly looking for SLTs to participate in the study. The project will be carried out under the supervision of Professor Triona Sweeney, Clinical Specialist/Adjunct Professor at Temple Street Children’s University Hospital.

Participation is completely voluntary, and you may withdraw your participation from the study at any time without consequence. We appreciate that you are all very busy but the benefits of participation in this study far outweigh any negative consequences. Your input would be of great value to this project and the process can be included in your yearly CPD profile for IASLT.

Although it is now acknowledged that perceptual assessment is the gold standard for assessment of speech disorders related to cleft palate and velopharyngeal dysfunction (Kuehn and Moller 2000), it is also acknowledged there are confounding problems with this approach (Sell et al 1999; Kent 1996; John et al 2006). For the perceptual assessment of nasality one area of controversy is the type of rating scale used. Recently, Whitehill (2010) reported good reliability for a visual analogue scale (VAS), and although this approach shows promise for research, the VAS can make clinical reporting of nasality and nasal airflow errors difficult; some members of the multi-disciplinary team may not understand the ratings. A descriptive scale for assessing nasality and nasal airflow, the Temple Street Scale (TSS), has been tested for validity, reliability and acceptability (Sweeney 2000; Sweeney and Sell 2008; Sweeney and Fennell 2009). This approach assesses nasality and nasal airflow errors, using detailed definitions of terms and descriptive definitions of scalar points. The aims of this project will be:

1) to compare inter-rater reliability for assessment of hypernasality for 10 patients using the Temple Street Scale and a Visual Analogue Scale and

2) to compare inter-rater reliability on both scales for trained and untrained listeners

To carry out this project, you will be provided with an encrypted memory key containing 10 video samples. You will be asked to watch and listen to these samples on a computer and will be required to provide ratings of hypernasality only. You will be required to complete TSS ratings on such samples on one occasion and to then complete the VAS ratings on the same 10 samples one month later. Both sets of ratings should take about 1 hour. You will then be required to post your ratings to
Appendix 3: Information Letter and Attached Consent Form for Trained Listeners

Children’s University Hospital, Temple Street, Dublin 1, Ireland.
Ospidéal Ollscolaíoch na Leanai, Sráid Temple, Baile Átha Cliath.
Telephone: 01-8784200   Fax: 01-8748355   Web: www.cuh.ie

Professor Triona Sweeney at Temple St. University Hospital for analysis. The results will be analysed using Kappa scores to assess reliability. Mean percentage agreement score for the above will also be calculated.

All information will remain entirely confidential and anonymous. To ensure confidentiality, all participants will be required to return encrypted memory keys (via hand or registered post) at the end of the study so that their contents can be destroyed. Although information may be published in an article or presented at an academic or conference in the future, nothing of a personal or distinguishing nature will be included.

If you agree to participate in this study, please sign the attached consent form and return via post or email by the 20th January. Please see contact details below. Please do not hesitate to contact either of us if you have any query regarding the study.

Many thanks for taking the time out to read this and we hope to hear from you soon.

Yours sincerely,

Sarah Jane Osborne         Martina Dwyer         Professor Triona Sweeney

_________________  ___________________  ___________________
Appendix 3: Information Letter and Attached Consent Form for Trained Listeners

Children’s University Hospital, Temple Street, Dublin 1, Ireland.
Ospidéal Ollscolaíoch na Leanáí, Sráid Temple, Baile Átha Cliath.
Telephone: 01-8784200  Fax: 01-8748355  Web: www.cuh.ie

Contact Details

Post:
Martina Dwyer c/o Geraldine Morrissey,
UL Speech and Language Therapy Clinic,
Clinical Therapies Department,
University of Limerick,
Castletroy,
Limerick.

Research Assistant 1: Sarah Jane Osborne
Email Address: sosbornes@gmail.com

Research Assistant 2: Martina Dwyer
Email Address: martinadwyer4@gmail.com

Principal Investigator: Professor Triona Sweeney
Email Addresses (please contact both):
Triona.Sweeney@cu.ie
sweeneyt@eircom.net
Consent Form

Research Study: To evaluate and compare the assessment of hypernasality, comparing the use of the Temple Street Scale of Nasality and Nasal Airflow with the use of a Visual Analogue Scale for rater-reliability in both the trained and untrained listener.

- I have read and understand the attached information sheet and by signing below, I consent to participate in this study
- I understand the aims of this study
- I understand that all information obtained will be kept confidential
- I understand that I also have the right to change my mind about participating in the study for a short period after the study has concluded
- I agree to keep all patient videos completely confidential and to only use the encrypted memory key for viewing samples (i.e. files will not be copied)
- I agree to return encrypted memory keys on project completion to ensure destruction

Date: _____________________________________________________________________________

Signed: ___________________________________________________________________________

Print Name (Block Capitals): _________________________________________________________

Children’s University Hospital, Temple Street, Dublin 1, Ireland.
Ospidéal Ollscolaíoch na Leanai, Sráid Temple, Baile Átha Cliath.
Telephone: 01-8784200   Fax: 01-8748355   Web: www.cuh.ie
Appendix 4: Demographic Questionnaire for Trained Listeners*

1. How long have you been qualified as an SLT?
   - 0 – 2 years
   - 2 – 5 years
   - Over 5 years
   - Over 10 years

2. How long have you been trained in the Temple St. Scale of Nasality and Nasal Airflow Errors?
   - 0 – 2 years
   - 2 – 5 years
   - Over 5 years

3. How long have you been working with children with hypernasality?
   - 0 – 2 years
   - 2 – 5 years
   - Over 5 years
   - Over 10 years

4. How many patients did you have on your caseload last year with hypernasality?
   - Less than 10 patients
   - Less than 30 patients
   - Less than 50 patients
   - Less than 100 patients
   - Over 100 patients

5. In which of the following settings do you work?
   - Community Care
   - Dublin Cleft Centre

6. Which of the following audio equipment are you using to rate the samples?
   - Seinheiser 90 headphones or equivalent
   - Genelec speakers
   - Other audio equipment

If other audio equipment is used, please indicate the model chosen:

_________________________

*such information was anonymised and collected via the following Survey Monkey web link: https://www.surveymonkey.com/s/7RNBTDD
Appendix 5: Information Letter and Attached Consent Form for Untrained Listeners

To Whom It May Concern;

We are MSc in Speech and Language Therapy students from the University of Limerick conducting our final year research project on the reliability of hypernasality ratings in general cleft assessments. This is a question that has been raised as a result of Cleft Audit Cleft Audit Protocol for Speech – Augmented (CAPS-A) and Americleft training. We are now eagerly looking for students to participate in the study. The project will be carried out under the supervision of Professor Triona Sweeney, Clinical Specialist/Adjunct Professor at Temple Street Children’s University Hospital.

Participation is completely voluntary, and you may withdraw your participation from the study at any time without consequence. We appreciate that you are all very busy but the benefits of participation in this study far outweigh any negative consequences. Your input would be of great value to this project and we really appreciate it. It will involve one hour of rating on the 27\textsuperscript{th} of February at 12am and another hour of rating approximately four weeks later (date to be decided).

Although it is now acknowledged that the perceptual assessment is the gold standard for assessment of speech disorders related to cleft palate and velopharyngeal dysfunction (Kuehn and Moller 2000), it is also acknowledged there are confounding problems with this approach (Sell et al. 1999, Kent 1996, John et al. 2006). For the perceptual assessment of nasality one area of controversy is the type of rating scale used. More recently, Whitehill (2010) reported good reliability for a visual analogue scale (VAS), and although this approach shows promise for research, the VAS can make clinical reporting of nasality and nasal airflow errors difficult; all members of the multidisciplinary team may not understand the ratings. A descriptive scale for assessing nasality and nasal airflow, the Temple Street Scale (TSS), has been tested for validity, reliability and acceptability (Sweeney, 2000; Sweeney and Sell, 2008; Sweeney & Fennell, 2009). This approach assesses nasality and nasal airflow errors, using detailed definitions of terms and descriptive definitions of scalar points.
Appendix 5: Information Letter and Attached Consent Form for Untrained Listeners

The aims of this project will be:

1) to compare inter-rater reliability for assessment of hypernasality for 10 patients using the Temple Street Scale and a Visual Analogue Scale and

2) to compare inter-rater reliability on both scales for untrained and trained listeners

To carry out this project, you will be asked to watch and listen to these samples on an overhead projector and will be required to provide ratings of hypernasality only. You will be required to complete TSS ratings on one occasion and to then complete VAS ratings on the same 10 samples four weeks later. Both sets of ratings should take about one hour. The results will be analyzed using Kappa and Intra-Class Correlation scores to assess reliability. Mean percentage agreement scores for the above will also be calculated. Mean reliability scores will be compared and analysed for the two rating scales.

All information will remain entirely confidential and anonymous. We will be using encrypted keys to show you the videos and all of the videos will be deleted at the end of the project by myself, Sarah Jane Osborne, and Martina Dwyer.

If you agree to participate in this study, please sign the attached consent form.

Yours Sincerely,

Sarah Jane Osborne           Martina Dwyer           Prof. Triona Sweeney
Contact Details

Research Assistant 1: Sarah Jane Osborne
Email: sosbornes@gmail.com

Research Assistant 2: Martina Dwyer
Email Address: martinadwyer4@gmail.com

Principal Investigator: Prof. Triona Sweeney
Email Addresses (please contact both):
Triona.Sweeney@cuh.ie
sweeneyt@eircom.net
Appendix 5: Information Letter and Attached Consent Form for Untrained Listeners

Research Study: To evaluate and compare the assessment of hypernasality, in The Temple Street Scale of Nasality and Nasal Airflow and using a Visual Analogue Scale for rater reliability in both the trained and untrained listener.

- I have read and understand the attached information sheet, and by signing below, I consent to participate in this study.
- I understand the aims of the study
- I understand that all information obtained will be kept confidential
- I understand that I also have the right to change my mind about participating in the study for a short period after the study has concluded.

Date: __________________________
Signed: ______________________________

Print Name (Block Capitals): _______________________________
Appendix 6: Definitions of Terms Provided at Training Course

1. HYPERNASALITY: The occurrence of excessive nasal resonance perceived during speech production resulting from the coupling of oral and nasal resonating cavities, where there is an increase in the ratio of nasopharyngeal port to oropharyngeal port size.

2. HYPONASALITY: The reduction/absence of expected nasal resonance associated with nasal consonants and adjacent vowels, where there is a decrease in the ratio of nasopharyngeal port to oropharyngeal port size.

3. NASAL AIRFLOW: The inappropriate escape of air through the nasal cavity during production of voiced and voiceless oral pressure consonants associated with incomplete velopharyngeal closure/palatal fistula.

4. NASAL EMISSION: Audible escape of air from the nasal cavity accompanying production of oral pressure consonants. Airflow is constricted within the nasal cavity, therefore it is audible. It has a frictional but no turbulent or snorting quality.

5. NASAL FRICATIVE: The frictional sound produced by air passing through nasal cavity when velopharyngeal sphincter is open. Complete or almost complete stricture in the oral cavity results in no audible oral release. The nasal fricative replaces a consonant, unlike nasal emission which accompanies the consonant.

6. NASAL TURBULENCE: A snorting or turbulent noise accompanying a sound. Approximation but inadequate closure of the superior border of the velum and the posterior pharyngeal wall. Results from friction produced when an airstream passes through a small velopharyngeal gap.

7. VELOPHARYNGEAL FRICATIVE: A snorting or turbulent noise which replaces another consonant. Approximation but inadequate closure of the superior border of the velum and the posterior pharyngeal wall. There is complete or almost complete oral stricture during sound production with no audible oral release.

8. FREQUENT Vs. INFREQUENT: Frequent nasal airflow indicates >10% of target phonemes with nasal airflow. Infrequent nasal airflow indicates up to 10% of target phonemes with nasal airflow.

9. CONSISTENT: If the degree/frequency of occurrence is the same across all levels of speech (words, sentences, automatic, conversational speech).

10. INCONSISTENT: If there is a difference in the degree/ frequency of nasal airflow errors across the different speech levels.
Appendix 7: Instructions for Opening Encrypted USB Keys

To all participants,

Many thanks for agreeing to participate in our research study on Hypernasality, comparing the use of the Temple Street Scale of Nasality and Nasal Airflow with the use of a Visual Analogue Scale for rater-reliability in both the trained and untrained listener.

Please note you should receive your encrypted keys in the next day or so with this letter of instruction on how to open the keys. Please see previous email for the password and below for instructions of use. You will have received only one of the below keys, either the DataTraveler or the SanDisk. Please refer only to the appropriate instructions.

Please ensure not to copy the files to your own computer. Collection of the keys will be organised at the end of the project to ensure all files have been deleted, where this is not possible you may be asked to return via registered post.

INSTRUCTIONS FOR OPENING KEYS:

**Data Traveler (Silver Key)**

1. Open the key
2. Run DTLplus_launcher.exp
3. Data Traveler icon will appear (this may take a minute or two to load)
4. Enter password (as per e-mail).
5. Open folder to view files.

(Please note if your key doesn’t load as above go into my computer; two drives will appear, one is the Launcher (shield icon) and the second contains the files, you need to use the Launcher first to enter the password, then you may need to go back into my computer and open up the files which our stored under a different drive)

**SanDisk (Slimline Shiny Keys)**

1. Open the key.
2. Click on RunSanDiskSecureAccess_Win.exe (this should be the last file on the drop down menu)
Appendix 7: Instructions for Opening Encrypted USB Keys

3. “Skip” registration.

4. Enter password (as per e-mail)

5. Click OK twice to close irrelevant windows.

6. You should see the files in the remaining open window.

If you have any issues opening the memory keys or any other queries, please do not hesitate to contact us. Thank you for your continued participation in this project.

Yours sincerely,

Sarah Jane Osborne  Martina Dwyer  Professor Triona Sweeney

____________________  __________________  __________________
To all participants,

Thank you for your continued participation in our study on hypernasality. You will be receiving your encrypted key, with video files that will need to be rated in the next day or so. The password for opening the key is:

******

Please see accompanying letter for instructions on opening the video files.

This is Part 1 of a two part project. This email will outline instructions for completion of Part 1. Part 1 will include some demographic questions as well as the first set of hypernasality ratings (Temple St. Scale ratings). Details pertaining to Part 2 (i.e. second set of ratings via a visual analogue scale) will be imparted to participants in three weeks’ time. Please allow approximately one to one and a half hours to complete Part 1.

Firstly please follow the link below to complete some demographic questions on Survey Monkey, choosing the most appropriate answer in the drop down menu,

https://www.surveymonkey.com/s/7RNBTDD

(If clicking on this link does not lead you to the survey, try copying and pasting the link into the search bar on your internet browser).

Next you will be required to watch the videos and complete the Temple St. Scale ratings. Please ensure not to copy the video files to your own computer. You need to watch and listen to the videos in numerical order and rate each one immediately after. You will need to click on the following link to complete the Temple St. Ratings:

https://www.surveymonkey.com/s/7XSL8L3

For example, to rate Video 1 you will need to open the video on the memory key (using the attached instructions), then watch the video, click on the above link, and finally choose the appropriate rating for Video 1. When you have completed Video 1, click ‘Next’ on the survey, open Video 2 on the memory key, watch the video and choose the appropriate rating for Video 2. Please continue to complete the ratings in this manner until you have finished rating Video 10. If you wish to change your mind about any previous ratings, please do not hesitate to click on ‘Prev’ and work your way back to the appropriate video. To complete the survey click on ‘Done’ when you have finished Video 10.
When listening to the samples, we recommend that you use Sennheiser 90 headphones (or equivalent). Genelec speakers are also appropriate. Where high-quality audio equipment is not available, the research assistants can arrange a loan of Genelec speakers from the Speech and Language –Therapy Department at Temple St Children’s University Hospital. Please do not hesitate to contact either of us if this is the case.

You will be required to complete this first set of ratings by the 21/02/2014.

When you have completed Part 1 of the study, we will then send you another email outlining details of Part 2 of the study in four weeks’ time. This email will contain visual analogue scales. You will then rate the same 10 samples according to the visual analogue scales. The second set of ratings should also take about one hour.

Please see the attached instructions for opening the videos on the encrypted memory keys (there will also be a hard copy of these instructions enclosed in the package you received with your keys).

Yours sincerely,

Sarah Jane Osborne            Martina Dwyer            Professor Triona Sweeney

______________________________  ______________________________  ______________________________
Appendix 9: TSS Survey Monkey Weblink and Screenshot

https://www.surveymonkey.com/s/SP7N3KF
### Appendix 10: Visual Analogue Scale Form

<table>
<thead>
<tr>
<th>Video 1:</th>
<th>________________________________</th>
<th>Normal</th>
<th>Severely hypernasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video 2:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 3:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 4:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 5:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 6:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 7:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 8:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 9:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
<tr>
<td>Video 10:</td>
<td>________________________________</td>
<td>Normal</td>
<td>Severely hypernasal</td>
</tr>
</tbody>
</table>
Appendix 11: Temple Street Scale – Hypermnasality (Paper Version for Students)

Video 1:
1. Absent
2. mild, evident but acceptable.
3. mild/moderate, unacceptable distortion, evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.
6. severe, evident on all vowels and most voiced consonants

Video 2:
1. Absent
2. mild, evident but acceptable.
3. mild/moderate, unacceptable distortion, evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.
6. severe, evident on all vowels and most voiced consonants

Video 3
1. Absent
2. mild, evident but acceptable.
3. mild/moderate, unacceptable distortion, evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.
6. severe, evident on all vowels and most voiced consonants

Video 4
1. Absent
2. mild, evident but acceptable.
3. mild/moderate, unacceptable distortion, evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.

6. severe, evident on all vowels and most voiced consonants

**Video 5**

1. Absent

2. mild, evident but acceptable.

3. mild/moderate, unacceptable distortion , evident on close vowels

4. moderate, evident on close and open vowels.

5. moderate / severe, evident on all vowels and some consonants.

6. severe, evident on all vowels and most voiced consonants

**Video 6**

1. Absent

2. mild, evident but acceptable.

3. mild/moderate, unacceptable distortion , evident on close vowels

4. moderate, evident on close and open vowels.

5. moderate / severe, evident on all vowels and some consonants.

6. severe, evident on all vowels and most voiced consonants

**Video 7**

1. Absent

2. mild, evident but acceptable.

3. mild/moderate, unacceptable distortion , evident on close vowels

4. moderate, evident on close and open vowels.

5. moderate / severe, evident on all vowels and some consonants.

6. severe, evident on all vowels and most voiced consonants

**Video 8**

1. Absent

2. mild, evident but acceptable.

3. mild/moderate, unacceptable distortion , evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.
6. severe, evident on all vowels and most voiced consonants

Video 9
1. Absent
2. mild, evident but acceptable.
3. mild/moderate, unacceptable distortion, evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.
6. severe, evident on all vowels and most voiced consonants

Video 10
1. Absent
2. mild, evident but acceptable.
3. mild/moderate, unacceptable distortion, evident on close vowels
4. moderate, evident on close and open vowels.
5. moderate / severe, evident on all vowels and some consonants.
6. severe, evident on all vowels and most voiced consonants
Appendix 12: Debriefing Report

Project Title: Evaluation of Assessment Scales of Hypernasality: The Temple St. Scale (TSS) of Nasality vs. the Visual Analogue Scale (VAS) in terms of Inter-Rater Reliability in the Trained and Untrained Listener

Researchers’ Names: Martina Dwyer, Sarah Jane Osborne

Supervisor’s Name: Adjunct Professor Triona Sweeney

Aims of Research:

Aim 1: To compare the inter-rater reliability of the descriptive scale provided within the TSS with that of a VAS when assessing hypernasality

Aim 2: To compare inter-rater reliability for both scales for trained and untrained listeners.

General Procedure: SLTs trained in the use of the TSS were recruited to the study via the Dublin Cleft Centre and the IASLT Cleft Palate SIG. SLT students, who were untrained in the use of the TSS were recruited via the University of Limerick MSc in Speech and Language Therapy programme. 10 recordings of speakers with cleft lip and palate and velopharyngeal dysfunction were anonymised and transferred onto encrypted USB keys. Such keys were sent to clinicians via registered post. Clinicians were required to listen to these samples on their own computers. Firstly, clinicians were required to rate the samples according to the TSS via a Survey Monkey web link. Four weeks later, clinicians were then required to rate the same samples according to a 10cm paper VAS and returned ratings via post or email. Students attended two listening sessions at a lecture hall at the University of Limerick. An encrypted USB key containing the samples was unlocked by the researchers and presented to students over an overhead projector. At the first session, students were required to rate samples according to the TSS via a Survey Monkey web link or via a paper version of the TSS. At the second session, four weeks later, students were required to rate samples according to a 10cm paper VAS. Paper versions of scales were collected at the end of both sessions. Listening sessions were separated by a four week gap to prevent the possibility of recall. Ratings were inputted to SPSS Version 22 where they were subject to intra-class correlation statistics. Mean percentage agreement among participants was also calculated.

General Findings: Intra-class correlation coefficients (ICCs) from the present study indicate that both the descriptive ordinal scale provided by the TSS and a VAS are highly reliable tools for the assessment of hypernasality. Although the TSS appeared clearly advantaged in its ability to provide high levels of inter-rater reliability, the VAS proved slightly more reliable. While no previously published research has focused on the use of a VAS in hypernasality assessment, the current study lends support to Cheng’s (2006) unpublished study demonstrating the success of a VAS in rating hypernasality. However, averaging data across clinicians can weaken important aspects of individuals’ perceptions (Kreiman et al 1993) and therefore the results of this study should only be generalised to individual clinical judgments with caution. Unexpectedly, untrained listeners demonstrated higher inter-rater reliability than trained listeners on both scaling methods.
Nevertheless, trained listeners demonstrated higher inter-reliability on the TSS than on the VAS. This was to be as expected as such listeners had not been trained on the VAS. Untrained listeners displayed higher ICC values on the VAS than on the TSS. This was also to be as expected as a VAS has been considered by researchers as an easy and straightforward scale to use (Cheng 2006). The role of training in enhancing the reliability of the rating of hypernasal speech therefore remains unclear. Reduced reliability scores for trained raters may be explained by the lack of standardisation of listening conditions for such raters. Untrained raters also completed their judgements in close proximity to each other, and this may also have inflated levels of reliability. Further examination of the role of training is required via larger sample sizes.