DOCTOR OF PHILOSOPHY

The Early Assessment of Speech Outcomes in 3-year old children with Cleft Palate ± Cleft Lip (The EASO Study)

Fitzpatrick, Beth

Award date:
2023

Awarding institution:
Coventry University

The Early Assessment of Speech Outcomes in 3-year old children with Cleft Palate ± Cleft Lip

(The EASO Study)

By

Beth Fitzpatrick

PhD

December 2022

The Early Assessment of Speech Outcomes in 3-year old children with Cleft Palate ± Cleft Lip

(The EASO Study)

Beth Fitzpatrick

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

December 2022

Certificate of Ethical Approval

Applicant:

Elizabeth Fitzpatrick

Project Title:

FULL/LONG TITLE OF THE STUDY
The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip

SHORT STUDY TITLE / ACRONYM
The EASO Study

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

23 February 2018

Project Reference Number: P68435

Certificate of Ethical Approval

Applicant:

Elizabeth Fitzpatrick

Project Title:

FULL/LONG TITLE OF THE STUDY
The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip-

SHORT STUDY TITLE / ACRONYM
The EASO Study

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as High Risk

Date of approval:

18 June 2018

Project Reference Number: P66325

Contents

List of Tables and Figures

Chapter 1

Tables

Figures

Chapter 3

Tables

Figures

Chapter 4

Tables

Figures

Chapter 5

Tables

Figures

Chapter 6

Tables

Figures

Abbreviation List

ANF: Active Nasal Fricatives

BCLP: Bilateral Cleft Lip and Palate

C: consonant

CAPS-A: Cleft Audit Protocol for Speech-Augmented

CAPS-A-AM: Cleft Audit Protocol for Speech-Augmented-Americleft

CEN: Clinical Excellence Network

CCUK: Cleft Care UK Study

CI: Chief Investigator

CL: Cleft Lip

CLISPI: Cleft Palate International Speech Issues

CP: isolated Cleft Palate

CP±L: Cleft Palate +/- Cleft Lip

CRANE: Cleft Registry and Audit Network

CSAG: Clinical Standards Advisory Group

CSC: Cleft Speech Characteristic

GOS.SP.ASS: Great Ormond Street Speech Assessment

ICC: Intraclass Correlation Coefficient

ICHOM: International Consortium for Health Outcomes Measurement

ICS: Intelligibility in Context Scale

JBI: Joanna Briggs Institute

KALPHA: Krippendorf's alpha

NAE: Nasal Airflow Errors

NHS: National Health Service

PCC: Percentage Consonants Correct

PIL: Participant/parent Information Leaflet

PPI: Patient and Public Involvement

R&D: Research and Development

RCLST: Royal College of Speech and Language Therapists

RWL: Restricted Word List

SD: Standard Deviation

SLT: Speech and Language Therapist

SMCP: Submucus Cleft Palate

SVANTE: Swedish Articulation and Nasality Test

ToPS Trial: Timing of Palate Surgery Trial

UCLP: Unilateral Cleft Lip and Palate

V: vowel

VAS: Visual Analogue Scale

VPC: velopharyngeal competency

VPC- Rate: Velopharyngeal Competence- Rate

VP Function: velopharyngeal function

VPI: Velopharyngeal insufficiency

WHO: World Health Organisation

WF: Word final

WI: Word initial

WM: Word medial

WMCLPS: West Midlands Cleft Lip and Palate Service

Abstract

Introduction

Speech assessments at age-3 years are needed to identify children with cleft palate ± cleft lip (CP±L) who are at risk of poor speech at age-5 years. It is important to identify children in this population who require further management, either therapeutic, surgical or a combination of both, as soon as possible. Currently, there are no published, valid, and reliable speech assessment protocols in English at age-3 years for the cleft palate +/- cleft lip (CP±L) population.

Aim

To propose a valid and reliable assessment framework to assess speech outcomes in 3-year old patients with CP±L. To achieve this through the examination of different speech samples, rating methods and scales on listener reliability and validity, and the evaluation of the acceptability and usability of the assessment to Speech & Language Therapists (SLTs).

Methods

Two speech samples were specifically developed for 3-year olds with CP±L and were used alongside an assessment framework, the Adapted CAPS-A, to assess speech outcomes at 3-years. Twenty-five 3-year olds participated (20 with CP±L; five without CP±L). Two speech samples were undertaken (Sample A: spontaneous speech and picture naming; Sample B: short sentence repetition). Completion rates of the speech samples were recorded. Seven SLTs from five UK cleft teams analysed video recordings of the speech samples for the following parameters of speech: Cleft speech characteristics (CSCs), phonology, resonance, nasal airflow errors (NAE), and a judgment of velopharyngeal (VP) function for speech. VP function and hypernasality were additionally measured using Visual Analogue Scales (VAS). Listener reliability was examined. SLTs completed an electronic questionnaire about the acceptability of the assessment methods used.

Results

Seventy percent of children in the CP±L group completed both speech samples in full. More children attempted some or all of Speech Sample A (85%) compared to Sample B (70%). Both speech samples had good reliability for resonance, NAE and CSCs. However, when only fully completed speech samples are considered Sample A had superior reliability for VPC-Rate and CSCs. Inter-rater reliability relating to the classification of phonological processes was suboptimal for both speech samples. The VAS had poorer reliability scores compared to ordinal scales for both speech samples (most notable for Speech Sample A). Children without a cleft were not rated as having speech difficulties associated with CP±L. The SLTs supported the use of the Adapted CAPS-A to measure speech outcomes at age-3 years.

Conclusions

It is possible to reliably assess speech in 3-year olds with CP±L when using speech samples and an assessment framework designed for use with this age group. Both of the speech samples designed in the study are appropriate for use when assessing speech in 3-year olds with CP±L. A new assessment framework to assess speech outcomes at age-3 years is proposed.

Chapter 1. Background and Preliminary Literature Review

1.1 Cleft Lip and Palate

Cleft lip and cleft palate are the most common craniofacial birth anomalies, ranking among the most frequent of all congenital birth defects (Cronin & McLeod 2019). A cleft lip refers to a gap in the top lip, and a cleft palate describes a gap in the roof of the mouth (the palate) each due to a lack of fusion (National Health Service [NHS] Choices 2019). A cleft lip and cleft palate can occur independently or together, with embryologic and epidemiologic data indicating that cleft lip ± cleft palate and isolated cleft palate are discrete forms of clefting (Jiang et al. 2006). Cleft lip ± cleft palate can be associated with other additional malformations, and over 400 syndromes involving cleft lip and/or cleft palate have been identified (Kummer 2021). Despite this, for children with cleft lip ± cleft palate, 70% have no other identified abnormalities (Calzolari et al. 2007) whereas for children with isolated cleft palate this percentage is lower, with only 50% occurring in the absence of any other abnormalities (Burg et al. 2016). Non-syndromic cleft lip and palate has a complex aetiology associated with a number of genetic and environmental risk factors which disrupt the attachment and fusion of tissue planes including and above the lip and/or the palate (hard and/or soft) (Dixon et al. 2011) during early embryologic development.

The incidence of cleft lip ± cleft palate has been reported to vary in different ethnic and racial groups, by geographic origin, and by socioeconomic status (Meng et al. 2009). El-Shazly et al. (2022: 30) reported that the incidence of cleft lip ± palate is highest in Asian populations (0.82-4.04 per 1000 live births), followed by Caucasian populations (0.9-2.69 per 1000 live births), and is lowest in African populations (0.18-1.67 per 1000 live births). Incidence has also been reported to vary by sex with a higher incidence of isolated cleft palate occurring in females, and cleft lip ± cleft palate generally occurring more frequently in males (El-Shazly et al. 2022: 30; Mossey et al. 2009).

Cleft lip ± palate has been described as ‘phenotypically diverse’ (Allori et al. 2017a :175) and is most often classified by describing the laterality and extent of the cleft. Clefts are commonly categorised as; i)- Cleft Lip (CL), ii)- Unilateral Cleft Lip and Palate (UCLP), iii)- Bilateral Cleft Lip and Palate (BCLP), iv)- isolated Cleft Palate (CP). Clefts are also described as being complete or incomplete, depending on the ‘presence of tissue across the line of the

cleft' (Peterson-Falzone et al. 2017: 1). Figure 1.1 shows the normal structures of the lip and palate. Figure 1.2 depicts isolated CP, UCLP and BCLP.

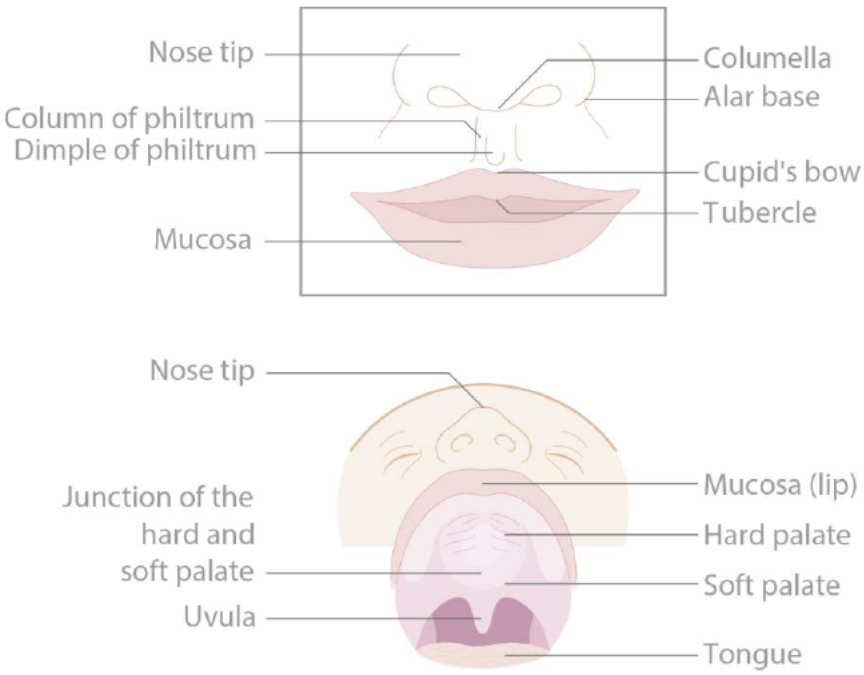


Figure 2.1 Normal anatomy of the lip and palate. Image used with permission of the West Midlands Cleft Lip & Palate Service

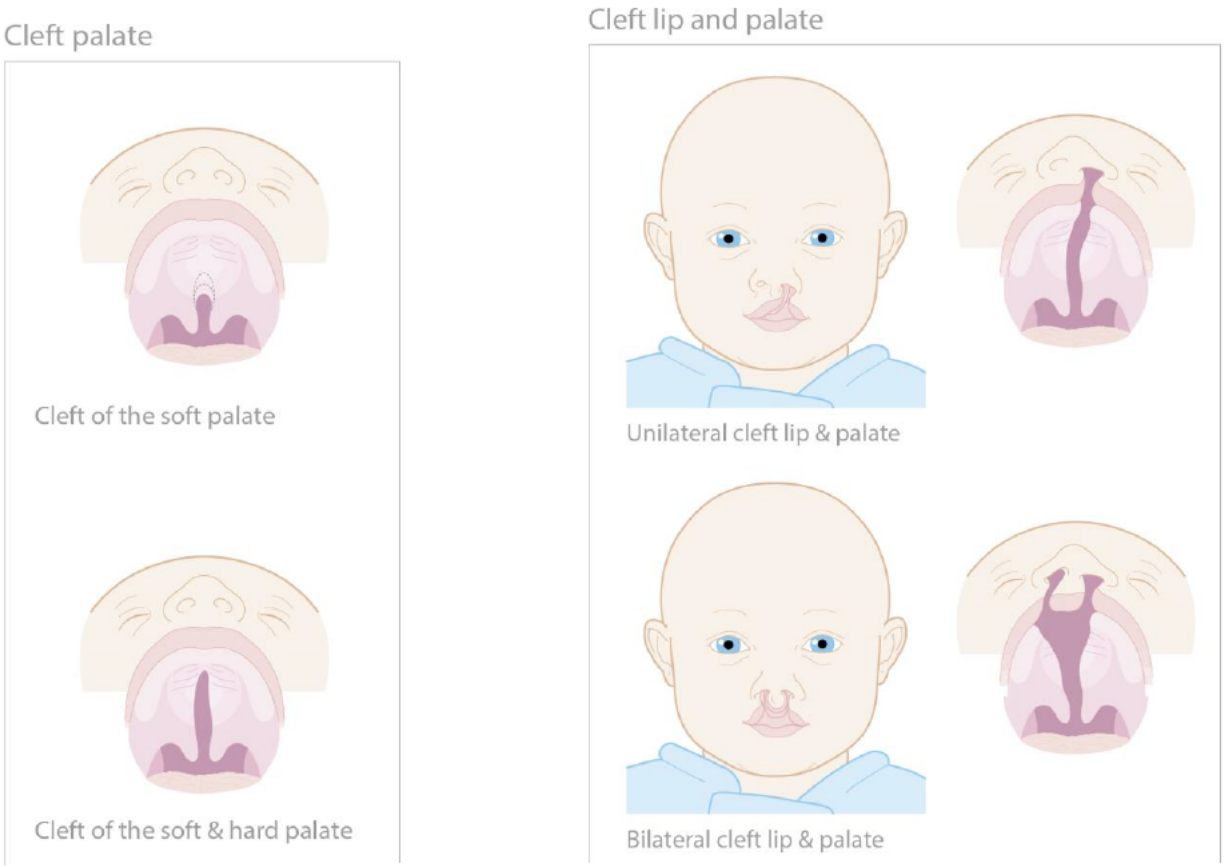


Figure 1.2 Illustration of Cleft Palate, Unilateral Cleft Lip and Palate (left sided) and Bilateral Cleft Lip and Palate. Image used with permission of the West Midlands Cleft Lip and Palate Service.

Evidence from the Cleft Registry and Audit Network (CRANE, 2016) indicates that for UK births between 2005-2014, isolated CP was the most frequent type of cleft, in 44.9% of all cleft births during this time period. Isolated cleft lip occurred in 24.1% of the cleft births, UCLP occurred in 21.5% and BCLP was the rarest form of cleft occurring in 9.5% of all cleft births. The frequency of different types of cleft in the UK is also reflected globally (El-Shazly et al. 2022: 30). This thesis has focussed on children with cleft palate ± cleft lip and does not consider children with isolated cleft lip given evidence that these children are at no greater risk of speech difficulties than the general paediatric population (Vallino et al. 2008).

1.1.1 Management of Cleft Palate ± Cleft Lip

Cleft palate ± cleft lip (CP±L) is associated with considerable morbidity (Dixon et al. 2011) and initially involves surgical repair of the cleft lip (if present) followed by palate repair. Palate repair is required to separate the oral and nasal cavities, achieve velopharyngeal closure during speech and support normal eating and drinking, facial growth, aesthetics, dentition, and middle ear function (Hoghoughi et al. 2021; LaRossa 2000). However, cleft care extends beyond the surgical repair: the impact of CP±L is wide-reaching and long-term multi-disciplinary care provided by specialist teams is recommended to help individuals achieve good outcomes relating to their emotional wellbeing, dental health, appearance, eating and drinking, hearing and speech (Lethaus et al. 2021; NHS England 2013; Bearn et al. 2001). In the UK, cleft care is centralised into 11 NHS centres or managed clinical networks that provide multidisciplinary care throughout childhood and into adulthood (Persson et al. 2015).

The NHS England Quality Dashboard (2018) sets standards by which the quality of care provided by NHS Cleft Services in England is measured. This includes guidance for the timings of surgical repair, with a target of 100% of infants receiving their lip repair before 6 months and their palate repair before 13 months. Palate repair involves the closure of the cleft defect and the realignment of the levator veli palatini, palatopharyngeous and palatoglossus muscles to enable normal movement of the soft palate during speech and eating and drinking (Stein et al. 2019). At the West Midlands Cleft Lip and Palate Service (WMCLPS), where this study was undertaken, lip repairs are typically undertaken when infants are 3-6 months of age, and palate repairs between 6 -12 months. In addition to surgical standards, other standards relate to the timeliness of specialist cleft nurse contact with the families of children with a cleft, targets for contact with cleft psychologists, hearing assessments, dental and orthodontic

outcomes, and speech assessment outcomes at age-5 years (NHS England Quality Dashboard 2018). No speech outcomes are formally reported in the UK before age 5-years.

1.1.2 The impact of cleft palate on speech development

Individuals born with CP±L are at a significantly higher risk of speech impairment than the general paediatric population (Sell et al. 2015; Law et al. 2000). There is a need to understand the role of the palate in speech production given that many, but not all, of the speech difficulties in this population originate from the anatomical and physiological impact of the cleft (Atkinson & Howard 2011). Both the hard palate and soft palate prevent air from entering the nose during normal speech; however, it is the soft palate which is dynamic and can be lowered or raised to alter the overall resonance of the voice. The nasal consonants /m, n, ŋ/ are the only consonants in English produced with the soft palate lowered, thus allowing air into the nose as shown in Figure 1.3. Raising and extending the soft palate brings the palate into contact with the posterior pharyngeal wall, thus preventing air from entering the nasal cavity and enabling the production of oral consonants, shown in Figure 1.4. In English these pressure consonants are the plosive sounds /p, b, t, d, k, g/ and the fricative and affricate sounds /f, v, s, z, ʃ, ʒ, tʃ, dʒ/ (McLeod & Baker 2017). Typical speech production requires air to be appropriately directed through the oral and nasal cavities, in order to produce the range of sounds of a given language (Howard et al. 2019). An unrepaired cleft palate leaves the nasal and oral cavities conjoined, and thus, except for the nasal consonants, the production of all consonant and vowel sounds is affected.

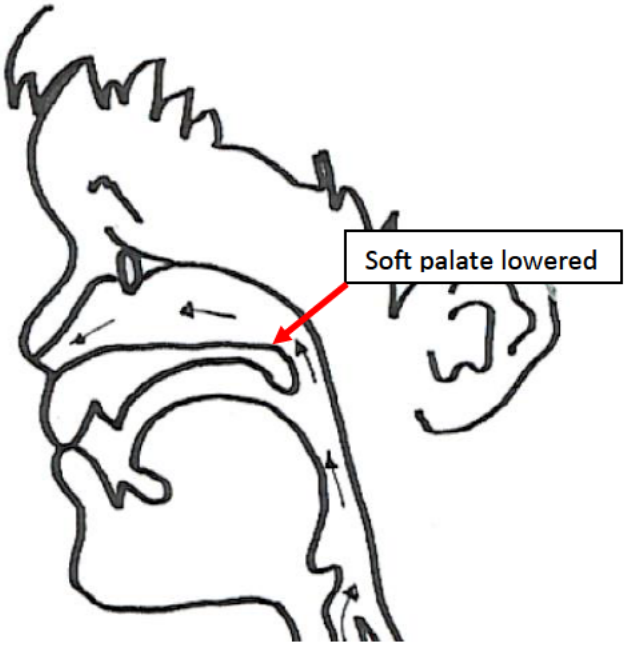


Figure 1.3 Soft palate is lowered for production of nasal consonants allowing airflow into the nasal cavity during speech. Image used with permission of the WMCLPS

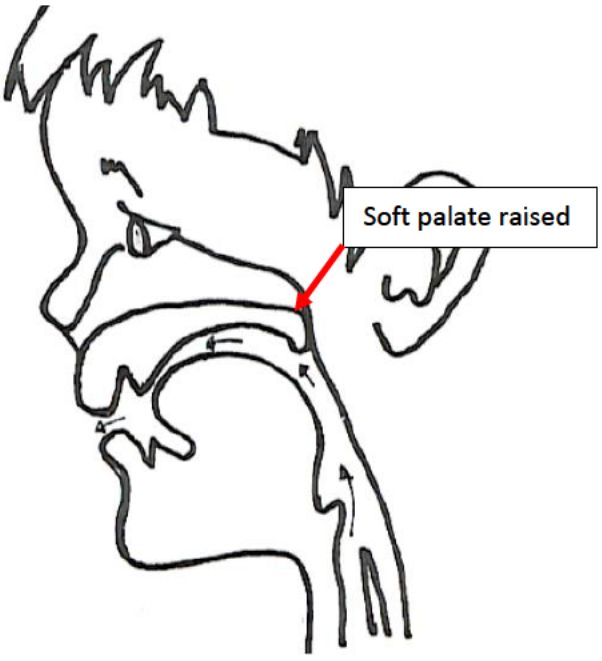


Figure 1.4 Soft palate is raised for production of oral pressure consonants preventing airflow into nasal cavity during speech. Image used with permission of the WMCLPS.

Following surgical repair palatal function may be inadequate for speech (Sell et al. 2015; Mani et al. 2010; Sell et al. 2001) which is termed velopharyngeal insufficiency (VPI) (Sell & Harding-Bell 2019). Both anatomical and physiological constraints of the velopharyngeal mechanism can result in VPI. VPI may be caused by a soft palate which is too short, does not elevate sufficiently, does not move dynamically or moves with a reduced range of movement, or due to a markedly deep pharyngeal space which the soft palate cannot close all of (which may be contributed to by small or even absent adenoids) (Howard et al. 2019). A hole, termed a 'fistula,' can also develop in the immediate postoperative period following palate repair. Rates of fistulae have been reported to vary; in a systematic review of the literature Hardwicke et al. (2014) reported they occurred in 5.4-17.9% of palate repairs, more recently Yang et al. (2020) reported that 31.3% of UCLP repairs had evidence of post-operative fistulae.

VPI can compromise the development of some speech sounds (Harding-Bell 2019). Speech signs of VPI include hypernasality (excessive nasal resonance), nasal emission (audible airflow from the nose during speech), nasal turbulence (distracting nasal snort/friction during speech) (Grunwell & Sell 2001a), and weak or nasalised consonants (John et al. 2006). To guide management decisions specifically relating to further surgical intervention, it is important that a differential diagnosis is made between airleak from the velopharyngeal mechanism and that from a palatal fistula. Sell & Harding-Bell (2019) outline several techniques to aid this differential diagnosis including the careful analysis of anterior versus posterior target sounds, the use of the mirror test (in which a dental mirror or reflector is placed under the nose to identify nasal airflow), fistula occlusion and oral examination. Surgical correction of the fistula may be necessary, and if VPI is indicated further surgery, often termed secondary speech surgery, may be required. The UK CRANE Database (2019) reported that 17.8% of all 5-year olds had already undergone secondary speech surgery to correct VPI, and that 16% of the 5-year olds continued to present with evidence of VPI.

It is also necessary to differentiate between passive and active cleft type speech characteristics (Harding & Grunwell 1998) to inform future management plans. Passive cleft type speech processes occur when there is no attempt made by the individual to compensate for VPI, and as such nasalised vowels and consonants, nasal replacements of consonants, hypernasality and nasal emission and/or turbulence occur (Chapman & Willadsen 2011). In

contrast, when active speech characteristics are used, there is an active change to either the manner or place of articulation or the direction of airflow. For example, the place of articulation may be moved outside of the oral cavity i.e. glottal or pharyngeal articulation, or the manner and direction of airflow may be actively directed into the nasal cavity as for active nasal fricatives (Howard et al. 2019: 36).

The types of articulation difficulties frequent in the CP±L population can also be described as Cleft Speech Characteristics (CSCs) (John et al. 2006; Sell et al. 1999). CSCs are typically grouped in relation to "structure and function" (John et al. 2006: 275) as shown in Figure 1.5. and summarised below in Table 1.1.

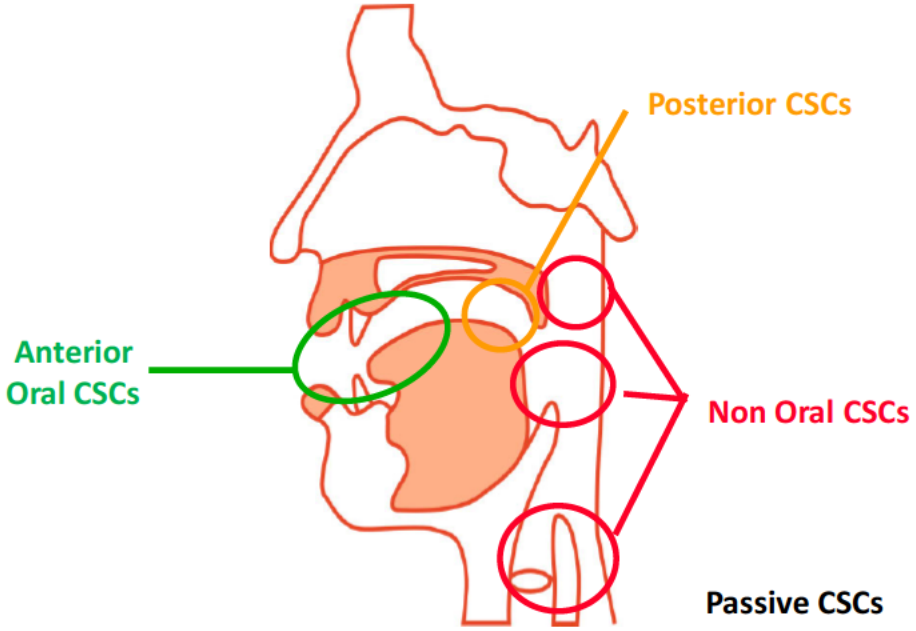


Figure 1.5 Categorisation of CSCs. Image used with permission of Debbie Sell from CAPS-A Team.

Table 1.1 CSCs occurring in English by group.

CSC group	Specific CSCs	Change to articulation/manner/direction of airflow
Anterior Oral CSCs	• Dentalisation/interdentalisation • Lateral/lateralisation • Palatal/palatalisation	Change of place of articulation at the front of the mouth.
Posterior Oral CSCs	• Double articulation • Backing to velar/uvular	Use of a retracted place of articulation within the oral cavity.
Non-oral CSCs	• Pharyngeal articulation • Glottal articulation • Active nasal fricatives • Double articulation with a glottal	Use of place of articulation outside of the oral cavity. Re-direction of airflow into the nasal cavity (Active nasal fricatives).
Passive CSCs	• Weak and nasalised consonants • Nasal realisations of plosives and/or passive nasal fricatives • Gliding of fricatives and affricates	Inappropriate nasal airflow during speech. Can result in change of manner (nasal realisations, gliding). Signs of VPI.

CSCs can result in persistent speech difficulties requiring long term therapy intervention (Sell et al. 2015). It has been estimated that 50-68% of children with CP±L require therapy intervention (Peterson-Falzone et al. 2009; Hardin-Jones & Jones 2005). Therapy before the age of 5-years is associated with improved speech outcomes (Sell et al. 2017). In the UK therapy intervention is provided by NHS regional cleft teams, NHS community speech and language therapy services and increasingly by privatised, independent speech and language therapy providers (Williams et al. 2021).

The presence of CSCs can also impact an individual's phonology. Phonology refers to the organisation and relationships of sounds in a language (Crystal 2002: 165) affecting sounds across a place and/or manner of articulation. This can result in a limited consonant inventory and a reduction in the ability to use sounds contrastively (Harding and Grunwell 1996). Harding-Bell (2019) reports that the following CSCs: backing to velar, glottal articulation, nasal realisation of plosives, nasal realisation of fricatives, active nasal fricatives,

and gliding of fricatives can also be considered cleft related phonological processes. As such the presence of a cleft palate can have a wide-reaching impact on an individual's speech sound system at both an articulatory and phonological level. It is also important to consider that children with CP±L can also present with articulation and phonological errors which occur in typical speech development (Klintö et al. 2014a), resulting in complex speech profiles.

The impact of the cleft on the growth of the jaws and irregular dentition in individuals with UCLP and BCLP can also impact the development of normal articulation (Howard 2011: 131-132). Children who have CP±L as part of a syndrome may have specific types of articulation difficulties that are characteristic of that syndrome. For example, D'Antonio et al. (2001) reported the predominance of glottal articulation in children with 22q11.2 deletion syndrome, which does not occur to the same extent in children with non-syndromic CP±L (Hardin-Jones & Chapman 2022). Persson et al. (2002) also reported that children with CP±L and a syndrome and/or other malformations had poorer speech in terms of indicators of VPI and disordered articulation when compared to those with CP±L alone, highlighting the additional impact the presence of a syndrome can have on speech outcomes.

Russell et al. (2022) reported the importance of early palate repair, specifically that palate repair before 13 months of age resulted in improved speech outcomes at age 5-years. Children aged 1-year or younger, who have an unrepaired or partially repaired cleft palate have also been found to produce fewer oral consonants than children without CP±L (Hardin-Jones & Chapman 2018), further highlighting the importance of early surgical correction of the cleft palate. It also demonstrates the need for early speech assessment and intervention given that young children with CP±L already show evidence of speech impairment. It is vital that cleft teams work to ensure that children achieve good speech outcomes as early as possible, not only because speech outcomes are one of the primary outcome measures of cleft care in this population (Grunwell & Sell 2001b: 68), but also because communication is central to quality of life; speech production difficulties can impact educational outcomes, social communication, and self-esteem (Law et al. 2000; Johnson et al. 1999).

There is evidence that speech outcomes can be poorer in children whose cleft is associated with a genetic syndrome. Basta et al. (2014) reported on speech outcomes in 132 children with cleft palate. Children with an additional diagnosis of 22q11.2 deletion syndrome consistently presented with borderline hypernasality and were three times more likely to

require secondary palatal surgery than children without this diagnosis. Seifert et al. (2021) reported that parents rated speech intelligibility significantly lower in children with identified syndromes when compared to children with CP±L in the absence of a syndrome. This is also supported by evidence from Feragen et al. (2017) who also reported poorer speech intelligibility in children who had additional conditions, including identified syndromes and ADHD. Persson et al. (2002) compared the speech outcomes of children with and without additional malformations (including identified syndromes) and reported that children with additional malformations had significantly higher rates of VPI, weak pressure consonants, hypernasality, glottal articulation and retracted articulation. This demonstrates that such additional malformations can impact not only on palatal function, but also articulation outcomes. As such, children with CP±L and identified syndromes have been routinely excluded in the UK audit of speech outcomes at age-5 years in the UK (Britton et al. 2016; Britton et al. 2014), given that this cohort may present with significantly different speech outcomes. However, given the number of different syndromes associated with CP±L (Kummer 2021), to date there is limited information relating to speech profiles for specific syndromes (other than for 22q11.2) which could be used to develop specific speech outcomes or standards for specific syndromes.

1.1.3 The impact of cleft palate on wider communication and developmental skills

Whilst it is specifically speech outcomes that are both a primary outcome measure for cleft care in the UK (NHS England Quality Dashboard 2022-2023; CRANE 2021) and the focus of this study, it is important to recognise the wider communication and developmental needs of the CP±L population. This is particularly relevant for younger children because broader communication or developmental impairments may impact upon their ability to engage and complete speech samples and assessments and thus measure speech outcomes.

Children with CP±L are at risk of conductive hearing loss which occurs due to otitis media with effusion which is described as almost 'universal' in young children with CP±L (Flynn & Lohmander 2014; Flynn et al. 2014). Whilst fluctuating conductive hearing loss is most prevalent in preschool children hearing generally improves with age (Fitzpatrick et al. 2021; Flynn & Lohmander 2014; Handžić-Ćuk et al. 1996). However, hearing loss at high frequencies can persist into older childhood and adolescence in the CP±L population (Flynn & Lohmander 2014; Flynn 2013). The impact of conductive hearing loss on cleft speech outcomes is not conclusive (Fitzpatrick et al. 2021), however broader consideration should be

given to the impact of hearing loss on other outcomes such as psycho-social development, language, and listening skills (Harman et al. 2015).

Children with CP±L have been found to have both poorer academic and neuro-developmental outcomes (Gallagher & Collett 2019; Bell et al. 2017). Hardin-Jones & Chapman (2014) found that early lexical development was delayed in the CP±L population, and this is supported by a scoping review by van Eeden & Stringer (2020) which concluded that there is evidence of both delayed linguistic and auditory processing skills in young children with CP±L. Specifically, van Eeden & Stringer (2020) concluded that in children with CP±L under 3-years of age, both expressive vocabulary development and mean length of utterance have been found to be significantly lower than for age matched controls. However, the authors reported that evidence of persisting language impairment in school aged children with CP±L is less clear. Whilst there is evidence that school aged children with CP±L performed significantly poorer on standardised language tests than their non-cleft peers, synthesis of the standard scores indicated that the CP±L groups were within one standard deviation of the norm. This conclusion is supported by the meta-analysis conducted by Lancaster et al. (2020) which concluded that differences in the language performance of individuals with CP±L and non-cleft controls decreases with age, with most significant differences in children 36 months of age or younger. In particular, disordered articulation has been reported to be associated with reduced expressive vocabulary (Hardin-Jones & Chapman 2014; Scherer et al. 2008; Chapman et al. 2003) in preschool children with CP±L.

Speltz et al. (2000) used the Bayley Scales of Infant Development to assess both expressive language and non-verbal developmental skills in CP±L and age matched peers at 24 months and reported that children with CP±L scored significantly lower in both areas, leading the authors to conclude that young children with CP±L did not present with specifically impaired language skills i.e. developmental language disorder, and that early expressive language impairments should be considered in the context of broader developmental delays. van Eeden & Stringer (2020) concluded that children with CP±L can have difficulties processing language against background noise, despite levels of normal hearing. This highlights another factor which should be considered when analysing the language profiles and understanding the aetiology of language impairment in young children with CP±L, which may be impacted by articulation, wider developmental delays and auditory

processing difficulties in background noise. Indeed, van Eeden & Stringer (2020) reported that in older age groups, it is challenging to specifically comment on the nature of any language impairment, as studies frequently report composite scores from standardised assessments, and thus do not provide specific information relating to subtests that would provide more detailed information about the types of language disorder in the cleft population. Certainly, whilst further research is required to map the language profiles of older children with CP±L and understand the relationship between language impairment and other comorbidities, SLTs working with young children with CP±L need to be aware of the potential for language impairments, particularly expressive language impairments, and consider this in assessment and treatment planning.

In addition to language impairments, in their retrospective review of over 30,000 children with non-syndromic CP±L, Khoshab et al. (2021) reported that children with isolated cleft palate had significantly higher rates of global developmental delay and intellectual disability than non-cleft controls. Those with cleft lip ± cleft palate had significantly higher rates of attention deficit disorder or attention deficit hyperactivity disorder compared to non-cleft controls. In addition, there is an association between poor speech outcomes and compromised early reading skills in the CP±L population (Chapman et al. 2011).

As there are over 400 syndromes associated with CP±L each with unique characteristics and profiles it is extremely challenging to summarise how each syndrome may impact on both communication and other developmental skills. However, there is evidence that children with syndromic CP±L present with more middle ear abnormalities and a higher prevalence of hearing loss (Flynn et al. 2014). In addition, Feragen & Stock (2014) reported that children with syndromic CP±L presented with increased emotional and psychological difficulties. Several syndromes associated with CP±L are also associated with receptive and expressive language impairments including 22q11 deletion syndrome (Solot et al. 2019), Apert syndrome (Kilcoyne et al. 2022) and Kabuki syndrome (Barry et al. 2022). SLTs working in CP±L need to consider the needs of each individual patient to comprehensively assess their speech and develop appropriate treatment plans. In addition, it is also important to consider the potential impact of a specific syndrome on an individual's hearing, emotional and psychological wellbeing, and language development in order to adapt assessments and treatment plans to the individual.

1.2 Speech Assessment and Outcome Measurement in children with CP±L

1.2.1 The Challenges of Perceptual Speech Assessment

Lohmander & Howard (2011) outlined that a reliable and valid assessment of speech is a primary requirement in cleft management and research. Perceptual assessment using narrow phonetic transcription has been described as the 'gold standard' for the assessment of cleft speech (Howard 2011; Sell 2005). Perceptual speech assessment typically involves the phonetic transcription of consonant production and the use of rating scales to assess those symptoms of speech associated with VPI (Lohmander & Howard 2011). Perceptual speech assessment allows an SLT to capture information about an individual's phonetic and phonological development, form a differential diagnosis, identify the need for therapeutic intervention, make judgements regarding how effectively the palate is working during speech, and explore the need for further surgery. Assessment for treatment planning focuses on each individual patient's needs, and as such a variety of different assessments may be used in combination. In the UK, the recommended speech assessment for children aged-5 years and over is the Great Ormond Speech Assessment (GOS.SP.ASS) (Sell et al. 1999) which "provides a systematic framework, based on a standard speech sample in which to assess and document the different speech parameters" (Sell & Harding-Bell 2019 pp.116). However, as Howard (2011) describes, it is important to consider the aspect(s) of speech production which SLTs aim to capture, and this may differ on an individual basis, for example to gain information on stimulability, or speech production in less or more demanding contexts. As Bates & Titterington (2021) advise, perceptual speech assessments should be of sufficient detail to inform clinical decision making, therapeutic targets and the approach to intervention. In contrast, when perceptual speech assessments are administered, and information is collected and recorded in a standardised format, these assessments can be used as a basis to measure speech outcomes. Whilst each speech outcome relates to the individual patient, it is therefore possible to present the outcomes of a cohort of patients, and thus compare speech outcomes, and measure these against agreed standards in audit. Cleft SLTs working in the UK are required to conduct both assessments for treatment planning and assessments to measure speech outcomes.

Perceptual speech assessment is not without its challenges, Kent (1996) reported that perceptual judgements are susceptible to error and bias. This, in turn, has implications for the

reliability of listener judgements, that is whether listener judgements consistently reflect the individual's speech. Shriberg & Lof (1991) reported that the severity of the speech disorder, the type of speech sample, the word position of the target sound, and whether narrow or broad phonetic transcription is used, all impact the reliability of listener judgements. There is, therefore, an inherent challenge in making reliable perceptual judgements of cleft palate speech given that CP±L is associated with complex and disordered speech profiles requiring narrow phonetic transcription e.g. to capture the distribution of nasal airflow errors (Howard 2011).

Given concerns relating to the subjectivity of perceptual assessment, Chapman et al. (2016) stated that there is a requirement for researchers to report on levels of listener reliability and agreement. Historically this has not been routine practice; Lohmander & Olsson (2004) reported that only 51% of the 88 articles they reviewed reported on levels of listener reliability. Chapman et al. (2016) have also described that is challenging to compare reliability results between studies because different statistical approaches and numbers of listeners may have been used. To reflect the need for reliable perceptual speech assessments, Sell & Sweeney (2020) have provided an updated statement that "narrow phonetic transcription and the reporting of intra- and inter-rater reliability are acknowledged as the gold standard" for speech assessments (Sell & Sweeney 2020: 143). A further recommendation can be made; whilst it is one thing to report on reliability, it is another to ensure that that standardised methods of perceptual speech assessment and analyses are used, meaning that any assessment practices result in sound levels of listener reliability.

The internal standards of a listener have been reported to impact upon listener reliability (Kreiman et al. 1993). Such internal standards refer to the comparison of a stimulus to a standard which has been developed and memorised, and thus can vary between listeners (Yamashita et al. 2018). One solution to this challenge is listener training, which has been shown to result in improved listener inter and intra-reliability (Lohmander et al. 2017a; Willadsen et al. 2017; Chapman et al. 2016; Lee et al. 2009; Sell et al. 2009). The development of assessment protocols, with agreed parameters of assessment that are consistently used by SLTs, such as the Cleft Audit Protocol for Speech- Augmented (CAPS-A) (John et al. 2006) may also favourably impact listener reliability. Another factor to consider is the impact of different speech samples e.g. connected speech (spontaneous, sentence repetition or reading), or

single words (spontaneous or repetition) on the reliability of listener judgements. Keuning et al. (1999) reported no difference in the reliability of listener judgements when listening to and subsequently analysing sentences read aloud, sentence repetition, and conversational speech in the cleft population. In contrast, Klintö et al. (2011), who investigated the impact of single word naming, a sample of connected speech, and sentence repetition on listener reliability, reported that single word naming resulted in the most reliable listener judgements. This highlights the potential of single word naming as a reliable speech sample, which may be particularly suited to younger children. However, alongside reliability, it is important to consider the validity of different speech samples in representing the speech of an individual. Howard (2013) exemplifies this in a case study detailing the speech of two children with CP±L (aged 9 years 5 months and 11 years 0 months) one of which had markedly more accurate speech productions at a single word level in comparison to connected speech. This suggests that a single word speech sample alone may not be a valid representation of an individual's speech.

A more recent trend in cleft research has been to question the use of different types of rating scales on listener reliability. In the past, equal interval appearing (EIA) or ordinal scales have dominated, reported by Lohmander & Olsson (2004) in 78% of the studies using perceptual speech assessments in the analysis of speech in the CP±L population. At present, existing cleft outcome protocols still utilise ordinal or EIA scales. However, an emerging debate regarding the cognitive processing behind the perception of resonance and nasal airflow characteristics has resulted in the examination of the reliability of ordinal measures. Stevens (1975) first argued that there are two dimensions to how we perceive perceptual phenomena, namely prothetic and metathetic. Prothetic dimensions involve quantitative changes, examining the extent to which the phenomena have changed, whilst metathetic dimensions measure qualitative changes regarding the type of change that has occurred. Stevens (1975) went on to argue that these different dimensions of perceptual phenomena require different rating scales, and recommended that prothetic dimensions be rated using magnitude measures such as direct magnitude estimation (DME) or visual analogue scales (VAS), and metathetic dimensions require partition scales as in ordinal or EIA scales.

Zraick et al. (2000) and Zraick & Liss (2000) have argued that 'nasality' is a prothetic phenomenon and thus is not ideally measured using EIA scales. Recent studies have found

the VAS to be as reliable (Bettens et al. 2018 [measuring hypernasality, nasal airflow, understandability], Castick et al. 2017 [resonance, nasal airflow, understandability, acceptability]) or even more reliable (Baylis et al. 2015 [hypernasality, nasal emission/turbulence]) than ordinal scales. As an alternative to VAS, Yamashita et al. (2018) used the BORG centi-MAX scale. The BORG centi-MAX scale is a vertical category-ratio scale with a continuous range of numbers and corresponding labels of intensity. Anchor descriptors are used along the scale. Yamashita et al. (2018) reported favourable reliability when rating hypernasality using the BORG centi-MAX scale compared to a 2-step ordinal scale and a visual sort and ordinal rating scale. The authors concluded that the superior reliability of the BORG centi-MAX scale was because it combined both a ratio scale with a labelled category scale. A further benefit of magnitude measures using continuous scales is that they also offer wider statistical analyses. However, these scales have not been adopted by SLTs working in cleft and challenges relating to their clinical interpretation need to be considered in the application and use of magnitude measures in clinical practice.

Another variable that may impact upon perceptual judgements of speech is the recording medium. In clinical practice, transcription is often 'live'; however, to report on listener reliability and to aid transcription, given the challenge of transcribing atypical speech in real time (Howard 2011: 130), recordings of speech assessments are typically made at the WMCLPS and are routinely made in the UK to facilitate speech outcome measures at age 5-years (Sell et al. 2009) as well as recording speech outcomes following secondary surgery. Sell et al. (2002) reported that video analysis may lead to more analytical evaluations of hypernasality, nasal turbulence, and consonant production than analysis using audio only. Klintö & Lohmander (2017) compared the impact of audio-only to audio with video recordings on the phonetic transcription of the speech of 3-year old children with CP±L. For the primary outcome measure, percentage consonants correct (PCC), the authors concluded that there was no significant difference in the analysis of audio-only or audio with video recordings. The authors also discussed the McGurk effect, when the listener visualises the production of one sound but hears another, impacting their perception of the sound (McGurk and Macdonald 1976). This effect is particularly relevant to the cleft population with regards to characteristic articulation patterns such as glottal reinforcement and double articulation. Klintö & Lohmander (2017) concluded that transcription was influenced by visual information and reported a statistically higher proportion of anterior articulations and nasal airflow

characteristics when analysing recordings with both audio and video. This suggests that video can be useful to make more critical listener judgements. Despite concerns regarding the McGurk effect, this evidence supports the use of audio with video in the analysis of speech in the CP±L population. The quality of the recording is also paramount (Sell 2005) and the use of high-quality equipment in recording and playback is recommended (Sell et al. 2009).

Achieving the 'gold standard' of a valid and reliable perceptual speech assessment is therefore not without its challenges. Existing assessment procedures and protocols, and those in development, need to address these challenges and ensure that such assessments are both valid and reliable. The following section considers alternatives to perceptual speech assessment.

1.2.2 Alternatives to perceptual speech assessment

Whilst perceptual speech assessment is the principal assessment method of cleft palate speech, it is important to acknowledge the role of other instrumental assessment methods to further assess both articulation and palatal function during speech. Electropalatography (EPG) is a computer-based instrument that captures lingual articulation (Lee 2021). An artificial EPG plate (custom made for each individual) is worn in the mouth with sensors in the plate attached to wires which externally connect to a processing machine, which in turn is connected to an EPG machine and then a computer. This produces a visualisation of the pattern and timing of contact between the tongue and the hard palate (Lee 2021: 339). When used in speech assessment, EPG can be used to "supplement auditory-perceptual judgments" (Lee et al. 2022) and has been used in the UK with the CP±L population. Lee et al. (2019) report that EPG can result in more critical ratings of PCC (specifically nasal stops) than perceptual judgements in children with CP±L. Gibbon & Crampin (2001) report that EPG supports the identification of labial-velar double articulations in the CP±L population which can be missed in perceptual speech assessment, as these errors are often perceived as labial only articulations.

Despite the advantages of EPG, relating to increased precision in phonetic transcription of speech, Cleland et al. (2020) reported that EPG has not been adopted as part of routine cleft assessments, due to the logistical and financial implications of creating bespoke EPG plates for each patient. Furthermore, EPG in the cleft population has been recommended for school aged participants or older children with intransient articulation

difficulties (Yamamoto 2020; Lee et al. 2019; Fujiwara et al. 2007; Gibbon et a. 2004) and is not typically used with pre-school children.

Cleland et al. (2020) proposed an alternative to EPG, Ultrasound Tongue Imaging (UTI), which Sugden & Cleland (2021) reported as a low-cost alternative to EPG, facilitating its use with a higher number of speakers. “An ultrasound probe is placed under the chin to show a dynamic image of the tongue in either mid-sagittal or coronal view” (Sugden & Cleland 2021: 3). Cleland et al. (2020) compared perceptual phonetic transcription in three conditions: live and then from audio recordings with ultrasound-aided transcription in children with CP±L aged 3 to 12 years. Using ultrasound-aided transcription, lingual errors were more reliably identified and more detail regarding double articulation, pharyngeal/uvular articulation and retroflexion of the tongue was gained using this method. As well as aiding transcription, this study demonstrates the potential use of ultrasound with pre-school children as the youngest participant was aged 3 years 7 months.

Instrumental assessments may also be warranted if perceptual assessment shows evidence of VPI. Nasendoscopy, videofluoroscopy and magnetic resonance imaging (MRI) are all used to visualise the structure and function of the velopharyngeal mechanism. However, there are limitations to all these investigations. Nasendoscopy is invasive which limits the patient groups with whom this investigation can be used, particularly young children (Perry et al. 2017). Videofluoroscopy uses ionising radiation and as such there are safety concerns relating to repeated or prolonged videofluoroscopy (Perry et al. 2014). MRI is non-invasive and free from ionising radiation; however its use is limited by significant financial and training needs and by the length of the investigation (Mason 2022), and in their US based protocol Kotlarek et al. (2021) only used this in children aged four and over. MRI is not routinely used in UK cleft centres.

An alternative instrumental assessment is Nasometry (Kay Elemetrics 2001), although this is an indirect method and does not visualise the velopharyngeal mechanism. Nasometry does, however, produce an objective measure of nasalance (Pereira et al. 2020), the relation between nasal and acoustic energy during speech production (Bettens et al. 2014) and can be compared to language specific norms. Differences in gender (Brunnegard & van Doorn 2009) have been reported and the type of vowels in the speech sample can also influence the results (Lewis et al. 2000). Correlations with perceptual speech assessment have been reported to

vary (Karnell 2011; Sweeney & Sell 2008; Keuning et al. 2002) and thus Sweeney & Sell (2008) recommend that nasometry should be used to supplement rather than replace perceptual speech assessment. Pressure flow techniques can also be used to objectively measure nasal airflow, such as the PERCI speech aerodynamic assessment system (PERCI-SARS, Microtronics 1994). A calculation of the size of velopharyngeal gap is made by measuring the rate of nasal airflow, intra-oral airflow and the volume of airflow through the velopharyngeal gap. However, specialist equipment is needed with catheters in both the mouth and nose. Bettens et al. (2014) report that such measures are "technically complex and require substantial cooperation" (Bettens et al. 2014: 176) which precludes the use of such assessments in clinical practice, particularly with young children.

A recent trend has also been to investigate the role of artificial intelligence in speech assessments in the CP±L population. Research in this area is currently in its infancy. For example, Yogendran et al. (2022) reported on machine learning outcomes when analysing sustained vowels produced by SLTs imitating hypernasality. They reported that at this early-stage measures of shimmer may prove useful when classifying the vowel /a:/ as produced with either oral or hypernasal resonance. Cornefjord et al. (2022) reported the development of three artificial neural networks to analyse velopharyngeal competence on a three-point scale. Audio recordings were used to 'train' the neural networks. One neural network, the convolutional neural network, proved the most reliable, able to determine the correct velopharyngeal competence score in 90% of the recordings analysed. Scherer et al. (2022) also reported the development of algorithms to assess hypernasality and speech acceptability. They reported high levels of correlation between the algorithm outcomes and perceptual evaluations of these speech parameters. Despite these recent advances none of these networks or algorithms are currently able to replace the role of the SLT in perceptual speech assessment.

Although instrumental assessment is an important part of cleft care it is currently used to supplement information or hypotheses based on perceptual speech assessment rather than replace it. As such perceptual speech assessment currently remains the 'gold standard' of speech assessments in the CP±L population.

1.2.3 The development of outcome measures at Age-5 years

The need to monitor patient outcomes has been at the forefront of cleft care in the UK since the review completed by the Clinical Standards Advisory Group (CSAG) (Sandy et al. 1998). The CSAG review was initiated by leading UK cleft researchers and supported by UK Health Ministers, following the publication of a report comparing patient outcomes across European cleft centres (Shaw et al. 1992). The report indicated that several UK outcomes, including speech, were amongst the poorest across the European nations studied. The CSAG review investigated a range of patient outcomes in 239 children with UCLP treated in the UK. Speech outcomes were inadequate and only 17% of patients were judged as having normal intelligibility (a measure of how well an individual's speech is understood by others [Whitehill et al. 2011]) at age-5 years (Sell et al. 2001). Following recommendations made by CSAG, cleft care in the UK underwent a major reorganisation, with a reduction from 57 centres treating patients with CP±L to 11 specialist cleft teams (Persson et al. 2015; Bearn et al. 2001). Other recommendations made by CSAG were; that individuals with CP±L should be cared for by specialist cleft clinicians, and cleft teams should regularly audit patient outcomes (Bearn et al. 2001).

In the UK all cleft centres have cleft specialist SLTs who undergo additional training to be involved in the audit of speech outcomes (Sell et al. 2009). Annual national clinical audits, (defined by the CSAG group as the "process of comparing care with baseline criteria and standards, with the aim of implementing changes to improve care" [Bearn et al. 2001: 38]), have become an integral part of the workload of UK cleft teams. In the UK speech outcomes are assessed using a standardised assessment protocol and routinely audited at age-5 years using the CAPS-A (Sell et al. 2009; John et al. 2006) which has been specifically designed for this purpose. The CAPS-A addresses the challenges of perceptual assessment by using an assessment protocol with agreed parameters of assessment, protocols for recording and playback, listener training, and the use of consensus listening. The audit process involves recording the patient's speech assessment, utilising a standard speech sample and reviewing the recording through a process of consensus listening. Specialist cleft SLTs come to a consensus about the patient's speech outcome using the CAPS-A tool. This process is referred to as consensus listening and has been adopted because low transcriber agreement has been reported in the analysis of complex, disordered speech profiles that are associated with CP±L

(Howard & Heselwood 2002), and so multiple listeners are utilised to improve the validity and reliability of the speech data (Kuehn & Moller 2000).

Speech outcomes are audited annually against National Audit Standards (Britton et al. 2016; Britton et al. 2014) which are published on NHS England's Quality Dashboard. The key outcome standards for speech (updated in 2016) are shown in Table 1.2.

Table 1.2. UK National Audit Speech Outcome Standards for speech for the CP±L population at age 5-years.

Standard	Outcome
1	By 5 years 0 months to 5 years 11 months over 60% of children with CP±L will have speech within the normal range
2	By 5 years 0 months to 5 years 11 months over 70% children with CP±L have speech with no evidence of a structurally related problem and have not had velopharyngeal surgery or fistula repair for speech
3	By 5 years 0 months to 5 years 11 months over 68% of children with CP±L have no cleft-related articulation difficulties requiring SLT and/or surgery

The publication of speech outcomes at age 5-years in the CSAG report (Sell et al. 2001) was the catalyst for the development of the current system of speech audit in the UK, particularly the selection of age 5-years as the audit point. The CAPS-A (Sell et al. 2009; John et al. 2006) was specifically developed as an outcome measure at age-5 and was validated for use at this age (although it has been used with other age groups [Pereira et al. 2021; Birch et al. 2021; Ahl et al. 2016]). Internationally, other robust, valid, and reliable speech assessments and outcome measures at this age have emerged; this has, in turn, consolidated age-5 years as a key assessment time point.

The CAPS-A has been used as a foundation for other national and language specific versions. Two of these versions have also been specifically developed for school aged children. The Americleft modification of the CAPS-A Americleft (CAPS-A-AM) (Chapman et al. 2016) was developed and tested with children aged 5-years and over and the Japanese version of the CAPS-A was validated on children aged 4-years 6 months to 7-years 0 months (Ogata et al. 2022). Interestingly Bruneel et al. (2020) took a slightly different approach, and

in the development of their Belgian Dutch version of the CAPS-A, they validated the tool using children aged 3 to 10 years, although the mean age of the participants was 6 years 5 months and 6 years 0 months in the two phases of their study. An alternative outcome measure to the CAPS-A is the Swedish Articulation and Nasality Test (SVANTE) (Lohmander et al. 2017b; Lohmander et al. 2005) and the Norwegian translation (SVANTE-N). The SVANTE was developed with assessment at age 5-years a key time point (although subsections of the outcome measure can be used with younger age groups including age 3-years).

The development of outcome measures for children aged 5 years and over has consolidated age-5 years as a key time point at which speech outcomes are reported in research studies (Morrison et al. 2021; Schölin et al. 2020; Chacon et al. 2017; Nyberg et al. 2014), and this age was also adopted by both the international Scandcleft Trial (Lohmander et al. 2017a; Willadsen et al. 2017) and then subsequently the Timing of Palate Surgery (ToPS) Trial (Shaw et al. 2019) as the primary age for reporting speech outcome measures. Given the dominance of age-5 years in terms of research and the development of speech outcome measures, it is not surprising that in 2017, the International Consortium for Health Outcomes Measurement (ICHOM) (Alliori et al. 2017b) adopted age-5 years as the first recommended timepoint for reporting speech outcomes. The selection of age-5 years by ICHOM may reflect the lack of speech outcome measures suitable for younger children.

Changes in service delivery and the focus on the assessment and audit of speech outcomes against agreed standards at age-5 years has led to demonstrable improvements in the majority of speech outcomes at age-5 years in the UK (CRANE 2020; Sell et al. 2017; Sell et al. 2015). Indeed, the speech outcome standards have been updated to reflect this improvement (Britton et al. 2016). However, despite such improvements, UK outcome data at age 5-years continues to indicate that a substantial proportion of children have significant speech needs. In the 2020 CRANE report, 60% of UK 5-year olds were evaluated as having “normal speech” (CRANE 2020: 55). This meets Speech Standard Outcome 1 (see above) (Britton et al. 2016) but it still means that 40% of all 5-year olds with CP±L do not have “normal speech.” Indeed, examining the outcomes by cleft type, only 27.1% of 5-year olds with BCLP had normal speech, and only 51.3% of children with UCLP (CRANE 2021). This is significant because all these 5-year olds will have begun their education, and the impact of having “not normal” speech may have wider implications for their learning and social interactions

(Bettens et al. 2020; Gallagher & Collett 2019; Bell et al. 2017). The Cleft Care UK (CCUK) study was set up to evaluate the effects of the CSAG recommendations on cleft care in the UK (Persson et al. 2015). This study reviewed speech outcomes in 5-year old children with UCLP and provided a further level of detail in relation to the speech outcomes of those children who do not have "normal speech" at age-5 years. The study highlighted that 20% of the 5-year old participants had the lowest scores for intelligibility/distinctiveness, described as speech which is "only just intelligible to strangers" or "impossible to understand" (Sell et al. 2015: 41).

In summary, there has been a significant improvement in speech standards since the CSAG report (CRANE 2020; Sell et al. 2017; Sell et al. 2015), and the UK now has a valid and reliable speech outcome measure at age 5-years, the CAPS-A (John et al. 2006). However, despite changes in service organisation and audit in the UK, a persistent number of children continue to have significant speech issues at age 5-years, particularly children born with UCLP and BCLP who are not achieving the outcome standards at age 5-years (CRANE 2021). The CCUK study concluded that 'early prediction' of poor speech outcomes and 'appropriate intervention is key' (Sell et al. 2015: 44). Indeed, it is logical that for children to achieve good speech outcomes at age 5-years and to continue to improve speech outcomes, speech assessment and intervention must precede this.

1.3 Speech Assessment at Age 3-years

1.3.1 Challenges of assessment

Whilst there are clear benefits of speech assessments before age-5 years, existing assessments can present challenges to SLTs. When assessing 3-year olds, cleft specialist SLTs need to consider not only the additional needs of some 3-year olds with CP±L, who may have global developmental delay or language delay, but also the wide developmental spectrum of 3-year olds. During their fourth-year (i.e. age 3-4 years) children are mastering new skills in the areas of attention and listening, play, and speech and language (Sharma & Cockerill 2014), which has implications for the type and number of speech assessment materials that can be used, and the duration of speech assessments. For example, 3-year olds are only beginning to learn to shift their attention without adult prompting and may require different levels of adult support than older age-groups (Buckley 2012). Thus, speech assessments at age 3-years

require SLTs to use a range of clinical skills to facilitate patient engagement and completion of the speech assessment.

At age 3-years, language skills develop rapidly; at their third birthday children will typically use around 700 words, use 'why?' questions (Buckley 2012), start to talk about the past and the future and use grammatical constructions such as possessive 's' and the uncontractible copula (Bowen 1998). However, between 3 and 4 years, children learn to use articles, regular past tense, third person regular present tense, the contractible copula, and auxiliary (Bowen 1998) and can combine four or more words to express ideas, feelings, problem solve and negotiate using their vocabulary of thousands of words (Buckley 2012). Thus, in their fourth-year, typically developing children show evidence of the rapid development of attention and language skills.

UK SLTs assessing children with CP±L at age 3-years do not have an agreed standardised speech assessment (Wren 2013). A variety of different speech samples are used including a variety of different single word naming assessments (pictures or objects), spontaneous speech samples or sentence or phrase repetition (Wren 2013). The variety of speech assessments used in the UK may reflect the increased risk of delayed language skills in children with CP±L in the preschool years (Tillman et al. 2018). As such spontaneous speech samples may differ in length from single words to more complex sentences, making it difficult to compare these speech samples. Language delay may mean that sentence repetition may not be possible, and for some 3-year olds even single word naming may be a challenge.

SLTs assessing speech at age 3-years need to analyse speech not only from a cleft but also from a developmental perspective, given that 3-year olds are also likely to be making developmental progress in their speech sound production during their fourth year. Indeed. McLeod and Baker (2017:202) reported that children learn to accurately produce the majority of consonant, vowel, and consonant clusters between the ages of 3 and 5. As such, assessing speech from a developmental perspective is more often required at age 3-years in comparison to age-5. Chapman and Willadsen (2011) outlined that whilst the source of speech difficulties may be cleft-related, in a developing sound system the cleft may also impact normal patterns of phonological development, which is particularly important to consider at age 3-years and highlights a further point of difference between speech assessments at age 3-years and those with older age groups.

There are few assessment protocols designed specifically to assess speech at age-3 years in the CP±L population, with no agreed protocol available in the UK. The SVANTE (Lohmander et al. 2017b; Lohmander et al. 2005) has been designed for use from 3-years of age and uses three different speech samples: single word naming, sentence repetition, and a spontaneous speech sample all in Swedish. Embedded within the SVANTE is the SVANTE-Mini, a minimum word set designed to facilitate inter-centre and cross-linguistic comparisons (Lohmander et al. 2017b). This cross-linguistic word list was developed as part of the international Scandcleft Trial (Lohmander et al. 2017a; Willadsen et al. 2017), which reported on speech outcomes across several languages, and was subsequently validated for 3-year olds in the development of SVANTE. The SVANTE highlights single word picture naming as a relevant speech sample in the assessment of 3-year olds with CP±L.

The Cleft Palate International Speech Issues (CLISPI) website (CLISPI 2017) (which was formed by countries who had participated in the Scandcleft Trial [Lohmander et al. 2017a; Willadsen et al. 2017]) sets recommendations for speech assessments at various ages, and recommends the following speech samples are used at age 3-years: single word naming (Scandcleft restricted word test in nations language, and the nation-specific phonology/articulation test); a sample of connected speech; rote speech (counting 1-5). However, CLISPI highlights that the priority should be to collect a single word speech sample, given evidence at age-5 years that single word production reflects a child's best production (Klintö et al. 2011). Klintö et al. (2011) (reporting on 5-year olds) also reported that if the aim is to assess connected speech then sentence repetition is both valid and reliable, and that consonant production in sentence repetition reflects other connected speech samples such as conversational speech and retelling a story. However, the Klintö et al. (2011) study was based on children aged 5-years and it is unknown if these findings apply to 3-year old children as the reliability of different speech samples has not been investigated at age-3 years.

Given the additional cognitive demands of sentence repetition compared to single word naming, younger children, particularly those with delayed language skills, may find it challenging to complete a sample of sentence repetition, which in turn may impact both validity and reliability ratings. There is historic evidence to suggest that perceptual judgements of nasality are more reliable in connected speech, followed by single words, then isolated vowels (Counihan & Cullinan 1970). The importance of the speech sample is further

underlined by Sweeney (2011:206). She stated that, when measuring resonance and nasal airflow errors, there can be variability in 'speech performance between single words and conversational speech' i.e. spontaneous speech. As such single word samples used in isolation may impact the validity of measures of resonance and nasal airflow errors, key measures in the assessment of cleft speech outcomes.

There is an inherent challenge when assessing speech at age 3-years to balance a child's ability to complete a speech sample with eliciting a speech sample which allows for the comprehensive assessment of their speech to facilitate treatment planning, particularly with reference to parameters of speech which are cleft specific, such as hypernasality and nasal airflow errors. In order to make comparisons between children and report speech outcomes for a cohort at age-3 years, there is also a need to investigate which speech samples children this age can usually complete. This speech sample needs to be a valid representation of the child's speech and result in reliable listener judgements.

If the speech sample impacts the validity of the assessment of resonance and nasal airflow errors, consideration should be given as to whether there may be another parameter which could be used. For example, an overall measure of speech characteristics associated with velopharyngeal function could be considered. Pereira et al. (2021) proposed a composite measure of velopharyngeal function using the CAPS-A, CAPS-A VPC Sum. The CAPS-A VPC Sum is derived from ratings of hypernasality, nasal airflow errors, non-oral CSCs and passive CSCs. Pereira et al. (2021) reported that the CAPS-A VPC Sum was both a valid and reliable measure of velopharyngeal function, although this was validated with participants with a mean age of 20 years 2 months and using a speech sample comprised of conversational speech, rote speech and sentence repetition. There is an inherent requirement in such a composite measure that the ratings of hypernasality, nasal airflow errors, non-oral CSCs and passive CSCs are also reliable; however this is an issue when reliability has yet to be established with a younger age group i.e. age 3-years.

Although reliability has been previously established for the CAPS-A (John et al. 2006) for 5-year olds, it has yet to be established whether the CAPS-A can be reliably used with children younger than 5-years, and the impact a single word speech sample may have on the validity and reliability of both the CAPS-A and the CAPS-A VPC Sum. VPC-Rate is an alternative measure of velopharyngeal function (Lohmander et al. 2017c; Lohmander et al. 2009) which

has been used with 3-years olds in the validation of the SVANTE (Lohmander et al. 2017b). VPC-Rate uses a 3-point scale to estimate velopharyngeal function and has been reported to be a good predictor of velopharyngeal competence or closure (Lohmander et al. 2017c). Subsequently VPC-Rate has been adopted as a key outcome measure at age 3-years in the ToPS international trial (Shaw et al. 2019). However, further research is required to investigate how the reliability and validity of listener judgements on an overall measure of velopharyngeal function are impacted by different speech samples when assessing speech in 3-year old children.

1.3.2 Current Assessment Practices in UK Cleft Centres

All Cleft Centres in the UK carry out speech assessments at age 3-years (Wren 2013). The selection of 3-years as an additional timepoint to measure speech outcomes is also a pragmatic decision given that this would not result in additional appointments at the Cleft Centres, and thus the burden of care for both cleft patients, their families and Cleft Centres is not increased. It is also hoped that this would support the adoption of a subsequent assessment framework at age-3 years, and support comparisons at age-5 years given that this is the next timepoint at which all UK Cleft Centres complete a speech assessment. The ToPS trial has demonstrated a reliable method using naturalistic listening in which the prelinguistic vocalisations of 12-month-old children with CP were assessed (Willadsen et al. 2020). There is, therefore, the potential to extend timepoints at which speech outcomes are measured to include younger children in the future.

At the WMCLPS the primary aim of assessment at this age is to identify children who require therapy intervention and/or may need further investigations, active monitoring, or surgical intervention relating to symptoms of VPI. At present, there is no standard protocol for the assessment used in the WMCLPS, and unlike practice at age 5-years, no one protocol for assessment at age 3-years has been adopted by all UK Cleft Centres (Wren 2013). As reported by Peryer et al. (2021), the WMCPLS often uses the PACS-TOYS (Grunwell & Harding 1985) single word naming test to assess speech at age 3-years. However, this is not a cleft-specific assessment which is why SLTs at the WMCLPS tend not to complete the assessment in full, instead selecting specific target words 'ad hoc.' In addition to the PACS-TOYS other single word assessments Wren (2013) reported are used in UK Cleft Centres includes the CLEAR Phonology Screening assessment (CLEAR Resources 2006), and Phonological Screening Assessment (PSA) (Stevens & Isles 2001). An alternative to a single word assessment at age 3-

years is the Great Ormond Street Speech Assessment (GOS.SP.ASS) (Sell et al. 1999), a sentence-based assessment that is used in its original form or is modified 'ad hoc' into shortened sentences using the established words. Play-based assessments of spontaneous speech are also used and the decision as to which type of assessment to use is made by the individual SLT, considering the child's developmental level, attention skills, and other child-specific factors such as interests and whether they are outgoing or shy and reticent to speak. A limitation of these speech assessments is that, with the exception of the GOS.SP.ASS (Sell et al. 1999), they have not been designed specifically for the cleft population, and it is not known if these assessments result in the reliable and valid assessment of parameters specific to cleft speech.

Wren (2013) conducted a survey of all UK cleft services regarding speech assessment practices at age 3-years both in terms of the assessments used and therefore the speech sample upon which judgements are made. The survey indicated that practices across the country mirrored those at the WMCLPS, with many centres utilising a range of assessments, with no single assessment or speech sample favoured across the cleft services. The majority of the speech assessments reported to be used were not cleft specific and thus this variety may reflect the need of SLTs to use different assessments, potentially in combination, in order to assess the relevant parameters of speech for treatment planning. The use of multiple different assessments within and between Cleft Centres also precludes the comparison of children's speech outcomes within and across cleft centres or report on the outcomes of a cohort (e.g. children born in the same year). Essentially this means that at present UK cleft services cannot report valid and reliable speech outcomes before age-5 years.

1.3.3 Rationale for Developing a UK Assessment Protocol and Outcome Measure at Age 3-years

The development of a cleft specific speech assessment which allows for the valid and reliable assessment of speech outcomes at age 3-years in the UK would have several benefits. This would facilitate the early identification of children most 'at risk' of a poor speech outcome at age 5-years. Cleft centres could focus their resources on children most in need and develop specific care pathways for these children in relation both to therapy and the investigation and management of VPI. Sand et al. (2022) reported that young children with CP±L (younger than 6-years) benefit most from therapy intervention. The early assessment and detection of patients at age 3-years who are most in need of intervention through the

use of a cleft specific speech assessment may facilitate SLTs in treatment planning, potentially reducing the need for teams to use a combination of speech assessments and support timely access to therapy.

Currently there is limited research to support one therapy approach over another when treating children with CP±L (Sand et al. 2022). Indeed, in their systematic review, Bessell et al. (2013) concluded that there was a lack of evidence to support the age at which therapy should begin. Almost 10 years later, the closest specification of an age to begin therapy is "below 6-years of age" (Sand et al. 2022: 570). In their meta-analysis, Sand et al. (2022) discuss that one of the challenges of evaluating the effectiveness of an intervention is that outcomes have been measured on statistical significance rather than clinically relevant improvements and change. Having a valid and reliable outcome measure at age-3 years would enable cleft teams to track improvements between the ages of 3 and 5-years. For example, improvements in articulation outcomes during this time-period may provide useful information about the effectiveness of therapy intervention, facilitate future research evaluating the effectiveness of different therapy approaches, and provide much needed information as to rate of progress between 3 and 5-years. Such information could contribute to one of the key objectives for cleft lip and palate care identified by patient, parent, and clinician stakeholders as part of the James Lind Alliance, which the cleft community are yet to decisively answer: "In individuals with a cleft of the lip and/or palate when is the most effective age to begin speech therapy?" (James Lind Alliance 2012).

A valid and reliable cleft specific assessment at age 3-years would provide an earlier indicator of velopharyngeal function for speech. Currently, until children are aged-5-years, UK cleft surgeons must wait for valid and reliable speech outcomes before they receive feedback about a cohort of their patients and the success of the initial palate repair. Valid and reliable speech outcomes at age 3-years would shorten this cycle, providing valuable information to the surgical team about outcomes related to specific types and methods of primary palate repair. Additionally, if assessment at age 3-years indicates VPI and the need for either secondary speech surgery and/or fistula closure, valid and reliable speech outcomes could be compared at age-3 and 5-years. Thus, valuable information regarding the success of this surgical intervention would be facilitated.

Furthermore, for both therapy and surgical interventions, the accurate identification of clinically relevant improvements could expedite the current audit cycle, providing an earlier indication of speech outcomes, and meeting the recommendation set in the NHS Five Year Forward View, that the NHS needs to 'learn much faster from the best examples' (NHS 2014: 16). Indeed, the CCUK review hypothesises that "the creation of an audit culture that encourages reflective practice" may have driven improvements in UK cleft care (Ness et al. 2015). With an earlier audit point, age 3-years, examples of good practice could be identified earlier and could be shared with other cleft services. There is a need for both therapy and secondary surgery intervention in the preschool years in order to achieve the national audit speech standards at age 5-years (Britton et al. 2016; Britton et al. 2014) and having a valid and reliable speech assessment at age 3-years would support this. Because both therapy and surgical intervention may take place between age-3 and 5 years it would not be appropriate to apply the same national audit speech standards which are used at age-5 years (Britton et al. 2016; Britton et al. 2014) at age-3 years. Therefore, separate standards will need to be developed in order to audit speech outcomes at age-3 years. Whilst a number of developmental speech norms are available (McLeod & Baker 2017) which could be used to develop speech standards at age 3-years and would facilitate comparisons with typically developing peers, such standards would not be cleft specific nor support comparisons with the national standards at age-5 years. As described in Section 1.2.3 the process of the national audit of speech outcomes by UK Cleft Centres has driven improvements in cleft care (CRANE 2020; Sell et al. 2017; Sell et al. 2015). Thus, it is also important that any new standards should remain cleft specific, reflecting those parameters of speech particular to this population in order to maintain and improve standards of care for 3-year old children.

In the wider context of the NHS, making efficiency savings has been and is a top priority (Anandaciva 2022; Jabbal et al. 2018). This is particularly relevant given that almost half of NHS paediatric SLTs report that they do not have the time or resources to adequately support children with communication difficulties (Royal College of Speech and Language Therapists [RCSLT] 2018). This is relevant not only to SLTs based in regional cleft centres, but also community-based services which are frequently involved in the provision of therapy for children with CP±L (Williams et al. 2021). SLTs offering intervention for children with CP±L have cited significant challenges relating to shortages of staff, resources, and equipment which has negatively impacted therapy provision (Williams et al. 2021). This has implications

for the quality and frequency of provision offered to children with CP±L, particularly in community speech and language therapy settings. It also adds further weight to the argument for focussing limited intervention resources on those children most at risk of poor speech outcomes. These children could be identified by a timelier audit cycle at age 3-years. In addition, being able not only to identify, at age 3-years, those children most at risk of poor speech outcome, but also to demonstrate the impact of intervention before age-5 years, may support discussions with NHS commissioners for increased resource.

There is also an argument that it is 'too late' to report speech outcomes at age 5-years, and that this timepoint needs to be brought forward. There is evidence that very young children with CP±L already display signs of disordered speech development in comparison with their non-cleft peers (Zajac et al. 2021; Ha & Oller 2021) and to ensure that children begin their education with good speech outcomes, speech assessment and subsequent intervention must take place at an earlier age. Why then wait until children are aged-5 years to report speech outcomes? Indeed, the aim set out by the RCSLT (2009) that speech and language therapy intervention in CP±L should be to 'promote normal communication by school entry' necessitates an earlier speech outcome measure. This is further substantiated by the systematic review of the impact of speech impairment on life activities by McCormack et al. (2010), and the association with difficulties in building and maintaining personal relationships.

In addition to the benefits from a social perspective, ensuring that children with CP±L begin school with good speech outcomes, is likely to support their reading development. In a study of the early reading skills of both non-cleft children and children with CP±L, Chapman et al. (2011) reported a significant difference between reading performance between the two groups, with the lowest scores achieved by the group with CP±L. Letter-sound knowledge was also poorer in the CP±L group which is important because letter-sound knowledge is a strong predictor of reading skills (Hulme & Snowling 2014). Chapman et al. (2011) further reported that within the CP±L group, those participants who performed better on the early reading skills assessment also had a better speech outcome. The impact of a poor speech outcome at age 5-years is clear; as such there is an argument not simply to report on such outcomes at age 5-years, but to bring forward the timepoint at which speech outcomes are reported with a view to identifying and improving these outcomes before children start school.

Valid and reliable speech assessment at age 3-years would also facilitate cleft teams in monitoring the speech outcomes of children who had delayed palate repairs due to the Covid-19 pandemic. Brierley et al. (2022) reported that whilst UK cleft teams prioritised initial palate repair surgeries during the Covid-19 pandemic, the average age at palate repair increased by 67 days during the first UK lockdown (March 2020- April 2021), from 320 days to 387. Whilst this falls within the UK national standard upper age limit of 396 days, at the WMCLPS some palate repairs were completed outside of this upper age limit. The impact of such delayed palate repairs, and the further cumulative impact of the pandemic with reference to delayed speech and language therapy assessment and intervention and audiological assessment and treatment (Arnaout et al. 2022) on speech outcomes is not yet known. It is therefore of great importance that the outcomes of these children, whose care was significantly disrupted by the pandemic, is closely monitored. Valid and reliable speech assessments at age 3-years would facilitate this.

Despite the challenges posed by perceptual speech assessments and assessing speech in 3-year old children with CP±L, there is a need to report valid and reliable speech outcomes at age 3-years. A potential solution would be to adopt the SVANTE which has been validated for use in a Swedish population at age 3-years (Lohmander et al. 2017b; Lohmander et al. 2005) and adapt the speech samples into English. However, one of the advantages of reporting outcomes at age 3-years is the potential to compare speech outcomes at age-3 and 5-years. It would not be possible to directly compare outcomes reported on the SVANTE with the CAPS-A (John et al. 2006) which is used in the UK at age 5-years. Whilst the outcome measures include similar parameters of speech, they use different scalar values. The SVANTE also uses PCC to report articulation outcomes unlike the CAPS-A. Such differences would preclude straightforward comparisons thus diminishing the benefit of introducing a speech outcome measure at age 3-years. There is a need, therefore, not only to investigate which speech samples can be completed by 3-year olds and also result in valid and reliable listener judgements, but also if outcomes can be reliably reported using an outcome measure based on the CAPS-A at age 3-years.

Chapter 2. Study Aim and Objectives

2.1 Study Aim

To propose a speech assessment framework to assess speech outcomes in 3-year old patients with CP±L in the UK by determining the validity and reliability of different speech samples, rating methods and scales, and evaluating the acceptability and usability of the framework to SLTs.

2.2 Study Objectives

The six study objectives were completed in two defined phases of the study both with separate ethical approval processes.

Study Phase 1

1. To undertake a scoping review of the literature to inform the parameters of speech that should be assessed in 3-year old children with CP±L, and the types of speech samples to be included in the speech assessment framework.

Study Phase 2

2. To determine the extent to which 3-year old participants with CP±L can complete different speech samples, and how this is different from participants without CP±L (acting as a control group).

3. To determine the impact of different speech samples on the validity and reliability of listener judgements.

4. To ascertain the impact of different rating methods and scales on the reliability of judgements made by listeners of the speech characteristics associated with velopharyngeal function.

5. To gain further information regarding the specificity of the speech assessment framework by using the assessment with three-year-old participants without CP±L and

any known speech difficulties (acting as a control group) and examining listener judgements.

6. To measure the acceptability and usability of the speech assessment framework and rating methods to the SLTs who act as listeners.

Chapter 3. Study Phase 1: Scoping Review

A scoping review methodology was chosen given that the aim was to broadly explore and review existing literature and evidence concerning the practices used to assess speech at age-3 years in the CP±L population. This is in keeping with the scoping review methodology (Tricco et al. 2016). This information was summarised to inform future research, including this research study, by identifying gaps in the existing knowledge base. The findings of the review were used to inform Phase 2 of the study, and the subsequent selection and development of the speech samples, the parameters of speech included in the assessment framework and the need to further investigate the impact of different speech samples, rating methods and scales on listener reliability in speech assessments at age-3 years.

The scoping review (Fitzpatrick et al. 2020) was published in The International Journal of Language and Communication Disorders (2020, volume 5, issue 2, 165-187) with the following authors: Beth Fitzpatrick (BF), Jane Coad (JC), Debbie Sell (DS), and Tanya Rihtman (TR). The author of this thesis (BF) is the first author on the paper, other authors were part of the academic supervisory team. BF and TR reviewed the articles and designed the review. BF drafted the manuscript. All authors critically reviewed the paper, read and approved the final manuscript and agreed with the description of their contribution prior to publication (Appendix A). The scoping review is presented in full in Appendix B, included with the permission of the publisher (Appendix C). Below is a summary of the methods and key findings from the review.

3.1 Methods

The methods are detailed in the full appended review in Appendix B.

3.1.1 Ethical Approval

Ethical approval was gained in February 2018 from Coventry University (study number P68435, Appendix D, Appendix E). Literature searches were completed in March-April 2018 and the first version of the scoping review was completed in June 2018.

3.1.2 Design

The scoping review was undertaken as per the guidance set out by the Joanna Briggs Institute (JBI) (Peters et al. 2015) and was completed in a systematised fashion.

3.1.3 Scoping review objectives

- To explore the types of speech samples used in the assessment of 3-year old children with CP±L.
- To examine the parameters of speech typically assessed in 3-year old children with CP±L, and to consider if they are core to the assessment of individuals with CP±L.
- To investigate the methods and rating scales used to assess the identified parameters of speech.
- To discuss how the parameters of speech assessment map onto different theoretical approaches to assessment.

3.1.4 Inclusion criteria (as per the JBI format)

- Participants: 3-year old children with CP±L.
- Concept: speech assessments, the types of speech samples and speech parameters assessed, including assessment methods.
- Context: sources published after publication of the Great Ormond Street Speech Assessment (Sell et al. 1999). Sources had to be in English or have available translation.
- Types of sources: sources had to be sufficiently detailed to extract information about assessment procedures, thus conference abstracts were excluded.

3.1.5 Search Process and charting of the results

Search terms were generated in relation to age, diagnosis and the concepts under investigation. The following databases were used: EMBASE, Medline, Cumulative Index of Nursing & Allied Health Literature (CINAHL), AMED and PsychINFO and medical subject headings were used. The study selection process was defined by the PRISMA Scoping Review Extension (PRISMA-ScR) flowchart (Tricco et al. 2016) (Appendix B). A total of 35 sources were reviewed in full and included in the scoping review.

3.2 Summary of key findings

3.2.1 Speech samples and assessment tools used at age 3-years

The review highlighted that there was no single preferred assessment tool or speech sample used to assess speech in the CP±L population at age-3 years in the UK or internationally. However, three different types of speech sample were identified: single word naming, a spontaneous speech sample, and short sentence repetition. Single word naming was most commonly used and of the single word naming assessments used, the Restricted Word List (RWL) (Lohmander et al. 2009) was used most frequently in a total of eight sources

(29% of the original scoping review articles) (Willadsen et al. 2018; Raud Westberg et al. 2017; Klintö et al. 2016; Klintö et al. 2015; Klintö et al. 2014a; Klintö et al. 2014b; Willadsen 2012; Willadsen & Poulsen 2012) all of which were associated with the Scandcleft Trial (Lohmander et al. 2017b; Willadsen et al. 2017) and originated from Scandinavia. The RWL was designed specifically for cross-linguistic assessment speech assessments and comparisons with several iterations available in seven different languages (CLISPI 2017). However, a limitation of this speech sample is that it was not developed to comprehensively assess all consonant sounds.

Spontaneous speech samples were used in 37.1% of sources, often in addition to a single word naming sample. In the UK, 64.3% of cleft teams reported that they used spontaneous speech samples at age-3 years (Wren 2013). Different nations had different preferences for speech samples. Chacon et al. (2017) from Australia was the only study outside of the UK to use a speech sample of sentence repetition. However, 85.7% of UK cleft teams reported using a speech sample of sentence repetition, specifically the GOS.SP.ASS (Sell et al. 1999), at age 3-years, highlighting this as a speech sample of importance in the UK at this age.

3.2.2 Speech Parameters included in assessment at age-3 years

The studies in the scoping review confirmed the inclusion of key parameters of speech for assessment at age-3 years: consonant production, resonance and NAE, and also highlighted the potential inclusion of a measure of velopharyngeal function at age-3 years. Consonant production was most frequently assessed in the studies reviewed. The review also highlighted the importance of considering developmental speech immaturities and phonological processes at age-3 years, taking into account the individual's developing speech sound system at this age.

Hypernasality and NAE were frequently assessed in the studies, whereas hyponasality was less routinely assessed, and was never assessed in the absence of hypernasality (Lohmander et al. 2017b, Swanson et al. 2017; El Ezzi et al. 2015; Klintö et al. 2014b; Wren 2013; Hamming et al. 2009; Persson et al. 2006; Chapman 2004; Lohmander-Agerskov 1998; Lohmander-Agerskov et al. 1998). An overall measure of velopharyngeal function featured in a number of studies (Lohmander et al. 2017b, Larsson et al. 2017; Swanson et al. 2017; El Ezzi et al. 2015; Dayashankara et al. 2011; Hamming et al. 2009; Lohmander et al. 2006; Persson et al. 2006; Zanzi et al. 2002; Gunther et al. 1998).

The importance of assessing intelligibility and voice is less clear based on the scoping review, as there was a discrepancy between the frequency with which these parameters were assessed in the studies included in the review, and how frequently they were used by UK cleft teams in their assessments at age-3 years. Intelligibility was assessed in 11 studies (Frey et al. 2018; Lohmander et al. 2017b; Safaiean et al. 2017; El Ezzi et al. 2015; Wren 2013; Willadsen & Poulsen 2012; Dayashankara et al. 2011; Hodge & Gotzke 2007; Lohmander et al. 2006; Lohmander-Agerskov 1998; Lohmander-Agerskov et al. 1998). However, less than half of the UK cleft teams reported assessing intelligibility at age-3 years (Wren 2013). There was also a difference in the frequency with which UK cleft teams reported that they assessed voice, 78.57%, compared to only 20% of the studies in the review (Swanson et al. 2017; El Ezzi et al. 2015; Wren 2013; Hamming et al. 2009; Gugsch et al. 2008; Gunther et al. 1998; Lohmander-Agerskov 1998).

3.2.3 Assessment methods and scales

Two principal methods to assess consonant production were used, both based on phonetic transcription- summary patterns (errors by place of articulation, passive or active characteristics) and PCC. Whilst the summary patterns focused on cleft specific patterns of articulation, PCC captured information not only relating to cleft specific articulation, but also developmental and disordered speech production. Of note was that UK studies did not use PCC in the assessment of 3-year olds (Wren 2013; Hattee et al. 2001).

A combined measure of NAE, in which nasal emission and nasal turbulence were combined into a single scale as per the SVANTE (Lohmander et al. 2017b; Lohmander et al. 2005) and CAPS-A-AM (Chapman et al. 2016) protocols, featured in some of the studies (Larsson et al. 2017; Lohmander et al. 2017b; Swanson et al. 2017; Lohmander & Persson 2008; Lohmander et al. 2006). This highlights the potential for a single measure of NAE, in which nasal emission and nasal turbulence are combined, at age 3-years.

The review highlighted the potential for broader assessments of velopharyngeal function at age 3-years, as such measures featured in several sources (Larsson et al. 2017; Lohmander et al. 2017b; Swanson et al. 2017; El Ezzi et al. 2015; Dayashankara et al. 2011; Hamming et al. 2009; Lohmander et al. 2006; Persson et al. 2006; Zanzi et al. 2002; Gunther et al. 1998). There was no preferred measure of overall velopharyngeal function with a mix of rating scales, composite scores, and clinical diagnoses used. Whether such a broad measure should

be used to complement or replace existing parameters has not been fully investigated at age 3-years. Categorical scales were most frequently used at age-3 years and the validity and reliability of alternatives to categorical scales has yet to be examined in 3-year old children.

3.2.4 Theoretical approaches to assessment

A linguistic approach to speech assessment, underpinned by phonetic transcription was used in the majority of sources. However, a developmental approach to assessment was often used, which took into consideration the developing speech sound system, highlighting the importance of a combined approach to assessment at age-3 years. The Wren (2013) survey undertaken in the UK indicated that only 50% of cleft teams assessed phonology at age-3 years. This suggests that UK cleft teams may concentrate on cleft specific speech outcomes. However, this limits the information gained from assessments at age- 3-years. Studies featured in the review by Chacon et al. (2017) and Hutters et al. (2011) reported that children with CP±L presented with more developmental phonological processes than their non-cleft peers. The scoping review suggests that both linguistic and developmental approaches should be used in speech assessments at age 3-years in order to assess speech from both a cleft and developmental perspective.

3.2.5 Reporting reliability

In their 2004 review of cleft palate literature, Lohmander & Olsson (2004) criticised the infrequency with which the reliability of listener judgements were examined or reported, concluding that reliability should be routinely reported in studies presenting speech outcomes. In total 60.0% (n=21) of the studies in the review reported on listener reliability (either inter-rater, intra-rater or both). Of those not reporting reliability, one study was a questionnaire, for which reporting reliability was not appropriate, another used computer bases analysis and thus did not report listener reliability, and two studies used consensus listening in which listeners reach an agreement about a speech outcome.

Inter-rater reliability for hypernasality varied across the studies reviewed. Persson et al. (2006) reported reliability in terms of complete agreement which was 44-55%. Lohmander & Persson (2008) reported poor inter-rater agreement for hypernasality, with 39% agreement in their study of 3–7-year-olds. In comparison, Raud Westberg et al. (2017) report higher reliability, with 80% agreement for hypernasality. Similar to inter-rater reliability, intra-rater reliability for hypernasality also varied across the studies. For example, Lohmander et al.

(2017b) reported that intra-rater reliability was very good in 42%, good in 54% and moderate in 4% of comparisons, whilst Chapman et al. (2008) reported 'very high' levels of correlation between the first and second listener judgements.

In contrast to hypernasality, when nasal emission and nasal turbulence were combined into a single measure, consistently strong inter-rater reliability was reported (Larson et al. 2017; Raud Westberg et al. 2017; Lohmander & Persson 2008; Lohmander & Persson 2008; Persson et al. 2006). For intra-rater reliability, Persson et al. (2006) reported 72-88% agreement, and Raud Westberg et al. (2017) reported 97.5%. This suggests that the use of a combined measure of NAE should be used in speech assessments at age-3 years to support listener reliability.

For articulation outcomes, inter-rater reliability was often reported as point-by-point transcription agreement using percentage agreement rather than by using correlation or kappa scores. Agreement ranged from 70%-97.5% (Willadsen et al. 2018; Chacon et al. 2017; Larsson et al. 2017; Raud Westberg et al. 2017; Klintö et al. 2016) indicating consistently strong inter-rater agreement across the studies. Intra-rater agreement was similarly strong across the studies, ranging from 80%-97.5% (Willadsen et al. 2018; Chacon et al. 2017; Larsson et al. 2017; Raud Westberg et al. 2017; Klintö et al. 2016).

In summary, the reported reliability ratings of those sources included in the review indicate that it is possible to reliably assess speech outcomes at age 3-years in the CP±L population. Although the use of different rating scales, numbers of listeners and statistical tests makes comparisons of reliability challenging, such ratings provided a benchmark and a comparator for the subsequent investigation of reliability in Phase 2 of this study.

3.3 Application of scoping review to Phase 2 of the study

The key findings from the scoping review directly informed the research objectives and methods undertaken to answer these within Phase 2 of the study, specifically the development of two speech samples for use with 3-year olds with CP±L and a proposed assessment framework for measuring speech outcomes at age-3 years. The following conclusions, listened below, were made from the scoping review and informed the objectives of Phase 2 of the study.

- There is no single speech sample which is favoured internationally for use in speech assessments. (Objective 2, Objective 3)
- Completion rates of different speech samples have not been reported. (Objective 2, Objective 5)
- There is no clear evidence to indicate if a particular speech sample results in more reliable listener judgements. (Objective 3)
- That there is a need to develop and compare the reliability of two different speech samples, taking into consideration that single word naming is most widely used internationally to assess speech, but that UK cleft teams show a preference for sentence repetition. (Objective 3)
- There is a need to measure speech outcomes from both a cleft and developmental perspective. (Objective 4)
- The use of a combined measure of velopharyngeal function for speech should be further investigated. (Objective 4)
- There is a preference in the literature to combine nasal emission and nasal turbulence into a single measure of NAE. (Objective 4)
- That the use of alternatives to ordinal scales to measure speech outcomes has not been investigated. (Objective 4)

3.4 Update to the scoping review

The scoping review (Fitzpatrick et al. 2020) directly informed Phase 2 of the study. However, for completeness and to reflect the most current literature and knowledge, an updated search was undertaken in August 2022. The results of the updated review therefore were not used to inform the methods used in Phase 2 of the study; however this information was used in the discussion of the results in Phase 2, and to inform directions for future research.

3.4.1 Methods

The updated review used the same search criteria and methods as the original scoping review but only the thesis author (BF) reviewed these additional sources. The timeline for inclusion in the review was also modified to include sources published from 2019, the year in which the original manuscript for the scoping review was submitted for publication. Figure

3.1 shows the PRISMA Scoping Review Extension (PRISMA-ScR) flowchart (Tricco et al. 2016) detailing the source review and selection process for the update to the review.

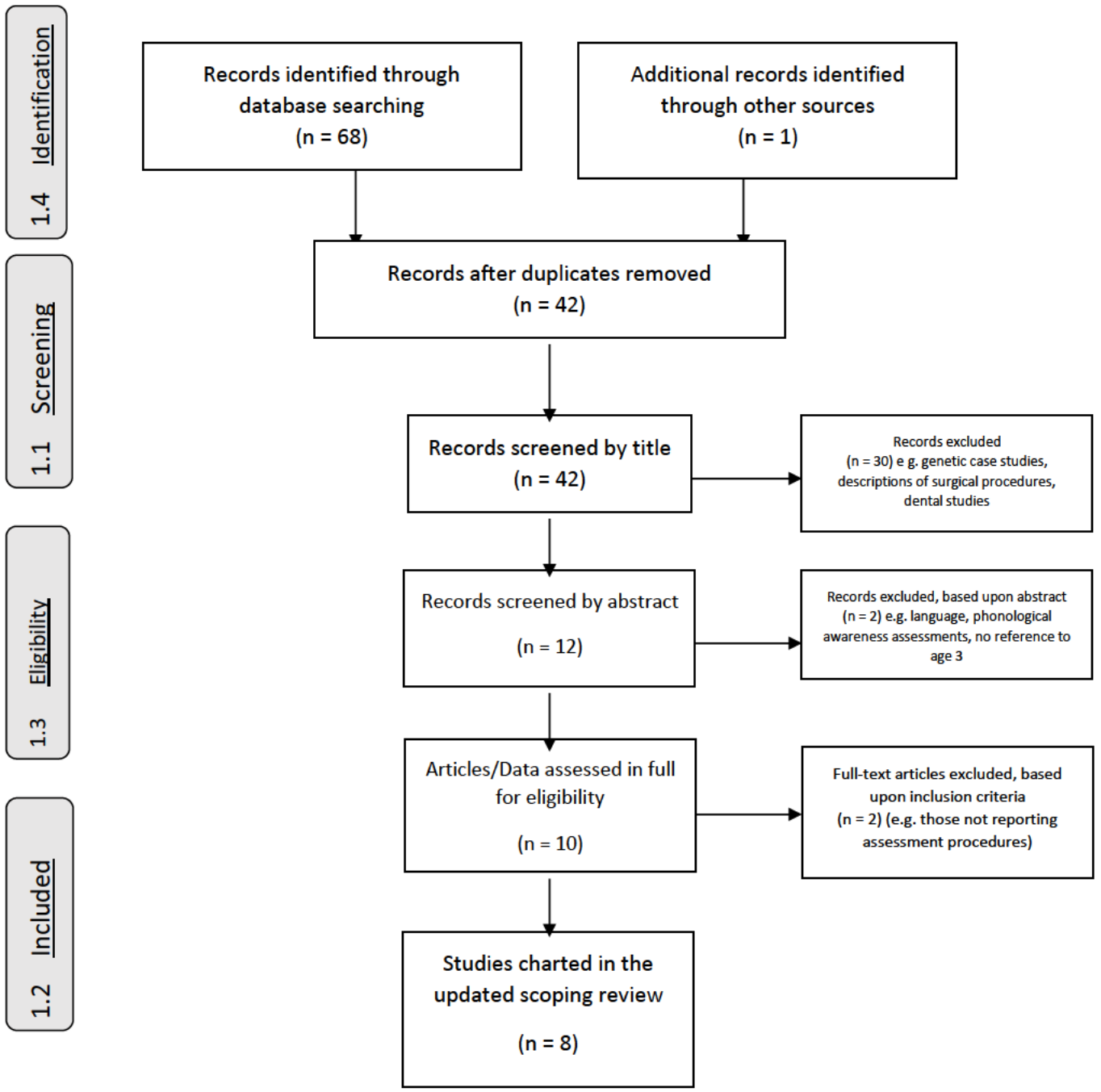


Figure 3.1. Adapted PRISMA Flow Diagram (Moher et al. 2009) showing the article selection process.

3.4.2 Demographic Information

The updated review identified an additional eight studies, all of which were published following the submission of the scoping review for publication. Details of the additional eight studies are presented in Table 3.1 and Table 3.2 (in the same format as Fitzpatrick et al. 2020). Table 3.1 outlines the articles by author and date, country of origin, methods and methodology, speech assessment, with additional information regarding reliability recording. Table 3.2 summarises the parameters of speech assessed in the studies.

The additional articles originated from five different countries, including two countries that did not feature in the original scoping review, Israel and Belgium. Thus, updating the review has extended the understanding of current assessment practices, giving a more international perspective. Three of the eight articles (37.5%) originated from the UK, although one of these, Shaw et al. (2019) details the protocol for an international Randomised Control Trial (RCT), the ToPS Trial. This study builds on the methods of speech assessment used in the Scandcleft Trial (Lohmander et al. 2017a, 2017c; Willadsen et al. 2017) much of which was reflected in many of the studies in the original review (Willadsen et al. 2018; Raud Westberg et al. 2017; Klintö et al. 2016; Klintö et al. 2015; Klintö et al. 2014a; Klintö et al. 2014b; Willadsen 2012; Willadsen & Poulsen 2012).

3.4.3 Key Results

At the time of thesis submission, no other study of 3-year olds with CP±L has compared the reliability of two different speech samples, nor reported on completion rates of the speech samples or investigated the potential use of an alternative to ordinal scales, as was investigated in Phase 2 of this study. Indeed, the results of the additional eight sources support the original conclusions of the scoping review and subsequent methodological decisions in Phase 2.

Examining the speech samples used in the additional eight sources, the results are in keeping with the original scoping review. Overall, single word naming or repetition was used most frequently in six of the studies (75%) (Rezaei et al 2022; Nachmani et al. 2021; Peryer et al. 2021; Zajac et al. 2021; Jorgensen & Willadsen 2020). Specifically, RWLs were used by both Jorgensen & Willadsen (2020) and Shaw et al. (2019).

Consistent with findings from the original scoping review, the same parameters of speech were frequently utilised in the additional studies. Except for the study by Seifert et al.

(2021), all the other studies assessed/reported on consonant production (n =7, 88%) and 50% (n=4) assessed phonology. Only 37.5% (n=3) of the additional studies assessed hypernasality and NAE. Two of the studies that did not report on hypernasality or NAE (Jorgensen & Willadsen 2020; Shaw et al. 2019) used an overall measure of velopharyngeal function. This suggests that at age-3 years an overall measure of velopharyngeal function rather than separately evaluating the parameters of speech indicative of this i.e. resonance and NAE, may be sufficient.

Bruneel et al. (2020) reported on the reliability of listener judgements using a Belgian Dutch Speech Outcome tool based on the CAPS-A. The authors reported that 20 children aged 3-10 years completed a speech sample comprised of spontaneous speech, automatic speech, and short sentence repetition. Single word naming or repetition was not used. However, the mean age in the cleft group was 6.5 years. No information is provided about how many 3-year olds participated in the study, nor how many of them were able to complete all aspects of the speech samples. Similarly, Nachmani et al. (2021) reported the use of several speech samples (repetition of sustained vowel phonation and consonant-vowel constructions, repetition of words and sentences, counting 1-20, spontaneous conversation (5 minutes), and a Hebrew articulation and phonology test). This was a retrospective study across a variety of ages including a subgroup of 134 children aged 3-4 years (mean 3.7 years). No information was provided as to whether all these speech samples were used in the analysis, and once again, no information was provided about completion rates of these speech samples by the 3-4 year-age group. Indeed, completion rates of the different speech samples were not reported in any of the studies.

In the original scoping review, it was suggested that the Intelligibility in Context Scale, (ICS) (McLeod et al. 2012) which uses parent/carer ratings of intelligibility may be well suited for use with the cleft population but had yet to be validated for use with a cleft population. The ICS is a valid and reliable parent or carer reported measured (for the non-cleft population) in which parents/carers report their child's intelligibility with different communication partners, validated and tested in 14 languages (McLeod 2020). Since publication of the original review, Seifert et al. (2021) have developed norm scores for the ICS within the CP±L population, specifically at age 3-years. This is the only study in either the updated or original scoping review to use a parent/carer reported outcome measure and constitutes a significant

development for measuring intelligibility outcomes at age-3 years in the CP±L population. Indeed, in the International Classification of Functioning, Disability and Health: Children and Youth Version (ICF-CY; World Health Organization [WHO] 2007), intelligibility is reported to be influenced by both contextual (environmental factors) as well as production factors (bodily functions) (McLeod et al. 2012: 649) which are arguably both captured in this parent/carer outcome measure. ICHOM (Allori et al. 2017b) have included the ICS in their recommended outcome set from age-5 years and above. Data from the Seifert et al. (2021) study indicates that the age at which the ICS is recommended by ICHOM could be brought forward, with the availability of the norm scores developed. This is particularly important because ICHOM brings to the fore the importance of Patient/Parent Reported Outcome Measures (PROMs), and movement away from the sole reliance on clinicians to evaluate the success or failure of interventions (Apon et al. 2021). Indeed, Sell & Pereira (2015) report that non-expert listeners are more valid judges of speech intelligibility. As such the ICS, a parent/carer outcome measure, should be considered in the future assessment of speech outcome measures at age-3 years.

Also, in keeping with the original scoping review was the prevalence of ordinal or dichotomous scales, and alternative scales such as VAS or the BORG cM scale did not feature. Whilst listener reliability was reported in 62.5% (n=5) of the studies (those not reporting reliability were retrospective studies or used the parent/carer outcome measure which would not have been appropriate), no studies compared listener reliability across different speech samples.

3.4.4 Conclusions based on the updated review

The aim and objectives of the present study remain highly relevant as no other studies to date have compared listener reliability when analysing different speech samples at age 3-years, nor have proposed an assessment framework for reporting speech outcomes in English at age-3 years. The overall conclusions of the original scoping review are also further substantiated by the sources in the updated review.

Table 3.1. Articles included in the review recorded by author and date, country of origin, methods and methodology, speech assessment, and information regarding reliability recording

Author & Date	Country of Origin	Age of Participants	Cleft Type*	Methodology and Methods	Speech Assessment	Reliability Reporting
Rezaei et al (2022)	Iran	3-year old- 7-year-old	UCLP BCLP CP SMCP	Retrospective Case Analysis	Single word repetition (Persian) Sentence repetition (Persian)	None
Nachmani et al. (2021)	Israel	Range of ages that included 3-4 years (Mean 3.7 years)	UCLP BCLP CP SMCP Occult SMCP	Retrospective Case Analysis	Repetition of sustained vowel phonation and consonant-vowel Repetition of words and sentences Counting 1-20 Spontaneous conversation (5 minutes) Hebrew articulation and phonology test (non-standardised)	✓ Inter and Intra Rater Reliability reported
Peryer et al. (2021)	UK	18 months 3-years 5-years	BCLP	Retrospective Case Analysis	Single word naming (PACS TOYS [Grunwell & Harding 1985])	None
Seifert et al. (2021)	UK	3-years	Cleft Lip only CP (syndromic/non-syndromic) UCLP BCLP	Prospective cohort study	Intelligibility in Context Questionnaire (completed by parent/carer)	N/A
Zajac et al. (2021)	USA	36 months (Range 34-41 months)	UCLP CP NCCG	Prospective cohort study as part of a larger research study	Goldman-Fristoe Test of Articulation Third Edition (GFTA-3; Goldman & Fristoe 2015) Spectral analysis of the sounds /t, k, s, ʃ/	✓ Inter and Intra Rater Reliability reported
Bruneel et al. (2020)	Belgium	3-10-year-olds	Various cleft types	Prospective reliability study	Spontaneous speech Automatic speech: counting 1-10, days of the week Short sentence repetition	✓ Inter and Intra Rater Reliability reported
Jorgensen and Willadsen (2020)	Denmark	3-year olds 5-year olds	UCLP	Subgroup analysis within larger multicentre Randomised Control Trial	Single word naming using the naming test developed in the Scandcleft study (Lohmander et al. 2009)	✓ Inter and Intra Rater Reliability reported
Shaw et al. (2019)	UK	12 months 3-years 5-years	CP	Randomised Control Trial	Single word naming based on the single word naming using the naming test developed in the Scandcleft study (Lohmander et al. 2009) (30 words) Spontaneous/continuous speech sample	

*UCLP= Unilateral Cleft Lip and Palate; BCLP= Bilateral Cleft Lip and Palate; CPO= Cleft Palate Only; SMCP= Submucus cleft palate; CPL= Cleft Palate +/- Cleft Lip (used when specific type of cleft is not stated); NCCG= Non-Cleft Control Group

Table 3.2. Summary of the parameters of speech assessed. Articles that utilised Percentage Consonants Correct were categorised as assessing both articulation and phonology.

Author & Date	Parameters of assessment							
			Resonance					
	Articulation	Phonology	Hypernasality	Hyponasality	Nasal Airflow Errors (NAE)	Velopharyngeal Function	Intelligibility	Voice
Rezaei et al (2022)	✓		✓		✓ Nasal Emission (none/present) Nasal Turbulence (none/present)			
Nachmani et al. (2021)	✓	✓	✓ (present/absent)		✓ Nasal Emission (present/absent)	✓ (but not perceptually, instead using instrumental analysis)		
Peryer et al. (2021)	✓	✓						
Seifert et al. (2021)							✓ Intelligibility in Context Scale	
Zajac et al. (2021)	✓							
Bruneel et al. (2020)	✓		✓	✓	✓ Nasal Emission Nasal Turbulence		✓ Understandability and Acceptability	✓
Jorgensen and Willadsen (2020)	✓	✓				✓ VPC-Sum		
	PCC- Obstruents: obstruent sounds specifically which were adjusted for age CSCs and Developmental Sound Characteristics							
Shaw et al. (2019)	✓	✓				✓ VPC-Rate Symptoms of VPI		
	PCC Percent correct placement Percent correct manner Nonoral errors Oral Consonant Errors							

Chapter 4. Study Phase 2: Methods- Development and testing of the Speech Assessment Framework

Phase 2 of the study relates to the completion of study objectives two-six, the development of the study protocol and associated ethical approvals. The study protocol is a full description of the research study which served as a 'manual' for the methods of the research study. Ethical approval for Phase 2 was gained from Coventry University (February 2018, June 2018) NHS Ethics, and the HRA (January 2019). Participant recruitment for the study opened on 06/03/19 and the final participant was recruited on 15/10/19. Data analysis began in January 2020. The key stages in the methods of Phase 2 are summarised in Figure 4.1 and are described in this chapter.

Figure 4.1 Phase 2: Key stages of Phase 2 Methods

4.1 Approval Process

The national Cleft and Craniofacial Conditions Clinical Studies Group is responsible for developing and supporting cleft research in the UK. The group is comprised of subject experts (cleft clinicians) and consumer representatives (e.g. service users and family members). A subject expert and a consumer representative reviewed the study protocol and provided feedback including suggested changes in June 2018. Suggested changes made by the reviewers and any subsequent changes to the protocol are detailed in Appendix F.

The protocol was then reviewed by Coventry University Ethics (P66325, Appendix G, Appendix H) and NHS Research Ethics (18/EM/0253, Appendix I) and Health Research Authority (242296, Appendix J), who provided further feedback which was incorporated into the study protocol, version 3.3. This version was approved by Coventry University, NHS Research Ethics, the Health Research Authority, and subsequently by the various NHS Trusts involved in the study. All 3-year olds participating in the study were recruited from a single cleft team, the WMCLPS, and ethical approval was granted by the NHS Trust (Birmingham Women's and Children's NHS Foundation Trust [18_BC_HNS_NO_134]). SLTs analysing the speech samples were recruited from five NHS Trusts and ethical approval was granted from all these trusts (Birmingham Women's and Children's NHS Foundation Trust [18_BC_HNS_NO_134], Cambridge University Hospitals NHS Foundation Trust [A095320], Newcastle upon Tyne Hospitals NHS Foundation Trust [09232], NHS Greater Glasgow and Clyde [GN19OD3009/242296)], Salisbury NHS Foundation Trust [P66325]). The study was then carried out following the final version of the protocol as described below.

4.2 Development of the Speech Samples

The scoping review highlighted two speech samples of importance at age-3 years; first a speech sample comprised of a single word naming test and a sample of spontaneous speech; and second a sample of sentence repetition. Internationally, the scoping review indicated that samples of single word naming are most common in speech assessments at age-3 years in the CP±L population, often in combination with a sample of spontaneous speech (Fitzpatrick et al. 2020). However, it was also important to consider typical UK practice, which underscored the importance of examining speech samples based on sentence repetition at this age (Wren 2013), as the overall aim of this study was to propose an assessment for use in the UK. The reliability of listener judgements for both speech samples, and completion rates of the speech

samples were compared in order to determine if one speech sample should be recommended over another in the speech assessment of 3-year olds with CP±L.

4.2.1 Speech Sample A: Single Word Naming and Spontaneous Speech Sample

Single word naming was selected to identify errors in speech sound production using phonetic transcription. This was used in combination with a sample of connected spontaneous speech so that judgements regarding the adequacy of palatal function for speech, specifically an overall judgement of velopharyngeal function, hypernasality, hyponasality, and NAE could be made. The rationale for the combination of single word naming and spontaneous speech into a single speech sample is based on evidence that nasality judgments are more reliable in connected speech (Counihan & Cullinan 1970) and because the perception of nasality and NAE may vary between single words and connected speech (Sweeney 2011:206). Indeed, CLISPI (2017) recommend that single word speech samples should be used alongside samples of spontaneous speech. Therefore, a combination of a spontaneous speech sample and single word naming was selected, referred to as Speech Sample A.

For assessment of spontaneous speech, a sample of two minutes of accumulative speech was collected following the recommendations made by CLISPI (2017), and with reference to Sell & Grunwell (1990) and Klintö et al. (2011) who both utilised spontaneous speech samples of two minutes in length. These two minutes of spontaneous speech were accumulatively based on the ratio of child to adult speech. The West Midlands Patient Voices Cleft Lip and Palate Association Group also provided ideas regarding suitable toys and activities to elicit the speech sample. The list of toys and activities used to elicit the spontaneous speech sample is detailed in section 4.4.5.iii.

For assessment of single words, the Restricted Word List (RWL) was selected forming the basis of the single word speech sample used in this study. The RWL was developed for use in the Scandcleft Trial (Lohmander et al. 2009), is recommended by CLISPI (2017) for use in single word assessments and was subsequently adopted in the ToPS International Trial (Shaw et al. 2019). The RWL was chosen for two primary reasons, firstly the scoping review highlighted the suitability of this speech sample for use at age-3 years, based on several studies, and secondly because the RWL facilitates the comparison of speech outcomes across

languages. There are currently seven RWLs available in different languages, one of which is the English version, with the potential to develop more RWLs using the same phoneme set (CLISPI 2017). The benefit of using a RWL is that speech outcomes can be compared internationally across other languages, examples of best practice can be more easily identified, and that these speech samples can be used in international research studies.

In the English RWL there are thirty words. Target phonemes are selected which appear cross-linguistically in several languages; however, being 'restricted' the sample does not include the full range of oral pressure consonants in English. The English RWL assesses one nasal phoneme /n/, and nine pressure consonants: six plosives /p, b, t, d, k, g/ and three fricatives /f, v, s/. Nine of these phonemes are assessed in word-initial (WI) position, except for /s/ which is assessed in word-final (WF) position only. For this study, an additional word list to be used in conjunction with the RWL was developed in order to assess a wider range of consonants not included in the RWL i.e. /m, ŋ, l, z, ʃ, tʃ, dʒ, h/ and the consonant clusters /sp, sn, sl/, and also consonants in word-final position. This would ensure a more comprehensive assessment of the 3-year old's sound system. Word-final position was included because of evidence that consonants in WF position can be easier for children with CP±L to achieve, because the vowel supports the use of an oral airflow and easier vowel-consonant blending (Russell & Albery 2017). A further benefit of assessing WF position is that this supports the assessment of both final consonant deletion and consonant harmony, both developmental speech sound processes which may be occurring at age 3-years (Dodd et al. 2003), and which should be considered in phonological assessments.

Whilst McLeod and Baker (2017) recommend that a word list of 100 words is utilised in single word speech assessments, they do not specify the age at which they would anticipate a child to complete an assessment of this length. Taking into consideration the age of the children in this study (3-years), it was important to consider the number of target words, and thus the length of the assessment. This was also important because typically speech assessments are completed in UK cleft centres in a single assessment session (Wren 2013), as families can often live at considerable distance from their regional cleft centre. As such, for practical and pragmatic considerations, the aim was for speech assessments in this study to be completed in a single assessment session. It was, therefore, important to strike a balance of designing speech samples that would provide maximal clinical information, but which could

also be completed by a 3-year old in a single assessment session lasting approximately one hour.

Several factors had to be considered to achieve this. Although Lohmander et al. (2009) and the CLISPI group (2017) recommend in the design of cross-linguistic word lists that target words should only contain the target pressure consonant, the practicalities of designing an assessment in which the target words would be familiar to 3-year old children resulted in the inclusion, within the original RWL in English (Lohmander et al. 2009), of some words containing more than one pressure consonant. To keep the number of words added to the original English RWL at a minimum, words that already contained more than one consonant were identified for inclusion in the new word list in this study (Table 4.1), hereafter known as The Combined Word List. For example, in the word 'feet' the target consonant in the RWL is /f/, however, /t/ was also identified as a target consonant reducing the necessity to include another word to assess WF /t/. This increased the number of phonemes assessed within the original English RWL from 10 different phonemes, each assessed 3 times as described by Lohmander et al. (2009), to 13 different phonemes each appearing 1-4 times without having to extend the number of words in the sample. In this way the English RWL was not changed, allowing this sample to continue to be used in cross-linguistic comparisons whilst also gathering maximal clinical information from the sample. It was not necessary to assess all additional phonemes identified in the original English RWL, for example, 'bus' was used to assess WI /b/ but not WF /s/ because WF /s/ was already assessed three times.

Despite identifying additional phonemes in the original English RWL all the relevant phonemes were not included, and thus additional words were added to the Combined Word List. The recommendations made by CLISPI (2017) regarding the design of single word assessments for use with the CP±L population were followed when selecting the additional words, as explained in Table 4.2. Specifically, high vowels were included in the Combined Word List to facilitate ratings of hypernasality given that these vowels are more susceptible to hypernasality (Dubey et al. 2018).

Table 4.1. Phonemes that appear in The Restricted Word List (Lohmander et al. 2009).

WI= word initial position, WM= word medial position, WF= word final position

Table 4.2. CLISPI (2017) Recommendations for single word assessments

CLISPI (2017) Recommendation	Combined Word List
Assesses all pressure consonants	The following pressure consonants were assessed: /p, b, t, d, f, v, s, z, ʃ, tʃ, dʒ, k, g/ Each pressure consonant was assessed x4 times in a variety of word positions (WI or WF). There is a consensus in speech development norms that /θ,ð/ are not acquired by age 3-years, and these sounds constitute sounds that are acquired last (McLeod & Baker 2017), hence why they were not included in the sample.
Assesses all or some low-pressure consonants	/l, h/ were assessed x4.
Assesses one or more nasal consonants	/m, n/ were assessed x4 /ŋ/ was assessed x1
Assessment includes assessment of all or some high and low vowels and some non-high/non-low vowels	The following high vowels were assessed: /i, ɪ, u, ʊ/ (the production of /ʊ/ may be accent dependent) The following low vowels were assessed: /a, æ/ Other vowels included a range of mid vowels.
High vowels should occur in approximately ten of the words which also have a test consonant in a 'strong position,' (which is defined as the position a test sound is in which is most distinctly articulated, most easily recognisable, and minimally influenced by context).	More than ten words had a high vowel and assessed a single test consonant in a strong position; this also facilitated the assessment of hypernasality (Lohmander et al. 2009).
Other pressure consonant sounds should be avoided within a word	25.86% of the words were CV structures and therefore only contained a single consonant (x2 examples were nasal-consonant constructions) 5 words contained the same consonant sound twice.
The consonant should occur in a 'strong position'	Test consonants did not always appear in a stressed position as analysis of consonant production in word-final position often involves the target consonant occurring in an unstressed, non-strong position.
Nasal consonants should be avoided within a word	No pressure consonant was assessed in a word containing a nasal consonant.
Vowels of different heights should be avoided within a word	84.74% of the words only contained one vowel (59.32% CVC structures and 25.42% CV structures) thus avoiding other vowel heights within a word.
Clusters with minimal loading (i.e. oral low-pressure consonants) should be assessed.	/sl/ was assessed in 'sleep.' McLeod et al. (2001a, 2001b) reported that at age 3 years children may be able to produce word-initial clusters containing /l/ or /s/. In addition, this vocabulary item is frequently featured in the first words of children speaking languages such as English, Polish (Rescorla et al. 2017), and Italian (Rescorla et al. 2014).
Clusters with maximal loading (i.e. nasal consonants) should be assessed.	/sn/ was assessed in 'snail.'
Other clusters should be assessed.	/sp/ was assessed in 'spoon.'
The test consonants and subsequently the words should be randomly ordered	The original word order of the RWL has been retained. Words added to this have been ordered as described below, to prioritise those sounds not assessed in the RWL.

Whilst the CLISPI recommendations refer specifically to the phonetic structure of the word list, it was also important to consider the items for inclusion on a lexical basis. McLeod & Baker (2017) described how most speech assessment tools have been developed in WEIRD societies (Western, Educated, Industrialised, Rich, Democratic). Whilst the speech samples were designed to assess speech in English, it was important to consider the language and cultural backgrounds of the children for whom the assessment was designed, and whether the lexical items selected were appropriate.

Lohmander et al. (2017b) advise that the words selected in a speech sample need to be well known to children at age-3 years. School pupils from the main city in which the WMCLPS is located speak 108 different languages and come from 87 different ethnic groups (Birmingham Community Safety Partnership 2013). The Department of Education also reports that 20.6% of primary school children in the UK speak English as an additional language (Department for Education 2017). To encourage independent naming, it was therefore important to ensure, as far as possible, that words included in the Combined Word List would be well known to all 3-year olds, not just those from certain language or cultural backgrounds.

To address the challenge of developing a target speech sample that could be used by a diverse population, consideration needed to be given to the phonetic requirements of assessment, the cultural experiences and language backgrounds of the children for whom the assessment was intended and the requirements of SLTs to complete a thorough speech assessment in a time-efficient manner. Several studies have investigated cross-linguistic differences and similarities in early language acquisition: Polish (Rescorla et al. 2017), Italian (Rescorla et al. 2014; Bornstein et al. 2004), Danish (Bleses et al. 2008), Greek (Papaeliou & Rescorla 2011), Korean (Rescorla et al. 2013; Bornstein et al. 2004) French, Hebrew, Dutch, Spanish (Bornstein et al. 2004) and Mandarin (Tardif et al. 1999). Noun dominance was reported in all these studies except for Mandarin.

The words contained in The Combined Word List and the word structure (consonants [C] vowels [V]) are shown in Table 4.3. In total, 91.37% of the word list (The Combined Word List) is comprised of common nouns. Rescorla et al. (2017) also reported that there may be a 'universality' in what young children talk about, given that many word matches (approximately 50%) have been identified in the top 100 words used by children speaking different languages. These nouns include, 'ball, nose, car, mommy, daddy, door, shoe, baby'

(Rescorla et al. 2017) and 'juice' (Rescorla et al. 2014), all of which have been included in The Combined Word List.

In addition, consonant-vowel (CV) constructions appear to be a universal syllable shape, emerging first in many languages, followed by consonant-vowel-consonant (CVC) constructions in languages such as English, Maltese, Hebrew, and Spanish, vowel (V) in Korean and vowel-consonant (VC) in Spanish and Hebrew (McLeod & Baker 2017: 213). In total, 84.74% of The Combined Word List is comprised of words with these constructions (Table 4.3). Sell et al. (1999) also recommended that the speech sample/target word should be easily represented by an image to facilitate independent naming. This was achieved in the majority of words with the exceptions of the proper noun 'Val' and the adjective/adverb 'very' in the RWL, and the concept 'high' in The Combined Word List, in which an image was used as a representation of this concept.

Following the development of The Combined Word List, it was necessary to develop the accompanying images. The pictures were produced using the website widgitonline.com to create simple coloured line drawings. This programme is frequently used to produce symbols used in symbol communication systems and was selected as it would provide representative images which are produced in a professionally consistent fashion. Feedback on the pictures was gained from the West Midlands Patient Voices Cleft Lip and Palate Association Group, comprised of teenage/adult service users and parents. There were six images each depicting the words juice, mum, chip, shop, sea, and high which the group felt would not be clear for 3-year olds and were subsequently changed. A copy of all the images is shown in Appendix K.

The Combined Word List was arranged to assess those words in the existing RWL first (allowing analysis of this to be easily conducted separately in future cross-linguistic studies), with the additional words ordered in such a way that phonemes not assessed in the RWL appear before those sounds previously assessed but in a different word position. By ordering The Combined Word List in this way SLTs would be able to gain maximally useful information about an individual's speech even if the child was unable to complete the full speech sample.

Table 4.3. The Combined Word List

		Vowel	Word	Structure	WI consonant	WF consonant
	Warm-up		**cat**			
	Warm-up		**bike**			
Restricted Word List	1.	high	**pooh**	CV	p	
Restricted Word List	2.		**puppy**	CVCV	p	
Restricted Word List	3.	high	**pea**	CV	p	
Restricted Word List	4.	second V high	**baby**	CVCV	b	
Restricted Word List	5.	high	**bee**	CV	b	
Restricted Word List	6.		**ball**	CVC	b	l
Restricted Word List	7.		**four**	CV	f	
Restricted Word List	8.	high	**feet**	CVC	f	t
Restricted Word List	9.		**five**	CVC	f	v
Restricted Word List	10.		**Val**	CVC	v	
Restricted Word List	11.	second V high	**very**	CVCV	v	
Restricted Word List	12.		**vase**	CVC	v	z
Restricted Word List	13.		**night**	CVC	n	
Restricted Word List	14.		**nose**	CVC	n	
Restricted Word List	15.		**knee**	CV	n	
Restricted Word List	16.	high	**tea**	CV	t	
Restricted Word List	17.	high	**two**	CV	t	
Restricted Word List	18.	high	**teeth**	CVC	t	
Restricted Word List	19.		**doll**	CVC	d	l
Restricted Word List	20.		**door**	CV	d	
Restricted Word List	21.	second V high	**daddy**	CVCV	d	
Restricted Word List	22.		**house**	CVC	h	s
Restricted Word List	23.		**horse**	CVC	h	s
Restricted Word List	24.		**bus**	CVC	b	s
Restricted Word List	25.	high	**kick**	CVC	k	k
Restricted Word List	26.	high	**key**	CV	k	
Restricted Word List	27.	low	**car**	CV	k	
Restricted Word List	28.		**gate**	CVC	g	t
Restricted Word List	29.		**girl**	CVC	g	l
Restricted Word List	30.	high	**geese**	CVC	g	
Combined Word List	31.	high (accent dependent)	**mum/mom**	CVC	m	m
Combined Word List	32.	high	**shoe**	CV	ʃ	
Combined Word List	33.	high	**juice**	CVC	dʒ	

Combined Word List	34.		**chip**		tʃ	p
Combined Word List	35.	high	**moo**	CV	m	
Combined Word List	36.	high	**spoon**	CCVC	sp	
Combined Word List	37.	low	**snail**	CCVC	sn	
Combined Word List	38.	high	**sleep**	CCVC	sl	
Combined Word List	39.		**mouse**	CVC	m	
Combined Word List	40.		**web**	CVC		b
Combined Word List	41.	high	**leaf**	CVC	l	f
Combined Word List	42.	high	**sea**	CV	s	
Combined Word List	43.	low	**hat**	CVC	h	
Combined Word List	44.		**leg**	CVC	l	g
Combined Word List	45.	high	**food**	CVC	f	d
Combined Word List	46.		**zip**	CVC	z	
Combined Word List	47.	high	**zoo**	CV	z	
Combined Word List	48.	high	**sheep**	CVC	ʃ	p
Combined Word List	49.	high	**fish**	CVC		ʃ
Combined Word List	50.	high	**chair**	CVC	tʃ	
Combined Word List	51.		**shop**	CVC	ʃ	
Combined Word List	52.		**beach**	CVC		tʃ

Combined Word List	53.	high	**cheese**	CVC		tʃ
Combined Word List	54.		**page**	CVC		dʒ
Combined Word List	55.		**bridge**	CCVC		dʒ
Combined Word List	56.		**sun**			n
Combined Word List	57.		**jelly**	CVCV	dʒ	
Combined Word List	58.	high	**ring**	CVC		ŋ
Combined Word List	59.		**high**	CV	h	

4.2.2 Speech Sample B: Sentence Repetition

Sentence repetition was used as the basis of the second speech sample, described as Speech Sample B. This speech sample was selected because sentence repetition forms the basis of the speech sample used in the CAPS-A (John et al. 2006), and, as highlighted by the scoping review, cleft teams in the UK frequently use the GOS.SP.ASS. (Sell et al. 1999) in speech assessments at age-3 years. The GOS.SP.ASS was published with a set of sentences, which were refined and extended for use with the CAPS-A (John et al. 2006) in the measurement of speech outcomes at age-5 years. Whilst the GOS.SP.ASS has become an integral clinical assessment in the UK over the last two decades, this was not designed or validated specifically for use with 3-year olds.

The GOS.SP.ASS. sentence based speech sample is comprised of 22 sentences, with the number of words in the sentences ranging from 3-7 words. Both the number of items in the speech sample and the length of the sentences are likely to be difficult for some 3-year olds to recall and repeat. Based on the scoping review, sentence-based assessments were not favoured outside of the UK at age 3-years, and it may be the length and complexity of these types of speech samples for 3-year olds, that may have been a barrier in the use of this type of speech sample internationally. Indeed, consideration needed to be given that a sentence repetition speech sample may have been particularly challenging for children with language impairment, given evidence that working memory difficulties (involved in remembering the sentences in order to repeat them) strongly intersect with Developmental Language Disorder (Gray et al. 2019). Furthermore, evidence from the meta-analysis by Schwob et al. (2021) concluded that poor non-word repetition could be used as a diagnostic tool to identify Developmental Language Disorder. However, Howard (2011) argues the importance of "gathering information on sound production in larger linguistic constructions" (pp.130) when assessing speech in the CP±L population in order to assess both intelligibility, resonance and articulatory abilities in more complex contexts. It was therefore important to adapt and simplify the existing GOS.SP.ASS sentences, so favoured by UK SLTs, in order to support completion of the assessment by 3-year olds, including those with language impairments, whilst maintaining the benefits of assessment of connected speech in short sentences.

A set of shortened sentences based on the original GOS.SP.ASS sentences for use with younger children were used in an intervention study by Sweeney et al. (2020). These shortened sentences formed the basis of the sentences for Speech Sample B in this study. However, given the objectives of this research study, to compare reliability and validity across Sample A (single words and connected speech) and Sample B (short sentences) there was a need to balance the number of times each phoneme was assessed in each of the two samples. Arguably a sample which includes more phonemes than another sample may be viewed as a more valid and comprehensive assessment. Another consideration was the impact this may have on the reliability of listener judgements if one phoneme was sampled more frequently in one of the speech samples.

Another challenge related to the number of sentences that the participants were asked to repeat. In order to reduce the length of the GOS.SP.ASS sentences (to suit 3-year olds), and to match the number of times phonemes that were assessed in the single word sample it was necessary to increase the number of sentences. It was also important to consider the impact of alliteration within a sentence, given that alliterating segments feature in 'tongue twisters' (Croot et al. 2010), which could have negatively impacted upon articulation at age 3-years and not been a valid representation of the participant's speech. Finally, as with the single word naming sample, the sentences had to be easily represented by an image (Sell et al. 1999).

With such challenges in mind, there were several iterations of the sentences for Speech Sample B. The first version sampled one phoneme per phrase (as per the GOS.SP.ASS and CAPS-A) but was then 34 sentences in length. The West Midlands Patient Voices Cleft Lip and Palate Association Group advised this was potentially too long for a 3-year old to complete. The second iteration was shortened to address this. However, the number of times each phoneme was sampled no longer matched Sample A. Furthermore, different phonemes were sampled in each phrase. This was a concern as when using the GOS.SP.ASS. or CAPS-A, SLTs in the UK are trained to analyse one phoneme per sentence. Asking SLTs to change their listening habits without training could potentially have impacted upon listener reliability, and thus it may have been difficult to discern whether it was the sentence speech sample, or how the listeners carried out the analysis which impacted the reliability ratings. The third iteration, the final version of Sample B, aimed to address all these issues.

The final version is comprised of 25 sentences (Table 4.4). Each phoneme is assessed four times in total, occurring in different word positions thereby matching the single words in Sample A. The consonant clusters 'sn' 'sl' 'sp', are assessed together with two sentences loaded with low-pressure sounds, as per the CAPS-A, to facilitate resonance judgements. In each sentence, no more than four phonemes are assessed, and the average sentence length is four words (3.6 average). Although different phonemes were assessed in the same sentence (which reduced the number of sentences required to assess the phonemes required), when the Listener SLTs were given their listening procedure (Section 4.4.6), they were only asked to listen to one phoneme at a time. For example, for the sentence '**D**a**ve** dri**v**ing a **v**an' (the sounds in bold were those to be assessed) the SLTs were asked to assess /d/ the first time they listened to the sentence and when they listened again to assess the production of /v/. In this way, the Listener SLTs did not assess more than one phoneme at a time. This replicates the way they had been trained to listen on the CAPS-A (Sell et al. 2009) which may have supported ecological validity.

In a similar way to the GOS.SP.ASS. an image was created to represent each sentence. These images were taken from Shutterstock which is a subscription graphic design website with a catalogue of images. The full list of images appears in Appendix L.

Table 4.4. Sample B Sentences (with the target phoneme shown in bold).

1.	**M**ary ca**me** **h**o**me**	13.	**Z**ebra li**v**es at the **z**oo
2.	**P**u**pp**y has a **p**a**p**er	14.	**Fish** and **ch**ips
3.	**B**ob i**s** a **b**a**b**y	15.	A **sn**ail **sh**ell
4.	**Ph**o**ne** **f**ell o**ff**	16.	The **ch**ick is **h**at**ch**ing
5.	**Dave** dri**v**ing a **v**an	17.	**Sl**u**g** i**n** the **s**alad
6.	**N**eil i**s** **t**e**n**	18.	**K**aren is **m**aking a **c**a**ke**
7.	I **like** the **ball**	19.	The **ch**ildren **j**u**gg**le
8.	**Tim** has a **h**a**t**	20.	**T**i**g**er in the **j**ungle
9.	**Dad** drinking oran**ge** **j**uice	21.	Ri**ng** the bell
10.	**G**irl wa**sh**ing her **h**ands	22.	A **sp**o**tt**y dog
11.	A **sad** fa**ce**	24.	Wear your welly

			(low pressure sentence)
12.	**Sh**e is o**n** the bu**s**	25.	Wow, a yo-yo (low pressure sentence)

4.3 Development of the speech assessment tool

The CAPS-A was used as the basis for the assessment tool and was selected over existing assessment tools, and over developing a new tool for several reasons. Firstly, the CAPS-A is very familiar to SLTs working in cleft in the UK (inclusion criteria for the SLTs involved in the study was that they had previously completed the mandatory two day CAPS-A training). Other reasons for its use relate to its wide use in cleft speech outcome studies (e.g. Baillie & Sell 2020; Ahl et al. 2016; Hardwicke et al. 2016; Choa et al. 2014; and Hens et al. 2013); the growing international use of the CAPS-A (Ogata et al. 2022; Bruneel et al. 2020; Chapman et al. 2016); and because it would facilitate comparisons of speech outcomes at 3-years and 5-years in the UK, providing much needed longitudinal data regarding the persistence or resolution of speech impairments in the CP±L population.

In the CAPS-A a traffic light speech outcome system is used. A green outcome on the traffic light scale (either dark or light green) is considered a 'normal' speech outcome and in this study all dark and light green traffic light scores were collapsed into a single green outcome as per the established method for presenting outcomes using the CAPS-A (Britton et al. 2014). Amber outcomes indicate a moderate level of impairment, and red outcomes on the traffic light scale indicate a severe speech impairment. Listeners could use 'unable to score' if they did not think the quality of the speech sample was sufficient for them to make a judgement for that parameter of speech. The scalar points (including their descriptions) and the traffic light colour system were all used in the version of the CAPS-A used in this study, subsequently referred to as the Adapted CASP-A.

The original intelligibility scale from the CAPS-A (John et al. 2006) was not included in the speech assessment tool given concerns about the validity of collapsing intelligibility and acceptability into a single scale (Whitehill 2002). Subsequently both Bruneel et al. (2020) and Ogata et al. (2022) used separate measures of intelligibility, acceptability, and understandability, as judged by SLTs, in subsequent versions of the CAPS-A. Whilst the use of these separate scales was considered for inclusion in the assessment tool, these measures

are based upon clinician's judgements of intelligibility (and associated parameters) and, as the ICF-CY (WHO 2007) indicates that intelligibility is influenced by both environmental and production factors, a PROM such as the ICS (McLeod et al. 2012) was determined to be a more valid measure. As such intelligibility was not included as a parameter of assessment in subsequent listener reliability testing as part of the Adapted CAPS-A.

4.3.1 Nasality and Nasal Airflow

Based on the scoping review, an overall measure of velopharyngeal function for speech was included in the Adapted CAPS-A. This parameter was included because it would provide the opportunity to specifically investigate and report on listener reliability of this parameter, and compare reliability with other measures of nasality and nasal airflow (i.e. hypernasality and NAEs). However, an overall measure of velopharyngeal function for speech was not part of the original CAPS-A design (John et al. 2006), and consideration was given as to the type of measure which should be used. Some studies in the scoping review favoured a composite score (Swanson et al. 2017; Lohmander et al. 2006; Gunther et al. 1998) and a composite score has recently been developed by Pereira et al. (2021) to be used with the CAPS-A. A limitation of composite scores is that they can be impacted by the reliability of the parameters used to form the composite score (Lohmander et al. 2017c).

An alternative to a composite score was to use an overall judgement of velopharyngeal function. The Velopharyngeal Competence-Rate (VPC-Rate) (Lohmander et al. 2009) has been reported to be both valid and reliable (Lohmander et al. 2017c), has been adopted as a key outcome measure in the ToPS trial, including age 3-years (Shaw et al. 2019), and has been recommended for use by ICHOM at other ages (Alliori et al. 2017b). Taking into consideration that there was no evidence concerning the reliability of listener judgements using the CAPS-A at age 3-years and the reliability of listener judgements for those parameters to be used in a composite score (e.g. hypernasality and NAEs) was an unknown, the decision was made to also examine a separate measure of overall velopharyngeal function, VPC-Rate.

In contrast to Lohmander et al. (2017c), which based VPC-Rate scores on connected speech, VPC-Rate in this study was measured on either Sample A (spontaneous speech and single words) or Sample B (short sentence repetition). This takes into consideration the age of the participants, who would be expected to produce a more limited connected speech

sample than the 5-year olds in the Lohmander et al. (2017c) study, particularly influencing judgements of this parameter for Sample A. The VPC-Rate categories were colour-coded (Table 4.5) based on the traffic light system used in the CAPS-A, given that the colour coding was familiar to UK Cleft SLTs and contributed to the development of an ecologically valid assessment. In addition, should the results have indicated that this measure was reliable, the consistent use of the traffic light system at 3-years and 5-years ages would facilitate the future comparison of outcomes between these ages for UK cohorts. The original CAPS-A traffic light scales were used for hypernasality and hyponasality; however, on the basis of the scoping review, and in consideration of the work of Bayliss et al. (2011) who argued that nasal emission and nasal turbulence are an expression of the same phenomenon, nasal turbulence and nasal emission were combined into a single measure (Table 4.6. Adapted CAPS-A Resonance and NAEs).

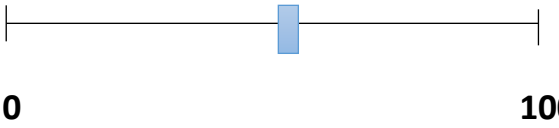
There is growing evidence to suggest that magnitude measures of speech characteristics associated with velopharyngeal function are more valid and may be more reliable (Bettens et al. 2018; Yamashita et al. 2018; Castick et al. 2017; Baylis et al. 2015) than ordinal scales. Despite this, the scoping review highlighted that categorical scales abound in assessments at age-3 years. In light of these concerns, a Visual Analogue Scale (VAS) was included in the assessment tool. Not all samples were re-evaluated using the VAS, in order to prevent listener recall, but a minimum of 20% of each speech sample was analysed again by the Listener SLTs using the VAS. The VAS was used to measure VPC-Rate and hypernasality to investigate this issue specifically in 3-year olds. This is the first time in which VPC-Rate has been measured with a VAS scale at age-3 providing a unique opportunity to compare reliability between an ordinal and VAS scale both for this measure and for hypernasality ratings. For consistency, the VAS scale (Figure 4.2, 4.3) has been developed using the same descriptors (for each end of the scale) as those used for the VPC-Rate ordinal scale and hypernasality on the Adapted CAPS-A.

Table 4.5. Adapted CAPS-A VPC- Rate (Lohmander et al. 2009)

*particularly active nasal fricatives are often misinterpreted as symptoms of velopharyngeal insufficiency (VPI) and are therefore mentioned.

Table 4.6. Adapted CAPS-A Resonance and NAE

Resonance			
Hypernasality			
Rating	Description	Unable to score	Score
0	Absent	8	
1	Borderline-minimal		
2	Mild-evident on close vowels e.g. z<u>oo</u>, thr<u>ee</u>, s<u>i</u>x / ũ, ĩ, ɪ̃/		
3	Moderate- evident on open and close vowels		
4	Severe- evident on vowels and voiced consonants		
Hyponasality			
Rating	Description	Unable to score	Score
0	Absent	8	
1	Mild- partial denasalisation of nasal consonants and adjacent vowels		
2	Marked- denasalisation of nasal consonants / m̃, ñ, ŋ̃/		
Nasal Airflow Errors (Nasal Emission & Nasal Turbulence)			
Rating	Description	Unable to score	Score
0	Absent on pressure consonants	8	
1	Occasional: pressure consonants affected <10% of the sample		
2	Frequent: pressure consonants affected >10% of the sample (judged highly pervasive or highly distinctive)		

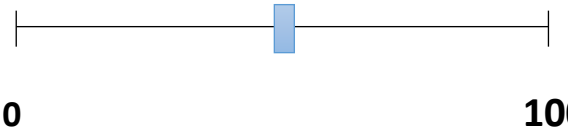


0 **Competent:** Can include active nasal fricatives*

100 **Incompetent:** Evidence of significant problems usually requiring surgical management

*particularly active nasal fricatives are often misinterpreted as symptoms of velopharyngeal insufficiency (VPI) are therefore mentioned.

Figure 4.2 VAS Score for overall judgement of speech characteristics associated with velopharyngeal function



0 **Absent:** No evidence of hypernasality

100 **Severe:** Evident on vowels and voiced consonants

Figure 4.3 VAS Score for Hypernasality

4.3.2 Articulation

When using the Adapted CAPS-A, the SLTs transcribed productions of the target sounds in the speech samples using narrow phonetic transcription. For Sample A transcription was completed on the single word naming, and for Sample B the target sounds in the sentences were transcribed. Based on their transcriptions the SLTs recorded the type and severity of Cleft Speech Characteristics (CSCs) using the Adapted CAPS-A (Table 4.7). As with measures of nasality and nasal airflow, the Adapted-CAPS-A uses a traffic light speech outcome system which considers both the type of CSC and the number of phonemes affected. A green outcome on the traffic light scale (either dark or light green) is considered a 'normal' speech outcome and all dark and light green traffic light scores were collapsed into a single green outcome (Britton et al. 2014). Amber outcomes indicate a moderate level of impairment, with red outcomes indicating a severe speech impairment.

For the purposes of this study, the label 'Anterior Oral CSCs' (used on the CAPS-A) was replaced with 'Anterior Oral Speech Characteristics', omitting the descriptor 'cleft'. This took into consideration the developing sound system of 3-year- olds as dentalisation/interdentalisation may occur developmentally at this age (Sell et al. 1999; Smit 1993) and it may be premature to categorise this as a CSC. To reflect this, the subsection dentalisation/interdentalisation was revised to dark green (indicating a normal result) even when 3 or more consonants were affected. This modification ensured that dentalisation/interdentalisation was analysed with reference to typical speech sound development whilst also allowing for comparisons at age 5-years. However, reliability results for dentalisation/interdentalisation were still presented alongside other CSCs to facilitate comparisons with other studies using the CAPS-A.

4.3.3 Phonology

A further adaptation made to the CAPS-A for this study, was an additional measure of phonological processes. Phonological processes were recorded in four categories (phonological processes present, age-appropriate phonological processes, delayed phonology, disordered phonology) which extended the presence/absence rating of non-cleft speech immaturities/errors in the original CAPS-A (John et al. 2006). This takes into consideration findings from the scoping review which indicated that not only articulation, but also phonology, are frequently included in the analysis of cleft speech at age 3-years, both in

the UK and internationally (Hutters et al. 2001, Chacon et al. 2017, Konst et al. 2003, Wren 2013, Klintö et al. 2016).

The inclusion of a measure of phonology at age-3 reflects that at this age, an individual's sound system is in a period of significant development (Dodd et al. 2003). The categories of age-appropriate phonological processes, delayed phonology, disordered phonology were not mutually exclusive as an individual could present with age-appropriate phonological processes, as well as delayed and disordered phonology (Bates & Titterington 2021). This information provides an important summary of an individual's phonology highlighting the complexity of their speech sound system.

Table 4.7. Adapted CAPS-A CSCs and Phonology

Cleft Speech Characteristics Summary				
Speech Characteristics		**Absent 0**	**1 or 2 consonants affected**	**3 or more consonants affected**
	Anterior Oral Speech Characteristics			
1	Dentalisation/inter-dentalisation			
2	Lateralisation/ lateral			
3	Palatalisation/ palatal			
	Posterior Oral CSCs			
4	Double articulation			
5	Backed to velar/uvular			
	Non-oral CSCs			
6	Pharyngeal articulation			
7	Glottal articulation			
8	Active nasal fricatives			
9	Double articulation			
	Passive CSCs			
10	Weak and/or nasalised consonants			
11	Nasal realisation of plosives &/or suspected passive nasal fricative			
12	Gliding of fricatives/affricates			
	Phonology			
Phonological processes present		**Age appropriate phonological processes**	**Delayed phonology**	**Disordered phonology**

4.4 Participant recruitment and participation

4.4.1 Settings

The main site for the study was The West Midlands Cleft Lip and Palate Service (WMCLPS), based at Birmingham Children's Hospital, Birmingham Women's and Children's NHS Foundation Trust. The WMCLPS is one of the largest single-unit cleft centres in the UK (CRANE 2020) and provides cleft care for patients and families across the West Midlands including Birmingham, Solihull and the Black Country, Herefordshire, Shropshire, Worcestershire, Coventry and Warwickshire and Staffordshire. The study was coordinated from the WMCLPS, and all 3-year old participants were recruited from this site and undertook the speech recordings at the site.

4.4.2 Participant Group: 3-year olds with CP±L

4.4.2.i Sampling

Three-year-olds are routinely seen for speech assessment at the WMCLPS as part of the standard care pathway. Convenience sampling was used in the context of the ongoing speech assessments taking place at this timepoint at the WMCLPS. This method of sampling was selected as the most feasible through which to recruit the participants with the least disruption to the protocol delivery of care at the WMCLPS. Two approaches were taken with regards to the sample size, the first was to consider the number of 3-year old participants required to answer the research objectives, and the second was to consider the number of listening incidents for each parameter of speech, which was particularly important in terms of reliability calculations. A sample size of twenty 3-year olds with CP±L was determined, taking into consideration existing literature, statistical analysis, feasibility, and the number of listening incidents as described below.

Firstly, the studies of John et al. (2006), Sell et al. (2009) and Chapman et al. (2016) who also aimed to develop standardised assessments and audit tools for the CP±L population were examined. In these studies listener reliability was calculated based on speech samples from 10 individuals. In their versions of the CAPS-A, Ogata et al. (2022) and Bruneel et al. (2020) calculated inter-rater reliability based on 20 speech samples. In all of these studies, only a single type of speech sample was used. In comparison, Klintö et al. (2011) recruited 20 participants with CP±L to compare the impact of speech samples on listener judgments at age-5 years. However, the number of participants who completed the different speech

samples ranged from 11-20. Taking into consideration that not all the 3-year olds participating in the study may have been able to complete the speech samples, a target sample size of 20 was used.

Calculations of inter and intra-rater reliability were analysed using weighted Cohen's kappa as described by Mandrekar (2011). Sample size calculations were calculated using R software (http://www.r-project.org/) and the 'kappaSize' package was used to guide the sample size. Such sample size calculations could only be used as a guide, given that at this point in the research process the number of categories used to assess each parameter of speech was unknown, and the spread of speech outcomes at age-3 years could not be accurately predicted. The CAPS-A tool (John et al. 2006) was used as a basis to determine the number of categories within each speech parameter. In the CAPS-A, the number of scalar points used to measure each parameter of speech varies between three and five, with three being used most frequently.

For three scalar points, assuming a percentage split across the scale points of 50:25:25, and using alpha =0.05, power =0.80, and an alternative hypothesis of kappa = 0.61 (bottom end of "substantial") and the null hypothesis of kappa =0.3 (midpoint of fair agreement as described by Landis and Koch [1977]) a sample size of 11 was suggested with a warning of small cell size. For five scalar points, assuming a percentage split across the categories of 25:25:17:17:16, and using alpha =0.05, power =0.80, and an alternative hypothesis of kappa = 0.61 (bottom end of "substantial") and the null hypothesis of kappa =0.3 (midpoint of fair agreement as described by Landis and Koch [1977]) a sample size of 7 was suggested with a warning of small cell size.

The sample size of participants with CP±L also had to be feasible, determined by the number of patients treated by the WMCLPS. The study aimed to open for recruitment in 2019, and therefore patients born in 2015 and 2016 (who would be turning 3-years in 2019) were eligible for inclusion. In 2015, 88 babies were born with CP±L and treated by the WMCLPS, and in 2016 this number was 83. A target sample of 20 participants with CP±L was therefore achievable considering that this was just over 20% of the 3-year olds with CP±L cared for at WMCLPS each year. The sample size also took into consideration that there were designated clinics for speech assessments at age-3 years at the WMCLPS and there was no backlog in waiting list for these appointments at the time of the study. As such a proportion of 3-year

olds would have already completed their speech assessment and thus would not be eligible for recruitment.

In addition, the sample size also had to be feasible in terms of the demands placed on the Listener SLTs taking part in the study. It was important to consider that if all the participants completed both speech samples, this would amount to 40 recordings for listening and analysis, in addition to potentially 10 recordings of the Control Group. In order to calculate intra-rater reliability, the Listener SLTs also repeated the listening and analysis task on a minimum of 20% of the speech recordings for each speech sample. Furthermore, the Listener SLTs were required to listen and analyse both hypernasality and VPC-Rate using the VAS for a minimum of 20% of the speech recordings. In total, this resulted in each listener analysing 81 speech samples, with a total of 648 speech samples being analysed in the study. Given that the Listener SLTs were completing the listening and analysis within their working hours and/or in their own time, with no financial remuneration, sample size decisions needed to be pragmatic considering the amount of time the listening/analysis would take (which using the CAPS-A at age 5-years takes approximately 15 minutes per speech sample [Ahl & Harding-Bell 2017]).

Calculations of listening instances also took into consideration the number of SLTs analysing the speech samples (see section 4.4.4.i). If the sample size of twenty 3-year olds with CP±L was achieved, with all participants completing the speech sample, and with 11 SLTs (maximum sample size of SLTs) completing the analysis, this would have resulted in a total of 220 listener judgements for every parameter of speech assessed for each speech sample. Should only 10 (half) of the 3-year olds have completed the speech sample, and six (the minimum number) of SLTs been recruited this would have resulted in 60 listener judgements for every parameter of speech for each speech sample. It was therefore important to consider the number of listening instances, and not just the number of 3-year old participants.

A sample size of twenty 3-year olds with CP±L was double that of the samples used in the studies by John et al. (2006), Sell et al. (2009), and Chapman et al. (2016) and was the same as that used by Klintö et al. (2015a), Bruneel et al. (2020) and Ogata et al. (2022). It also fulfilled the cautionary sample size calculations. Pragmatically, this number was also achievable in terms of numbers of children age-3 years at the WMCLPS, and the time it would take the Listener SLTs to complete the analysis. This sample size would have also resulted in

a maximum of 220, or a conservative estimate of 60, listener judgements for every parameter of speech assessed.

4.4.2.ii Recruitment

As part of the standard care pathway, all children with CP±L treated at WMCLPS are routinely offered a 3-year assessment appointment in the Speech and Language Therapy department. Children aged 36-47 months with CP±L who met the inclusion and exclusion criteria (Table 4.8) were recruited to the study and seen for a study assessment during a 10-week recruitment window. The parents/guardians of eligible 3-year old children with CP±L were sent the Participation Information Leaflet (PIL) (Appendix M) and letter to parents/guardians inviting their child to participate in the study (Appendix N) in the post by the cleft administration team, together with their appointment letter. One week after the appointment letter had been sent, using a prescribed telephone script (Appendix O) the Chief Investigator (CI) (a member of the direct care team) telephoned the parent(s)/guardian(s) to ascertain interest in their child's participation in the study. Parent(s)/guardian(s) who did not wish for their child to participate in the study were seen for the standard 3-year appointment as per the appointment letter which was completed by one of the Cleft SLTs working at the WMCLPS (not the CI).

Those parent(s)/guardian(s) who agreed for their child to participate, were consented to the study by the CI at the time of the study assessment using the CP±L Consent Form (Appendix P). A copy of the completed consent form and another copy of the PIL were given to the parent(s)/guardian(s). A screening log, consent log, and participant identification log were completed by the CI as per Good Clinical Practice guidelines (NIHR CRN 2013). These were the only forms that contained participant identifiable information. The participant's cleft consultant was informed by letter that their patient had consented to participate in the research study as per guidance from the NHS Trust. Participant demographic information i.e. date of birth, gender, and cleft type (required to report the demographics of the CP±L Group participants) were recorded. Figure 4.4 shows the flowchart of the recruitment and assessment process for 3-year old children with CP±L. In total 34 children were screened against the inclusion and exclusion criteria (8 not eligible; 4 declined; 2 did not attend) and 20 were recruited to the study.

Following recruitment, the 3-year olds were randomised to the order in which they completed the speech samples (further detailed in Section 4.4.5.i). The use of randomisation considered the participant's age and the potential for fatigue i.e. if the 3-year olds could only maintain their focus and attention for long enough to complete one speech sample. Therefore, randomisation to the order the participants completed the speech samples ensured that there were examples of both speech samples which could be subsequently analysed by the Listener SLTs.

Table 4.8. CP±L Participant Inclusion/Exclusion Criteria

Inclusion Criteria	**Rationale**
Children diagnosed with CP±L.	The study examined the impact of different speech samples and rating methods, and the validity and reliability of listener judgements in patients with CP±L.
Children treated at the WMCLPS.	Children were recruited from the main study site.
Children aged between 36-47 months during the assessment period.	The study specifically aimed to examine the impact of different speech materials and rating methods on participant completion of the speech sample, and the validity and reliability of listener judgements in children with CP±L at age-3 years. Typically, children are seen for a 3-year assessment at WMCLPS at any age between 36-47 months and thus the age range in the study reflected this.
Patients who were eligible for a standard three-year assessment.	As part of standard care at the WMCLPS, patients with complex medical needs and/or significant developmental delay were offered a cleft consultant appointment, rather than an SLT appointment, as this met their needs more appropriately. This ensured that it was appropriate for the patient to be offered a 3-year SLT assessment.
Exclusion Criteria	
Patients with Submucous Cleft Palate (SMCP).	Patients with SMCP are excluded from speech audit results at age-5 years (CRANE 2016). SMCP also differs anatomically from other forms of palatal cleft (Sommerlad et al. 2004), with not all

	patients with SMCP requiring surgical management.
Patients with an identified genetic syndrome.	Specific speech patterns associated with syndromic clefts can differ from the wider cleft population (D'Antonio et al. 2001) and may have impacted the reliability of listener judgements. Children with genetic syndromes are also at greater risk of language delay and may not have been able to complete the speech samples or engage with the assessment materials in the same way (Kilcoyne et al. 2021; Scherer et al. 1999).
Children from a non-English speaking family.	The speech assessment was administered in English. For children who had not been exposed to English, it was more appropriate to assess their first language with the support of an interpreter as per the established pathway at the WMCLPS.
Failure/technical problem with the recording of the speech assessment.	Recording quality may have impacted reliability scores, and high-quality recordings are reported to enhance listener ratings (Sell et al. 2009).

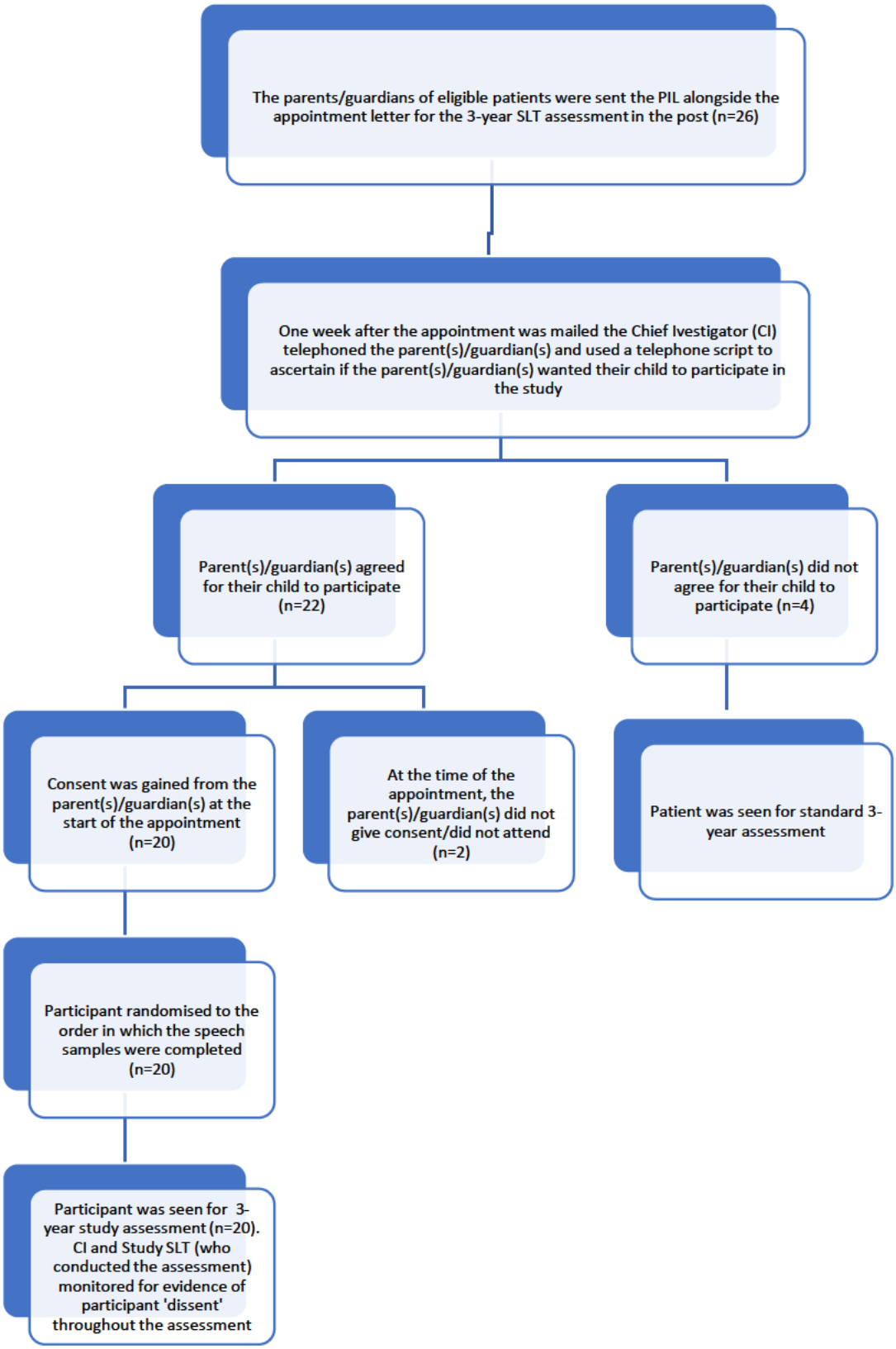


Figure 4.4 Flowchart of the recruitment and assessment process for 3-year old children with CP±L

4.4.3 Participant Group: Control Group of 3-year olds

4.4.3.i Sampling

A convenience sample was used to recruit participants to the Control Group. Given that the motivation for recruiting a Control Group was to provide information about the specificity of the assessment, convenience sampling was both time and resource-efficient (Jager et al. 2017), particularly because there was no requirement for random or stratified sampling. A sample size of a minimum of two and a maximum of five Control Group participants was used. This sample size was based on existing literature.

In the CAPS-A study (John et al. 2006) only one control speech sample was used to contribute to the specificity of the assessment and the Americleft study did not use any samples from a control group (Chapman et al. 2016). However, in the development of the Belgian Dutch outcome tool, Bruneel et al. (2020) used 10 speech samples from a control group in phase one of the study, and one speech sample from a control group in phase two. The SVANTE (Lohmander et al. 2017b; Lohmander et al. 2005) was developed with a different design strategy, to develop normative scores based upon the non-cleft population, and thus 102 non-cleft 3-year olds were recruited. In the current study, a minimum of two participants was selected to ensure that an example of speech Sample A and B was completed to allow specificity to be compared across the two speech samples.

In this study, the Control Group was specifically recruited to investigate the specificity of the assessment framework; that Listener SLTs did not identify children without CP±L as having those speech characteristics associated with CP±L. Specific consideration was given to the number of listening instances required to identify an issue with the specificity of the assessment framework. Had the maximum number of children in the Control Group been recruited (five) and the maximum number of SLTs (11), this would potentially have resulted in a maximum of 55 listening instances for each parameter of speech i.e. hypernasality would have been assessed 55 times, and in total there would have been a maximum of 880 listening instances for cleft specific speech parameters for each speech sample. This number of listening instances was judged to be adequate to determine if there was a specificity issue in relation to each cleft parameter of speech.

4.4.3.ii Recruitment

Control Group participants were recruited within a 10-week recruitment window. Members of the Therapies Department (Physiotherapy, Occupational Therapy, Speech and Language Therapy)/WMCLPS acted as gatekeepers, through which 3-year old children (without a cleft palate) of the gatekeeper's friend or family members were informed about the study and invited to contact the CI if they were interested in participating. Members of the Therapies Department/WMCLPS were sent an email (Appendix Q) outlining the study with the Control Group PIL attached (Appendix R). They were asked to forward the email and attached PIL to any of their family or friends who may be interested in participating.

Parent(s)/guardian(s) contacted the CI if they wanted their child to participate. The CI discussed with the parent(s)/guardian(s) whether their child met the inclusion criteria. The CI then invited individuals who had read the PIL, met the inclusion criteria, and agreed for their child to participate in the study to attend a study assessment at the WMCLPS. A total of six families contacted the CI, all met the inclusion/exclusion criteria, but one could not attend an appointment during the study window, and thus five children were consented and recruited to the study.

At the study assessment parent(s)/guardian(s) were asked to consent to their child's participation in the study by the CI (Appendix S). A copy of the completed consent form and another copy of the PIL were given to the parent(s)/guardian(s). A screening log, consent log, and participant identification log were completed as per Good Clinical Practice guidelines (NIHR CRN 2013). These were the only forms that contained participant identifiable information. Participant demographic information i.e. date of birth and gender (required to report the demographics of the Control Group participants) was recorded as outlined in Table 4.9. Following consent, the Control Group participants completed the speech samples. A short report based on the assessment was sent to the parent(s)/guardian(s) following the assessment. Figure 4.5. shows a flowchart of the recruitment and assessment process for 3-year old children in the Control Group. Following recruitment, the Control Group participants were randomised to the order in which they completed the speech samples (further detailed in Section 4.4.5.i).

Table 4.9. Control Group Participant Inclusion/Exclusion Criteria

Inclusion Criteria	**Rationale**
Control Group participants were aged between 36-47 months at the time of assessment.	The study investigated speech assessments in 3-year-olds.
Control Group participants had no current or previous involvement with Speech and Language Therapy, nor were waiting for a Speech and Language Therapy initial assessment.	To minimise the possibility of a Control Group participant presenting with a communication difficult, thus ensuring that the Control Group participants were suitable as a control.
Control Group participants had no medical condition(s) associated with communication impairments e.g. diagnosed developmental delay, hearing impairment, syndromes associated with speech and language delay e.g. Downs Syndrome, 22Q11 Deletion Syndrome.	
The parent(s)/guardian(s) did not have any concerns about their child's communication.	
The CI did not have any concerns about the child's communication before the assessment.	
Exclusion Criteria	
Failure/technical problem with the recording of the speech assessment.	Recording quality may impact reliability scores, and high-quality recordings were required to enhance listener ratings (Sell et al. 2009).
Children from a non-English speaking family.	The speech assessment was administered in English. For children who had not been exposed to English, it would not have been appropriate to assess their speech in English.

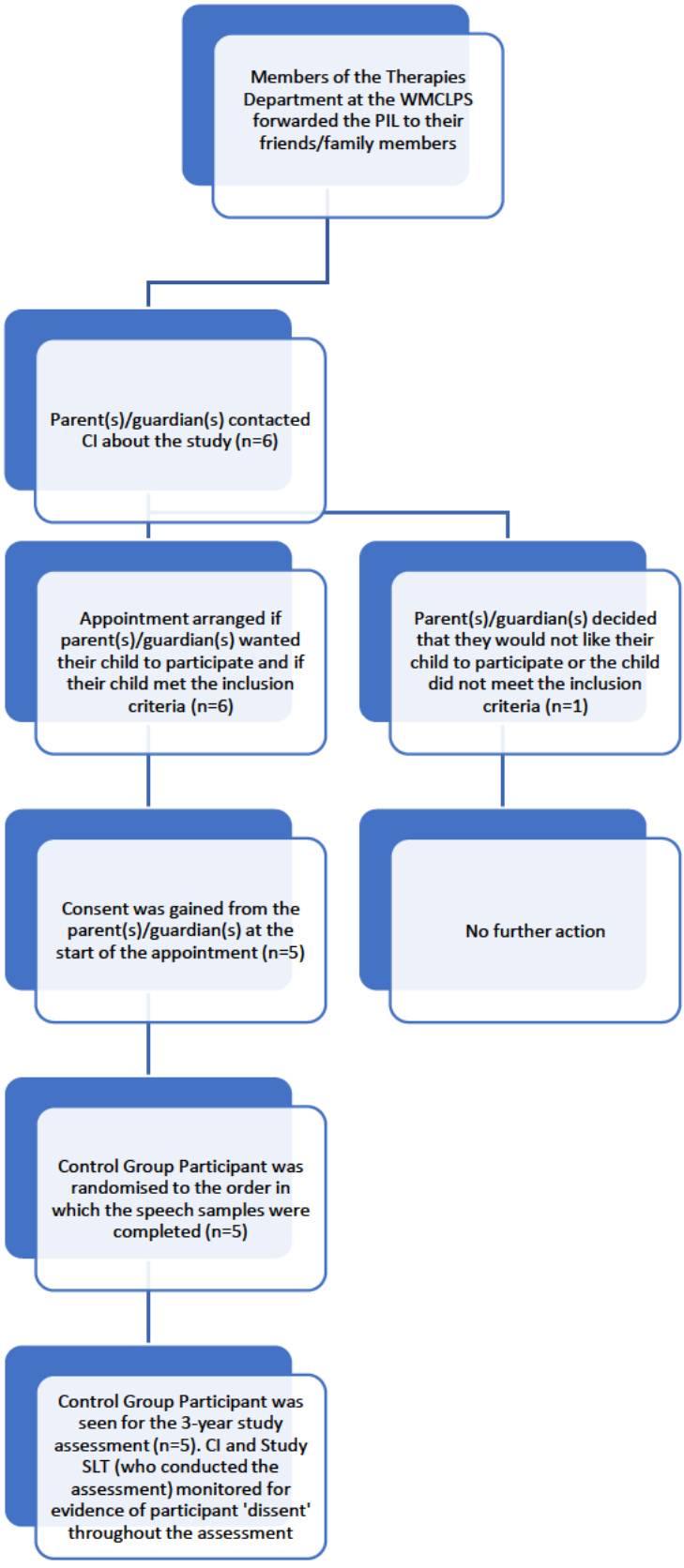


Figure 4.5. Flowchart of the recruitment and assessment process for 3-year old children in the Control Group.

4.4.4 Listener Group: Specialist Cleft SLTs

4.4.4.i Sampling

A sample size of a minimum of six Listener SLT participants was utilised based on existing literature and the feasibility of recruitment. The target sample took into consideration the number of listeners previously used to determine the reliability and validity of listener judgments in the development of speech assessment tools. In the CAPS-A study, 10 listeners were used (John et al. 2006), in the Americleft study, reliability ratings were based on nine listeners in the first part of the study and six listeners in the second part (Chapman et al. 2016), in the development of the Belgian Dutch outcome tool, two listeners were used in phase one of the study and four in phase two (Bruneel et al. 2020), and the CAPSA-JP used six listeners (Ogata et al. 2022).

There are 11 cleft services in the UK. As per assessment practices at age 5-years (which requires Specialist Cleft Team SLTs to listen to and analyse a minimum of 10% of the speech samples of another cleft centre [Britton et al. 2014]), it was intended to recruit SLTs from other NHS cleft teams as well as the WMCLPS. The inclusion of SLTs from other cleft teams protects against 'listener drift;' SLTs who regularly work together in a cleft centre develop listening characteristics that differ from other cleft centres (Kent et al. 1999). Widening the recruitment procedure to include SLTs from other cleft teams controlled for potential differences in 'listening' across cleft teams and enhanced the validity and reliability of the listener judgements. A convenience sample was utilised over a stratified sample, given that the demands on cleft teams meant it was likely to be difficult to recruit an SLT from each NHS cleft team, particularly due to the time commitment involved in completing the listening (either personal time or work time agreed with their manager). Whilst a target sample size of 6-11 Listener SLTs compared well to other CAPS-A studies, this number was higher than that used in the reporting of reliability in studies specifically of 3-year olds i.e. two listeners were used in the studies by Chacon et al. (2017), Raud Westberg et al. (2017), Klinto et al. (2016), and four in the Willadsen et al. (2018) study. In other outcome studies using the CAPS-A e.g. Baillie & Sell (2020) and Sell et al. (2015) only two Listener SLTs were used. The higher number of Listener SLTs in this study could potentially have led to increased inter-rater variation, subsequently impacting reliability ratings compared to studies utilising fewer listeners. However, because the aim of this study was specifically to test for Listener SLT reliability at age-3 years and make comparisons with other studies validating the CAPS-A with older age

groups it was judged as essential to include a similar number of listeners to those studies and to include listeners from a variety of UK Cleft Centres.

As discussed in Section 4.4.2.i, having a minimum sample size of six listeners ensured that if all 3-year olds with CP±L completed the speech sample there would be a minimum of 120 listener judgements for every parameter of speech assessed, for each speech sample. If only half of the 3-year olds completed the speech sample, there would still be a minimum of 60 listener judgements for each parameter of speech, supporting the validity of any conclusions relating to listener reliability.

4.4.4.ii Recruitment

The Cleft Clinical Excellence Network (CEN) acted as a gatekeeper for this component of the study. An email was sent to all members of the Cleft CEN (Appendix T), along with the PIL (Appendix U). SLTs wishing to participate were asked to contact the CI. The CI spoke to SLTs on the phone confirming they had read the PIL, answered any questions that they had, and checked that they met the inclusion criteria as described in Table 4.10. If the inclusion criteria were met, the CI contacted their local Research and Development (R&D) office and sought approval to recruit a participant from the NHS Trust. Once approval was gained from the local R&D office the SLT was emailed a consent form (Appendix V) (and another copy of the PIL) which they were asked to complete and return by email to the CI.

A total of 8 SLTs were screened for inclusion in the study and all met the inclusion criteria set (Table 4.10). All were invited to take part and consented to participate in the study. Seven SLTs completed the listening and analysis. The seven Listener SLTs were recruited from five different NHS cleft teams, including the following geographic areas: the North-East of England, Scotland, the West Midlands, Wiltshire and Cambridgeshire. Three SLTs were recruited from the WMCLPS. Figure 4.6 shows the Recruitment Procedure for Listener SLTs.

Table 4.10. Listener SLT Inclusion/Exclusion Criteria

Inclusion Criteria	**Rationale**
The individual was currently working as an SLT.	The research investigated listener judgements made by Cleft SLTs and required phonetic transcription. Lewis et al. (2003) reported that SLTs are more reliable judges of nasality than other professional groups or students.
The SLT had designated sessions working with paediatric Cleft Patients in an NHS Cleft Centre.	The Listener SLT was working as a specialist in cleft, was in current NHS employment and

	would, therefore, follow NHS guidelines regarding confidentiality.
The SLT had completed mandatory two day CAPS-A training and had participated in consensus listening in the audit assessment of children at five-years of age.	Experienced, trained listeners are more reliable than inexperienced listeners (Gooch et al. 2001). The CAPS-A training has shown consistent SLT judgments and good levels of agreement between SLTs on many aspects of the tool (Sell et al. 2009). Ensuring that the SLTs had completed CAPS-A training and participated in the audit of speech outcomes speech at age 5-years was used as an indicator of both a trained and experienced listener.
Exclusion Criteria	
The SLT carried out a speech assessment with the CP±L Participant in the last year.	Specifically relating to Listener SLTs recruited from the WMCLPS, this minimised a potential source of bias relating to prior knowledge of the participants' speech, which could have influenced listener judgements (Day & Altman 2000).

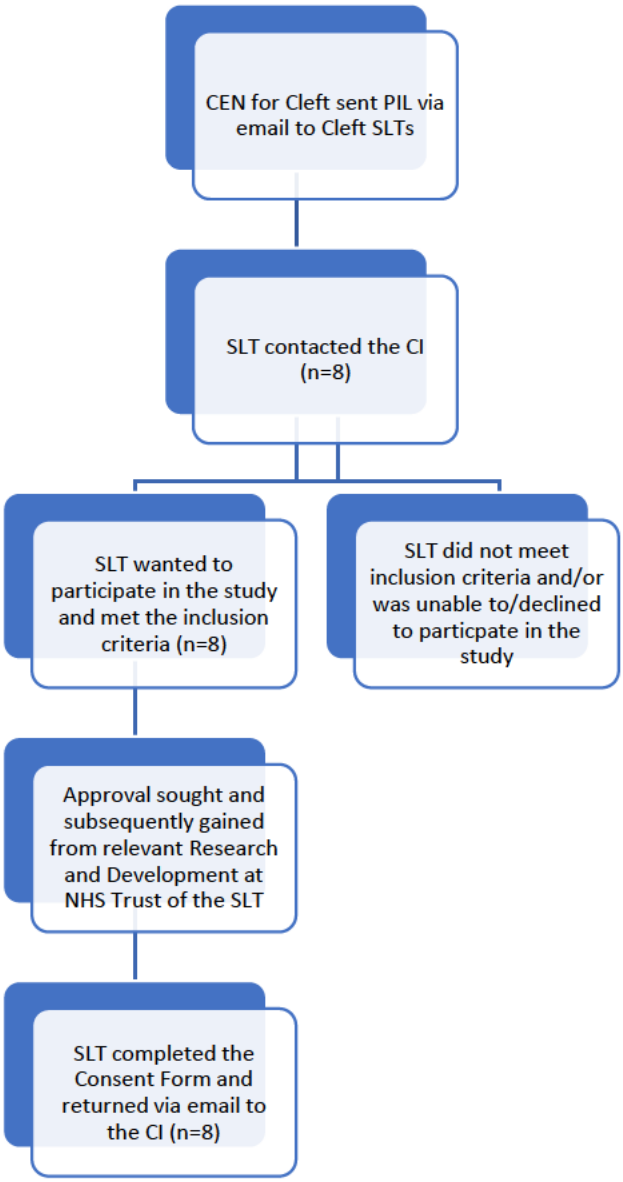


Figure 4.6 Flowchart of the recruitment process for Listener SLTs

4.4.5 Procedure: Completion of the Speech Samples

4.4.5.i Randomisation

The 3-year old participants were randomised as to the order of completion of the speech samples using a process of constrained randomisation (Li et al. 2016). Those randomised to Group 1 completed Speech Sample A first (spontaneous speech sample and single word naming), followed by Speech Sample B (short sentence repetition), and those randomised to Group 2 completed Speech Sample B first followed by Sample A. The Control Group participants were randomised separately to the group with CP±L. The process of constrained randomisation was selected to avoid (as far as possible) imbalance between the groups given the sample size. Each week of the 10-week recruitment window formed a 'block' containing a specified number of randomly generated group assignments based on the number of appointments (potential participants) for that week (see Appendix W for an example of the block randomisation for Week 1). Randomisation was completed using the online programme 'sealedenvelope.com'. In the CP±L group, 60% were randomised to complete Sample A first, and 40% to complete Sample B first. One participant who was randomised to Speech Sample B was switched to Speech Sample A based on the clinical judgement of the SLT who completed the assessment (Study SLT) as being more appropriate for this individual.

4.4.5.ii Consent

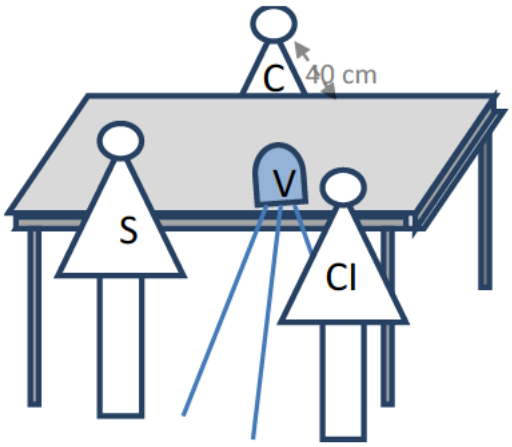
Specific itemised consent for video recordings of the assessment session was recorded on the consent forms. Additionally, for the CP±L group, parent(s)/guardian(s) were asked to sign a Speech and Language Therapy Video/Audio Request Form in line with departmental policy. Video recordings were selected over audio-only based upon evidence from Klintö & Lohmander (2017) that both lingual-labial, interdental articulation, and audible nasal airflow are more critically analysed using video recordings rather than from audio alone. This is supported by evidence from Pereira (2012) that dentalisation can be missed when only audio recordings are used.

4.4.5.iii Completion of the speech samples

The 3-year old participants attended their speech assessments at the WMCLPS. The same SLT, referred to as the Study SLT (who did not participate in the study as a listener and was not the CI) carried out all the assessments. All the speech assessments were video recorded by the CI. This allowed the Study SLT to focus on the participant and administration

of the speech assessment. This was important as for the CP±L group, the study visits were used also as their standard care pathway speech assessment at age 3-years. The CI was able to focus on operating the recording equipment (framing of the picture, live monitoring of audio quality, etc.) This aimed to minimise the number of recordings excluded from the study due to technical problems with the recording and ensure that high quality recordings were made which were a valid representation of the participant's speech. This arrangement also mirrored that used in the international multi-centre ToPS Trial to achieve high-quality speech recordings for analysis (Shaw et al. 2019).

As per Sell (2005) and the recommendations made by CLISPI (2017), recordings were made in a standardised way. Speech assessments were video recorded using a Panasonic W850 high-definition video camera with an internal microphone. The sound quality was monitored live using Sennheiser HD203 headphones. Video recordings were saved onto a dedicated memory card for this study. Figure 4.7 shows the set-up of the recording equipment. The camera was set up pointing directly at the child's head, and the child was recorded against a neutral background to minimise any visual distractions during the subsequent listening task/ analysis. The child's face and upper neck were framed in the picture to facilitate clear observations of the child's articulators as per Sell et al. (2009). When possible, the speech materials were presented in a way that supported a neutral head position of the child, encouraging them to face towards the camera. Following the assessment session, the videos were uploaded by the CI onto the secure drive at the NHS Trust designated for video recordings and were deleted from the memory card. All videos were reviewed by the CI to check audio and video quality. Videos were given a unique identification number. Only the CI knew which video corresponded to which participant (as recorded in the Site File).



C = Child participant
S = Study SLT speech therapist
CI = CI- camera operator
V = video recorder on tripod at child's level

Figure 4.7 Set up for the recording equipment

Study assessments lasted for approximately 1 hour which is in keeping with the standard appointment time for speech assessments at age 3-years at the WMCLPS. This included a case history (not part of the study), completion of the speech sample(s) (part of the study), additional assessments as indicated by the Study SLT (not part of the study), feedback to parents/guardians and future planning (not part of the study). The toys and equipment used to elicit the speech samples in the recordings were standardised for the study and are detailed below. All toys and equipment utilised in the study met the NHS Trust at WMCLPS' requirements for Health and Safety and Infection Control.

Spontaneous Speech Sample (Sample A): The spontaneous speech sample was elicited using a variety of different toys which the participants could choose. This included the Usborne First 100 words book, magnetic scenes featuring Paw Patrol characters, Peppa Pig, farm animals, transport, Mr Potato Head, and safari 'Happy Land' toys. The Study SLT asked the children to talk about what they could see in the book or comment on what they were playing with to elicit the speech sample and the CI timed the accumulative length of the spontaneous speech sample to ensure that the two-minute target had been achieved (if possible).

Single Word Naming (Sample A): In the single word naming task, each word was presented on a separate card depicted by a picture. To familiarise the participants with the task two 'warm-up' pictures were presented before the test items. Activities to encourage participation in the activity were used e.g. posting the picture in a box after naming, laying out the pictures upside down and having the participant turn them over and name them. The Study SLT used their clinical judgement to determine if such activities to encourage participation were required. If

the participant was unable to name the picture spontaneously an elicitation hierarchy (adapted from Lohmander et al. 2009) was used to elicit the target word as detailed below.

i- Semantic prompting

ii- Forced alternative (when the target is the first word)

iii- Repetition of SLT

Short Sentence Repetition (Sample B): For sentence repetition, each sentence was depicted by a picture stimulus. The Study SLT showed the participant the picture stimulus and introduced each sentence with, 'This is Bob…' 'This is Neil', etc. The SLT then said the target sentence and asked the participant to repeat after them. The Study SLT was able to repeat the sentence multiple times to support the 3-year old participants if necessary.

Discontinuation and Participant Dissent: During the assessment sessions the Study SLT and the CI monitored the participants for signs of any difficulty completing the assessment, distress during the session, or signs of dissent (e.g. the participant refused to complete the assessment; the participant stopped speaking; the participant became upset/started to cry). The Study SLT or the CI used their clinical judgement and discontinued the assessment session as necessary.

The CI recorded the extent to which the participant completed the speech sample on the Assessment Recording Form (Appendix X). For the spontaneous speech sample this was based on the length of the speech sample, for single word naming and sentence repetition this was based on the number of words or sentences completed. Full completion was defined as completing > 90% of the speech sample, partial completion as completing >10-90% of the speech sample, and not completing the speech sample as <10% of the speech sample. The CI also recorded the parameters of speech that the Study SLT had reported on in the medical record for participants in the CP±L group using the Speech Parameters Recording Form (Appendix Y). Following the study appointments for the CP±L group, the Study SLT liaised with other relevant professionals and wrote an assessment report to parents and professionals as per established practice at the WMCLPS. For participants in the Control Group the CI wrote a short report which was sent to the participant's parent(s)/guardian(s) about their child's

performance. None of the participants in the Control Group required an onward referral to speech and language therapy.

4.4.6 Procedure: Analysis of the Speech Recordings

3-year old participant speech recordings were collated and transferred to Listener SLTs using NHS encrypted and password-protected USBs. Also included on the USBs were the listening instructions, the speech sample forms for phonetic transcription, and spreadsheets for the Listener SLTs to record their analyses and judgements.

4.4.6.i Practice Listening Task

A practice listening of two speech recordings, an example of Speech Sample A and one of Speech Sample B, took place before the main listening task. The practice listening results are not included in the study results. The practice listening session aimed to support listener calibration and familiarise the Listener SLTs with both the speech samples, the listening procedure, the assessment forms and methods and informed the details of the final guidance provided to the Listener SLTs.

The practice listening highlighted some important findings before the main listening task could start. Group feedback was given to all the Listener SLTs, to clarify the scoring of VPC-Rate, calibration for hypernasality, NAEs, and passive CSCs, and to provide clarification and guidance on the rating of phonological processes. In addition, each Listener SLT was given individual feedback, highlighting areas in which their judgement was not in agreement with the majority. This individual feedback varied across the listeners, with NAE ratings differing from the majority as the most frequent discrepancy for both speech samples. The Listener SLTs were then asked to listen back to their specific examples comparing and calibrating their responses with that of the other listeners. The Listener SLTs were also sent an excel spreadsheet allowing them to see their responses in the context of the other Listener SLTs (anonymised). For ordinal ratings, the spreadsheet was colour coded (using the traffic light system shown in Tables 4.5, 4.6, 4.7) to support comparisons. For VAS ratings the SLTs could compare numerical (0-100) scores. Examples of the feedback on the practice listening is provided in Appendix Z.

4.4.6.ii Main Listening Task

The main listening was divided into three listening sessions as outlined in Figure 4.8. Every Listener SLT analysed the same recordings in the same session. However, to reduce the

potential impact of fatigue on the analysis, each Listener SLT listened to and judged the recordings in a unique order. Only one speech sample from each participant was analysed within the same listening session to prevent listener recall. For each session, every Listener SLT had an excel spreadsheet in which to input their ordinal judgements. To reduce the potential for error the excel spreadsheets were designed with limited options (in a drop-down menu) which the SLT would select for each judgement. In addition, each judgement was colour coded (using the traffic light system) as a visual prompt to reduce inputting errors. The third listening session included VAS ratings and this was completed using Qualtrics, an online survey platform that is approved by Coventry University for data collection, because of the high-security measures offered by the platform.

All the Listener SLTs were given specific written instructions and guidance for each listening session to aid their analysis. Listeners were able to structure their listening sessions as convenient for them but were advised not to listen to a recording more than 3 times, following the advice set out by Shriberg et al. (2005). To calculate intra-rater reliability a proportion of the speech samples were analysed by the Listener SLT for a second time with a minimum of 4 weeks between each listening session to prevent recall.

Listener SLTs were provided with information regarding the age range of the participant i.e. whether they were aged 3:0-3:05 months, or 3:6-3:11 months. This supported the Listener SLTs in making age-related judgements about phonology. To ensure that the Listener SLTs were all using the same speech acquisition norms, they were all provided with the set of norms by Dodd et al. (2003). These norms were developed based on English children. An outline of each listening session is provided below and a flowchart of the listening process is shown in Figure 4.8.

Listening Session 1: Ordinal Rating Scale

- Total: 24 video recordings
- Sample A: 17 video recordings
- Sample B: 7 video recordings

Listening Session 2: Ordinal Rating Scale

- Total: 24 video recordings

- Sample A: 12 video recordings (including x5 previously analysed in Listening Session 1 to calculate intra-rater reliability)
- Sample B: 12 video recordings

Listening Session 3: Ordinal and VAS Scale

- Ordinal: Sample B: 5 video recordings (including x5 previously analysed in Listening Session 1 or 2 to calculate intra-rater reliability)
- VAS: 11 videos recordings (Sample A: 6 video recordings, Sample B: 3 video recordings)

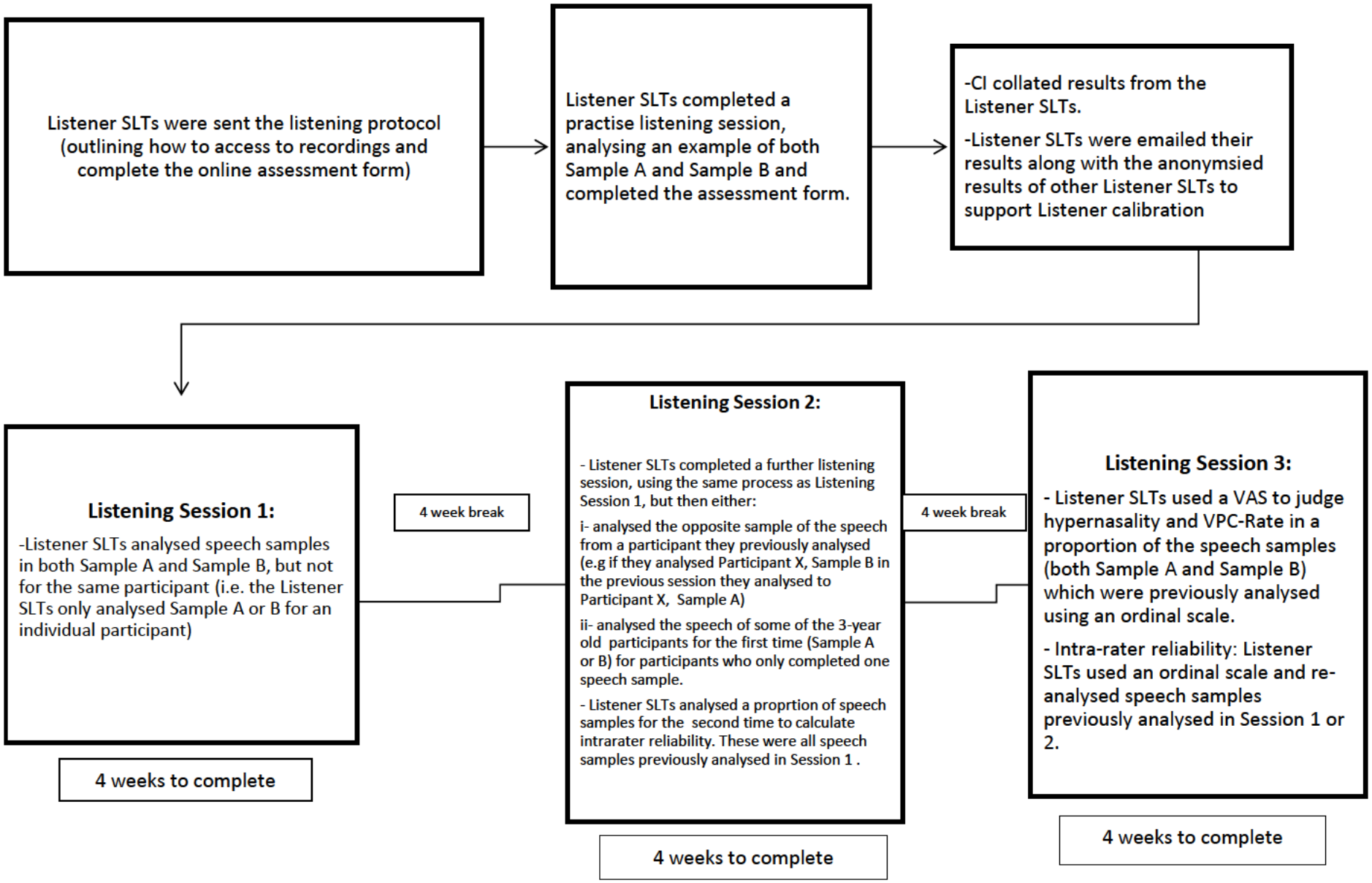


Figure 4.8 Flowchart of the Listening Process

4.4.7 Procedure: Completion of the Listener SLT Questionnaires

After the Listener SLTs had completed all three listening sessions, they were asked to complete an electronic questionnaire (Appendix AA). The objective of this questionnaire was to gain information about the acceptability and usability of the speech assessment and rating methods including the use of VPC-Rate and the VAS. The questionnaire was designed using Qualtrics software and was completed online by the Listener SLTs. The questions primarily used multiple choice or a Likert-scale to obtain the SLT Listeners' feedback. Optional free-text questions were also used after each multiple choice or Likert-scale questions to allow the Listener SLTs to expand upon their answers or add any additional relevant comments.

The questionnaire was designed using elements of a 'Likert-type' scale which, as described by Kaptein et al. (2010), is appropriate for use in studies of usability, attitudes, and judgements, and was therefore well suited to the objective of this questionnaire. For all questions on the Likert scale, 0 represented the most negative answer and 10 the most positive, except for the two questions relating to the amount of time it took to analyse the speech samples. In this case, 0 referred to "too long" and 10 "too fast"- with 5 being the ideal. The results of the questionnaire were collated and presented using descriptive statistics to summarise the responses. The free text responses were synthesised to define key topics, and the responses were presented with reference to these topics.

4.5 Data Analysis

Statistical Analysis was carried out using IBM SPSS version 26 and Microsoft Excel 2016. To achieve the aims and objectives of the study both descriptive and inferential statistics were used. For inferential statistics, a significance level of $p < .05$ was adopted.

4.5.1 Validity

4.5.1.i Content Validity

The parameters of speech analysed by the Listener SLTs were compared with the parameters of speech utilised by the Study SLT and recorded on the Speech Parameters Recording Form. The results were used to highlight differences and similarities between the parameters assessed and discuss the validity of the assessment tool. Information from the Listener SLT questionnaire and descriptive statistics were used to discuss similarities between the speech parameters utilised in this study and those the Listener SLTs use in their clinical practice.

4.5.1.ii Construct Validity

The assessment tool and the speech samples were predominately designed to measure cleft speech. Completion of Phase 1 of the study informed the parameters of speech subsequently included in the assessment tool contributing to the construct validity of the tool. To gain information about the specificity of the assessment, and the construct validity of the assessment, the number of Control Group participants judged by the Listener SLTs as having speech characteristics associated with CP±L was examined.

Taking into consideration the strict inclusion and exclusion criteria for the Control Group participants (which aimed to reduce the likelihood of 3-year olds with any communication difficulties being recruited to the Control Group) it was hypothesised that the number of Control Group participants judged as having speech characteristics associated with CP±L would be minimal. This would be reflected in a 'green outcome' on the Adapted CAPS-A. Conversely, given that at age-5 years 50% of individuals with CP±L still have speech characteristics associated with CP±L (Britton et al. 2014), it was hypothesised that at least 50% of the 3-year old participants with CP±L would be judged by the Listener SLTs as having speech characteristics associated with CP±L. Descriptive statistics were used to examine the number of Control Group participants judged as having speech characteristics associated with CP±L.

4.5.2 Reliability

The judgements made by the Listener SLTs were statistically analysed to measure both inter and intra-rater reliability. Reliability statistics were then compared between the speech samples. To examine the impact of different scales on reliability scores for VPC-Rate and hypernasality, a comparison was made between reliability scores on both an ordinal scale and VAS for these parameters.

Descriptive statistics were used to describe inter and intra-listener agreement. Total percentage agreement, where there was 100% agreement for all Listener SLTs, and majority agreement when 6/7 Listener SLTs agreed, were both used. Percentage agreement was presented alongside inferential statistics due to the presence of a statistical artifact previously reported in studies using the CAPS-A (Baillie & Sell. 2020; Bruneel et al. 2020) which can occur due to a lack of variability in the use of the full CAPS-A scale.

To calculate inter-rater reliability for both ordinal and VAS data, single-measures Intraclass Correlation Coefficients (ICC) were used. ICC scores compared the variability of different judgements for the same participant to the total variation across all judgements in the wider participant group. The following parameters of speech were analysed using ICC: VPC-Rate, hypernasality, hyponasality, NAE, and all CSCs. For nominal data, Krippendorf's alpha (KALPHA) was used, as this approach supported dichotomous categorical data. The following phonological measures were used: phonological processes present; age-appropriate phonological processes; delayed phonological processes; and disordered phonological processes. Inter-rater reliability scores and percentage agreement were interpreted using Altman's (1991) descriptors as shown in Table 4.11.

Table 4.11. Reliability descriptors used for inter-rater reliability (Altman 1991).

To calculate intra-rater reliability paired *t*-tests were used to measure the correlation between SLT judgements on the first and second occasions (with exception when there was insufficient variability in the data). For Sample A, 26.3% (5/19 video recordings) were analysed twice to calculate intra-rater reliability. For Sample B, 35.7% (5/14 video recordings) were analysed twice to calculate intra-rater reliability. Correlation was interpreted using the descriptors set out by Schober et al. (2018) as presented in Table 4.12.

Table 4.12. Correlation descriptors used for intra-rater reliability (Schober et al. 2018).

4.5.3 Acceptability and Usability

The results of the Listener SLT questionnaire were automatically collated by the Qualtrics software and exported into an excel spreadsheet for analysis. The free text responses were synthesised to gain feedback on the SLT Listener's perceptions and experience of the speech samples, methods and scales used, and their views on speech assessment at age 3-years. This feedback was impressionistic and was not analysed with a qualitative approach (Mills & Birks 2014). The CI organised SLT Listener feedback around emerging topics based on the impressionistic comments.

4.5.4 Completion Rates

The number of CP±L group and Control Group participants randomised to each speech sample is reported in the results section. Information regarding the level of completion of each sample at the end of the assessment session was collated. Descriptive statistics were used to describe the completion rates, links to randomisation, and any differences noted between the CP±L group and Control Group participants.

Chapter 5. Phase 2: Results- Validity and Reliability of the Speech Assessment Framework

5.1 Participants Demographic Information

5.1.1 CP±L Group

Participants were aged between 3 years 0 months and 3 years 10 months (mean age 3 years 4 months ±standard deviation [SD] 3.28 months), range 10 months, median 3 years 5 months, modes: 3 years 1 month, 3 years 2 months and 3 years 7 months.

Participants (n= 20 children with CP±L; n=13 male) were born in 2015 and 2016 with 25% bilingual (languages spoken other than English: Polish, Punjabi, Urdu). All the participants lived within the West Midlands and all cleft types were represented (Figure 5.1)

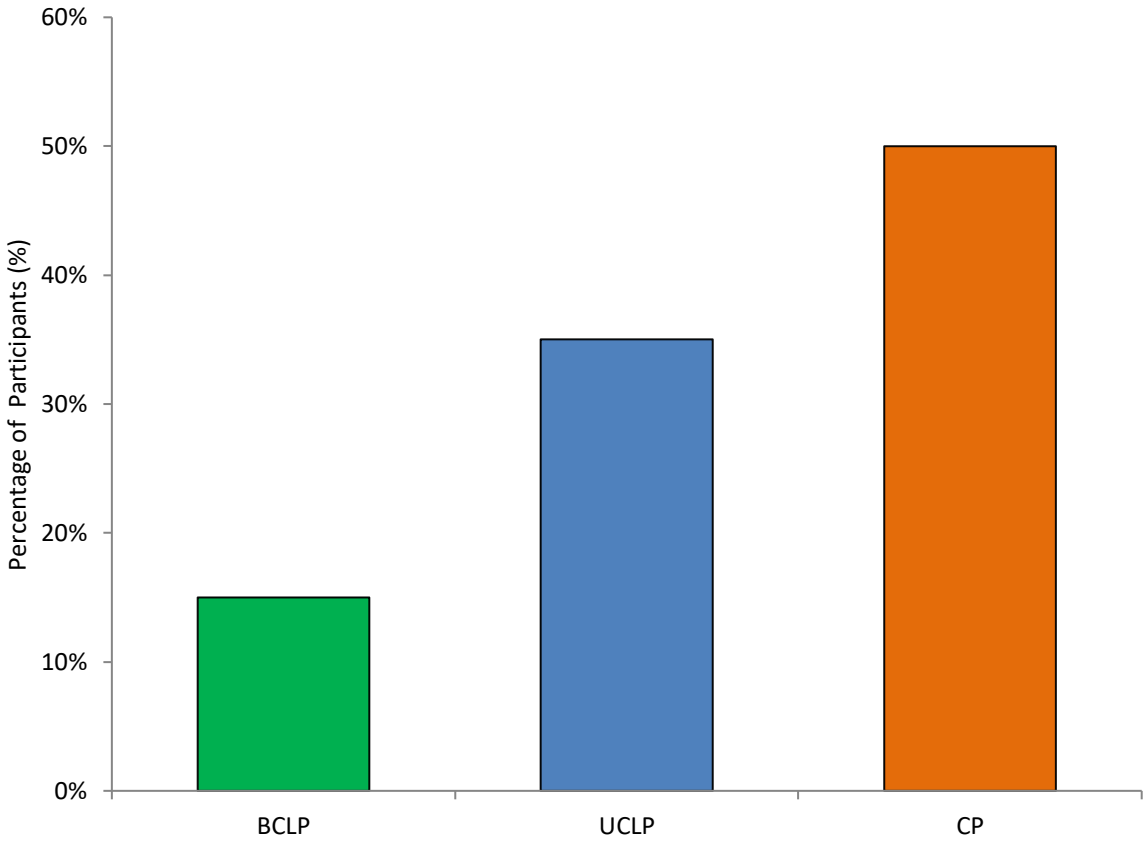


Figure 5.1. Percentage of CP±L Group by Cleft Type

5.1.2 Control Group

The Control Group participants (n=5) were not matched for gender, ethnicity or bilingualism. All 5 children (n=4 male) were from monolingual families and lived within the West Midlands. The participants were aged between 3 years 1 month and 3 years 7 months

with a range of 6 months. The mean age in this group was 3 years 4 months (±2.39 months), the median was 3 years 3 months and there was no mode.

5.2 Speech sample completion rates

In the control group 100% of the participants completed both speech samples in their entirety within the single assessment session. In contrast, 14 participants (70%) in the CP±L group completed both speech samples in their entirety within the assessment session. One participant (5%) with CP±L did not attempt any of either speech sample. Completion rates for the speech samples by the participants with CP±L are shown in Figure 5.2.

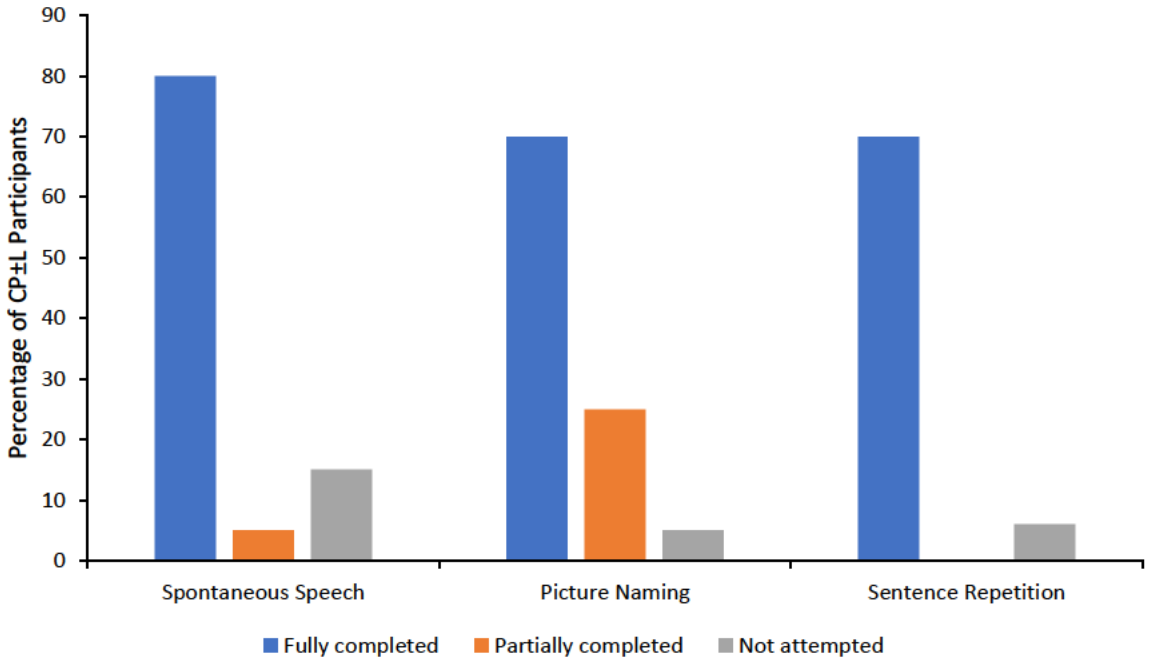


Figure 5.2 Completion rates of the speech samples by the CP±L group

As a result in the variation of completion rates in the CP±L group, completion rates for Sample A are separately reported for spontaneous speech and picture naming. Three participants did not attempt the spontaneous speech sample (15%), compared to only one participant (5%) who did not attempt any of the picture naming. Some participants attempted but were unable to complete all of the picture naming (n= 5, 25%). Similarly, one participant (5%) attempted the sample of spontaneous speech but was not able to complete the full two-minute sample. In total, 14 of the participants (70%) completed all of the picture naming, and 16 participants (80%) completed all of the spontaneous speech sample. Overall, 17 participants (85%) completed some of both the spontaneous speech and picture naming tasks. The same 14 participants (70%) who completed both the spontaneous speech and

picture naming samples in their entirety also completed all the sentence repetition sample (Sample B) irrespective of randomisation. The participants who only managed to partially complete either the spontaneous speech or picture naming samples could not complete Sample B. For the five participants (25%) who only partially competed Sample A, three presented with a language delay, one had an attention delay, and one was described by parents as being particularly shy. The participant not completing either speech sample was bilingual and their parents reported at the assessment that they were going through a silent period (a recognised stage of bi/multilingual development [Harris 2019]) as well as feeling unwell. Of these six participants, four were randomised to complete Sample A first, and two to complete Sample B first and their ages ranged from 3 years 0 months to 3 years 7 months, with a mean age of 3 years 3 months (±3.21 months).

The mean length of time for participants in the CP±L group to complete Sample A was 17 minutes 20 seconds (± 2:58 minutes and seconds). The mean time to complete Sample B was 5 minutes and 54 seconds (± 1:23 minutes and seconds). The difference in the time taken to complete Sample A versus Sample B was significant for the CP±L group (t (13) = 19.417 , p <0.001). For the Control Group the mean length of time for participants to complete Sample A was 15 minutes and 02 seconds (± 2:36 minutes and seconds). The mean length of time for Sample B was 5 minutes and 06 seconds (± 01:21 minutes and seconds). The difference in the time taken to complete Sample A versus Sample B was also significant for the Control Group (t (4) = 10.599 , p <0.001). There was no significant difference in the mean length of time taken for the CP±L group to complete Sample A compared to the Control Group (t (4) = 1.561 , p = 0.194) or to complete Sample B compared to the Control Group (t (4) = 0.448 , p = 0.677).

5.3 Specificity analysis: Control Group

For Sample A there were 35 ratings for each parameter of speech (totalling 560 listening instances), for Sample B there were 34 ratings due to a playback issue for one of the video recordings which occurred for one Listener SLT only (totalling 544 listening instances). The key outcome for specificity was a 'green' outcome on the adapted CAPS-A traffic light scale (with dark and light green outcomes collapsed as per Britton et al. [2014]).

5.3.1 Nasality, nasal airflow and Cleft Speech Characteristics

For Sample A, fifteen of the sixteen cleft specific parameters of speech were judged by all the Listener SLTs to be 'normal' (green outcome) (93.8%). For Sample B, fourteen of

these parameters were judged to be 'normal' (87.5%) by all of the Listener SLTs. This demonstrates very high levels of listener agreement relating to absence of these cleft specific parameters of speech. For those parameters of speech in which not all of the Listener SLTs agreed on a normal outcome, agreement relating to the absence of these parameters was >90% for hypernasality (Sample A) and Palatalisation/ palatal and Backed to velar/uvular (Sample B), indicating that most Listener SLTs agreed on the absence of these parameters of speech, and that those Listener SLTs identifying the presence of these parameters were outliers.

The percentage of green outcomes for each parameter for both samples in outlined in Table 5.1 below. Whilst the only possible outcome on the traffic light scale for Dentalisation/inter-dentalisation was a green outcome, the rating 'absent' was only used in 71.4% of Listener SLT judgements for Sample A, and 64.7% of Listener Judgements for Sample B, highlighting the increased frequency of this CSC in the speech of participants in the Control Group.

Table 5.1. Percentage of 'green' outcomes for each parameter of speech as judged by Listener SLTs

	% 'green' outcomes	
	Sample A	**Sample B**
VPC-Rate	100	100
Hypernasality	97.1	100
Hyponasality	100	100
NAE	100	100
Dentalisation/inter-dentalisation (all outcomes scores as green)	100	100
Lateralisation/ lateral	100	100
Palatalisation/ palatal	100	97.1
Double articulation	100	100
Backed to velar/uvular	100	94.1
Pharyngeal articulation	100	100
Glottal articulation	100	100
Active nasal fricatives	100	100

Double articulation with glottal	100	100
Weak and/or nasalised consonants	100	100
Nasal realisation of plosives &/or suspected passive nasal fricative	100	100
Gliding of fricatives/affricates	100	100

5.3.2 Phonology

The reliability of Listener SLT judgements and percentage of scores relating to the presence of phonological processes is shown in Table 5.2. Whilst the Krippendorf's alpha score was poor for each variable, this was impacted by the lack of variability in the data as majority agreement scores were relatively high, except for age-appropriate phonological processes for Sample B. The percentage of the ratings indicating the presence of phonological processes illustrates that only a small percent of participants in the Control Group had no evidence of phonological processes, 11.4% on Sample A and 3.0% for Sample B. Age-appropriate phonological processes were frequently reported for both speech samples, 88.6% in Sample A and 88.3% in Sample B. Delayed phonological processes were rated as present more frequently in Sample B, 20.6% compared to 11.4% for Sample A, however disordered phonological processes were rated as occurring infrequently, in 5.7% of Sample A ratings and 5.8% of Sample B ratings.

Table 5.2. Reliability (Krippendorf's alpha: KALPHA), percentage agreement, and percentage of scores marked as present for the control group.

	Sample A					**Sample B**				
	n	**KALPHA**	**Interpretation**	**% Majority Agreement**	**% present**	**n**	**KALPHA**	**Interpretation**	**% Majority Agreement**	**% present**
Phonological processes present	35	0.1774	poor	80	88.6	35	0.0286	poor	100	97.0
Age-appropriate phonological processes	35	0.1774	poor	80	88.6	35	0.0093	poor	60	88.3
Delayed phonological processes	35	-0.0968	poor	100	11.4	35	0.2119	poor	80	20.6

Disordered phonological processes	35	-0.0303	poor	100	5.7	35	0.0185	poor	100	5.8

5.4 CP±L group: Validity & Reliability

5.4.1 Comparison of the speech parameters in the Adapted CAPS-A and clinical assessment

There are strong parallels between those parameters of speech in the Adapted CAPS-A and those which featured in the Study SLT's clinical assessment in the CP±L group. Phonological processes and CSCs featured most frequently in the Study SLT's assessment of the participants' speech, with phonological processes reported in 100% of the clinical assessments and CSCs in 94.7% of assessments. Both VP Function and hypernasality were also frequently used in 68.4% of the Study SLT's assessments. In comparison, hyponasality was only featured in 10.5% of the Study SLT's assessments. Neither voice nor intelligibility featured in the Study SLT's assessments and were not used in the Adapted CAPS-A.

5.4.2 Reliability: CP±L Group

There are frequent examples in the results of high agreement of 'green' outcomes by the Listener SLTs. The lack of agreement for outcomes other than 'green' and thus the lack of variability in the use of the full Adapted CAPS-A scales by the Listener SLTs for some parameters impacted the ICC, Kripperndorf's alpha and correlation coefficient, resulting in poor ICC scores but high percentage agreement[1]. As a result of this statistical artifact there are instances in which SPSS could not return the ICC statistic, Krippendorf's alpha, or correlation coefficient.

5.4.2.i Resonance and NAE

The seven Listener SLTs analysed 19 video recordings for Sample A, and 14 video recordings for Sample B. Data was missing for one rating of NAE for one listener only for Sample B. For Sample A, there were a maximum of 133 listening instances for each parameter. There were three instances in which a Listener SLT used 'unable to score' for hypernasality, and two for NAE. There is one video recording of Sample A, which was part completed, in which a single Listener SLT used 'unable to score' for every parameter. For Sample B, there were a maximum of 98 listening instances for each parameter of speech. For Sample B,

[1] Personal correspondence with Head of Statistics at SIGMA: Mathematics & Statistics Support Service, Coventry University (April 2021)

'unable to score' was only used once, for a single hypernasality rating. These ratings were subsequently removed from the ICC analyses.

To ensure that measures of reliability were not impacted by the higher number of video recordings and partially completed speech samples in Sample A (n=19), analysis of the scores for the 14 speech samples for Sample A which were fully completed was also undertaken. This allowed for comparisons with all the recordings for Sample A, to determine if partially completed speech samples impacted reliability scores. It also allowed for a direct comparison with Sample B, as these were the same participants, and thus the only difference was the speech sample itself.

5.4.2.i.a Inter-Rater Reliability

Inter-rater reliability ICC scores and percentage agreement for hypernasality, hyponasality and NAE are shown in Table 5.3 for both speech samples, and the matched samples i.e. those participants who fully completed both Sample A and Sample B. For each parameter the full spectrum of scalar points were utilised by the Listener SLTs for Sample A and Sample B. In the Sample A matched recordings, the full scale was used for hypernasality and NAE, but the rating of 'marked' hyponasality was not used.

For hypernasality, the ICC was good for Sample A and only moderate for Sample B, but majority agreement levels are broadly comparable, 73.7% for Sample A and 78.6% for Sample B. For hyponasality, both ICC scores are very low, despite high majority agreement scores for both speech samples. ICC scores for this parameter were impacted by the majority of scores given by the Listener SLTs corresponding to a 'green' traffic light outcome. This statistical artifact is illustrated as scatterplots (Figures 5.3 and 5.4) which show all of the Listener SLT scores for each video recording for the parameter of hyponasality. For NAE, the ICC scores were good for both Sample A and Sample B, alongside high levels of majority agreement for both samples. Based upon ICC scores only, Sample A had stronger reliability for resonance and NAE, however, majority agreement scores are comparable across the speech samples with strong levels of agreement for all these variables.

Table 5.3. Inter-Rater Reliability ratings for each speech sample for Resonance and NAE

	Sample A (19 video recordings)				Sample A (14 matched video recordings)				Sample B (14 matched video recordings)			
	ICC	Interpretation	% Total Agreement	% Majority Agreement	ICC	Interpretation	% Total Agreement	% Majority Agreement	ICC	Interpretation	% Total Agreement	% Majority Agreement
Hypernasality	0.790	good	42.1	73.7	0.793	good	57.1	85.7	0.489	moderate	28.6	78.6
Hyponasality	0.167	poor	89.5	94.8	-	-	100	100	0.000	poor	92.9	100
NAE	0.694	good	78.9	94.8	0.741	good	78.6	92.9	0.718	good	78.6	92.9

Missing ICC values: correlation could not be computed, and SPSS did not return a value

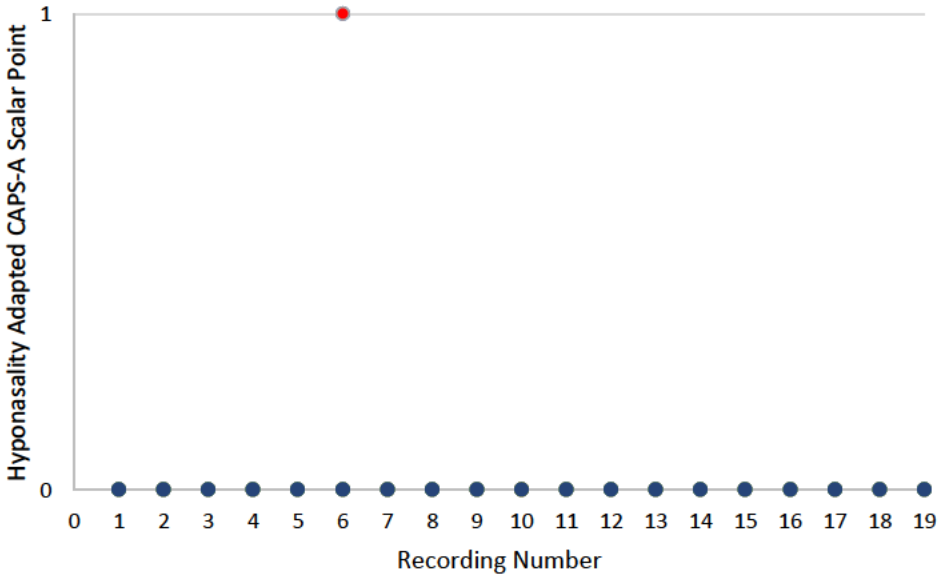


Figure 5.3. Listener SLT scores for Sample A Hyponasality

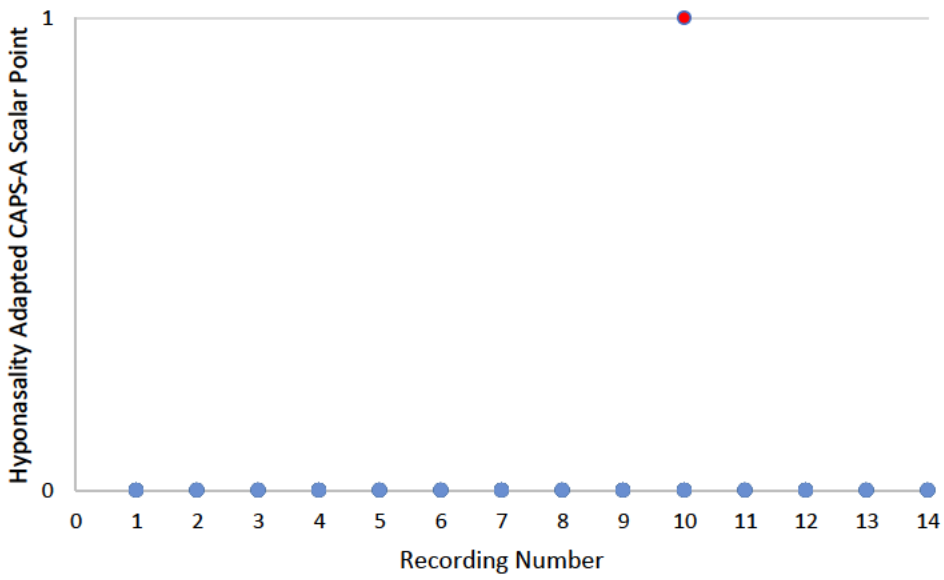


Figure 5.4. Listener SLT scores for Sample B Hyponasality

Examining Sample A, ICC scores are comparable between the group of all the participants, and those participants with matched video recordings for both hypernasality and NAE (Hypernasality: 0.790 for all participants, 0.793 for matched group. NAE: 0.694 for all participants, 0.741 for the matched group). Majority agreement scores were higher for the matched samples in every parameter except NAE, in which there was a slight decrease (94.8% for all participants, 92.9% in the matched group). For hyponasality 100% of outcomes in the matched group were green, highlighting that all participants fully completing Sample A were judged to have a normal speech outcome for this parameter. When comparing the sample of 14 participants who fully completed Sample A and Sample B, both ICC and majority agreement are higher in Sample A for hypernasality. For the 14 participants who completed all of Sample A and Sample B the parameters of hyponasality and NAE were judged consistently, irrespective of the speech sample i.e. majority agreement for hyponasality was 100%, and majority agreement for NAE was 92.9% for both speech samples.

5.4.2.i.b Intra-Rater Reliability

There was no missing data, although the rating 'unable to score' was used in both samples, more frequently for Sample A as reflected in the number of ratings used to calculate intra-rater reliability, shown in Table 5.4.

Table 5.4. Intra-rater Reliability: correlation scores for each speech sample for VPC-Rate, Resonance and NAE

	Sample A (19 videos)				Sample B (14 videos)			
	n	***r***	**Interpretation**	**% Agreement**	**n**	***r***	**Interpretation**	**% Agreement**
Hypernasality	32	0.891	strong	71.9	34	0.889	strong	94.1
Hyponasality	34	0.696	moderate	97.0	35	-	-	100
NAE	32	-	-	96.9	35	-	-	100

Missing r value: correlation could not be computed and SPSS did not return a value

Intra-rater reliability was good for both samples, demonstrating that the Listener SLTs applied consistent internal standards for these parameters of speech irrespective of the speech sample. Hypernasality had consistently strong correlation for both samples, although percentage agreement was higher for Sample B. Percentage agreement was >90% for both hyponasality and NAE for both samples, but this agreement was 100% for Sample B. Overall,

Sample B showed stronger intra-rater reliability. However, there were more examples of 'green' outcomes in Sample B, which indicates that the Listener SLT's were more consistent in their internal judgements as to when the parameter was absent.

5.4.2.ii Cleft Speech Characteristics

5.4.2.ii.a Inter-Rater Reliability

All CSCs in the assessment framework were judged to be present in the Sample A recordings. In Sample B, the only CSC which was not rated as present was that of double articulation. Inter-Rater reliability was calculated for twelve CSCs, this is presented in Table 5.5. For Sample A only three CSCs had good reliability scores based on the ICC (backed to velar/uvular, glottal articulation, weak and/or nasalised consonants). For Sample B, only one CSC had good reliability based on the ICC (backed to velar/uvular). However, when considering majority agreement, 11/12 CSCs had percentage agreement >61% (exception being dentalisation/inter-dentalisation) for both speech samples.

For Sample A, 6/12 CSCs had majority agreement >81% (double articulation, pharyngeal articulation, active nasal fricatives, double articulation with a glottal, nasal realisation of plosives &/or suspected passive nasal fricative, gliding of fricatives/affricates). The same six CSCs had majority agreement >81% for the matched recordings of Sample A (Table 5.5). For Sample B, 4/12 CSCs had percentage agreement >81% (double articulation, pharyngeal articulation, nasal realisation of plosives &/or suspected passive nasal fricative, gliding of fricatives/affricates). This indicates that for the following CSCs: double articulation, pharyngeal articulation, nasal realisation of plosives &/or suspected passive nasal fricative, gliding of fricatives/affricates, that there were high levels of Listener SLT agreement irrespective of the speech sample. Overall, Sample B has seven CSCs with higher majority agreement scores than Sample A, compared to four CSCs with higher majority agreement in Sample A.

Majority percentage agreement was higher in the matched recordings for Sample A for 9/12 CSCs when compared to all of the recordings for Sample A. This suggests that the Listener SLTs were less reliable in their analysis of these nine CSCs when analysing the part completed speech samples. When compared to Sample B, majority agreement was higher in 6/12 CSCs for the matched Sample A, and agreement was the same in 5/12 CSCs (backed to velar/uvular, pharyngeal articulation, active nasal fricatives, double articulation with a glottal,

gliding of fricatives/affricates) highlighting that only 1/12 CSCs had higher majority agreement for Sample B. ICC scores were more variable between the matched Sample A recordings and Sample B.

Table 5.5. Inter-Rater Reliability ratings for CSCs

		Sample A (19 video recordings)				Sample A (14 matched video recordings)				Sample B (14 matched video recordings)			
		ICC	**Interpretation**	**% Total Agreement**	**% Majority Agreement**	**ICC**	**Interpretation**	**% Total Agreement**	**% Majority Agreement**	**ICC**	**Interpretation**	**% Total Agreement**	**% Majority Agreement**
Anterior Oral Speech Characteristics	Dentalisation/inter-dentalisation	0.394	fair	21.6	32.2	0.389	fair	21.4	21.4	0.477	moderate	14.3	21.4
	Lateralisation/ lateral	0.463	moderate	42.1	73.7	0.484	moderate	35.7	64.3	0.167	poor	57.1	78.6
	Palatalisation/ palatal	0.279	fair	42.1	63.2	0.352	fair	50.0	71.4	0.195	poor	64.2	64.3
Posterior Oral CSCs	Double articulation	-0.020	poor	78.9	94.7	-0.013	poor	85.7	100	-	-	100	100
	Backed to velar/uvular	0.631	good	42.1	63.2	0.695	good	42.9	64.3	0.764	good	57.1	64.3
Non-oral CSCs	Pharyngeal articulation	0.000	poor	89.5	100	-	-	100	100	0.000	poor	92.9	100
	Glottal articulation	0.644	good	52.6	68.4	0.464	moderate	64.3	78.6	0.570	moderate	64.3	71.4
	Active nasal fricatives	0.192	poor	57.9	84.2	0.172	poor	57.1	85.7	0.170	poor	78.6	78.6
	Double articulation with glottal	0.167	poor	89.5	94.7	-	-	100	100	0.000	poor	78.6	78.6
Passive CSCs	Weak and/or nasalised consonants	0.807	good	68.4	79.5	0.758	good	78.6	78.6	0.415	moderate	42.9	64.3
	Nasal realisation of plosives &/or suspected passive nasal fricative	0.441	moderate	73.7	84.2	0.130	poor	85.7	100	0.000	poor	85.7	100
	Gliding of fricatives/affricates	0.442	moderate	78.9	84.2	0.667	good	92.9	100	0.568	moderate	78.6	85.7

Missing ICC values: correlation could not be computed, and SPSS did not return a value

In both samples, several CSCs were impacted by a lack of variability in the data, which resulted in low ICC scores but high percentage agreement. For example, for pharyngeal articulation, the ICC for both speech samples was less than 0.000 indicating poor agreement. However, majority agreement was 100% for both speech samples. For Sample A, the full spectrum of each scale was used by the Listener SLTs for 9/12 CSCs, but only green and amber ratings were used for the CSCs active nasal fricative and double articulation with a glottal, and only green and red ratings were used for pharyngeal articulation. For Sample B, for 8/12 CSCs, the full spectrum of each scale was again used, but only green ratings were used for both lateralisation/lateral and double articulation, and only green and amber ratings used for pharyngeal articulation and nasal realisation of plosives &/or suspected passive nasal fricative. Of all the CSCs dentalisation/inter-dentalisation had the lowest levels of percentage agreement. Majority percentage agreement was 32.2% for Sample A, 21.4% for the matched recordings for Sample A, and 21.4% for Sample B.

5.4.2.ii.b Intra-Rater Reliability

Intra-rater reliability was calculated using paired *t*-tests. There was no missing data, although the rating 'unable to score' was used in Sample A by one Listener SLT. Correlation and percentage agreement for CSCs is presented in Table 5.6. Like intra-rater reliability scores for resonance and NAE, both correlation and percentage agreement are higher for Sample B. However, there was greater variability in the scalar points used in Sample A than Sample B. Percentage agreement was >70% for every CSC for both samples, indicating strong reliability overall, although agreement was >90% for 6/12 CSCs for Sample A, and 10/12 for Sample B, highlighting stronger intra-rater reliability for Sample B.

Table 5.6. Intra-rater Reliability: correlation scores for each speech sample for CSCs

	Sample A				Sample B			
	n	***r***	**Interpretation**	**% Agreement**	**n**	***r***	**Interpretation**	**% Agreement**
Dentalisation/inter-dentalisation	34	0.622	moderate	73.5	35	0.771	strong	77.1
Lateralisation/ lateral	34	-	-	100	35	-	-	100
Palatalisation/ palatal	34	0.672	moderate	91.2	35	0.718	strong	91.4
Double articulation	34	-	-	97.1	35	-	-	100
Backed to velar/uvular	34	0.674	moderate	73.5	35	0.858	strong	88.6
Pharyngeal articulation	34	-	-	97.1	35	-	-	100
Glottal articulation	34	0.893	strong	82.4	35	-	-	100
Active nasal fricatives	34	0.684	moderate	94.1	35	-	-	100
Double articulation with glottal	34	0.696	moderate	97.1	35	-	-	100
Weak and/or nasalised consonants	34	0.853	strong	82.3	35	0.946	very strong	97.1
Nasal realisation of plosives &/or suspected passive nasal fricative	34	0.907	very strong	88.2	35	-	-	100
Gliding of fricatives/affricates	34	0.498	moderate	85.2	35	0.852	strong	97.1

Missing r value: correlation could not be computed and SPSS did not return a value

5.4.2.iii Phonology Outcomes

5.4.2.iii.a Inter-Rater Reliability

The results for phonology outcomes are presented in Table 5.7. For phonological processes present, there are high levels of majority agreement for both samples but the KALPHA score is poor for both Sample A and Sample B; relating to the high frequency with which an "phonological processes present" was indicated by the Listener SLTs. KALPHA scores are <0.50 for all phonology measures for both speech samples. Based on the majority agreement scores, Sample B has higher agreement scores for age-appropriate phonological processes and delayed phonological processes, whilst Sample A has higher agreement for disordered phonological processes.

Table 5.7. Inter-Rater Reliability: Phonology Outcomes (Krippendorf's alpha: KALPHA)

	Sample A (19 video recordings)					Sample B (14 video recordings)				
	n	**KALPHA**	**Interpretation**	**% Total Agreement**	**% Majority Agreement**	**n**	**KALPHA**	**Interpretation**	**% Total Agreement**	**% Majority Agreement**
Phonological processes present	133	0.0522	poor	84.2	100	98	0.1656	poor	78.6	92.9
Age Appropriate Phonological processes	133	0.0719	poor	26.3	31.6	98	0.1567	poor	28.6	50
Delayed Phonological processes	133	0.3034	fair	26.3	57.9	98	0.4021	fair	28.6	78.6
Disordered Phonological processes	133	0.4902	moderate	41.1	67.4	98	0.1257	poor	14.3	35.7

5.4.2.iii.b Intra-Rater Reliability

Intra-rater reliability for phonology outcomes is presented in Table 5.8. There was no missing data. Like the inter-rater reliability score relating to the presence of phonological processes, the intra-rater percentage agreement was high for this parameter for both speech samples, 94.3% for Sample A and 88.6% for Sample B. However, correlation was negligible for Sample A, and was impacted by the prevalent rating of "phonological processes present," although this did result in consistent agreement, as seen in the percentage agreement score. For Sample B, the Listener SLTs more frequently rated "absence of phonological processes" but percentage agreement indicates they were less consistent in their judgements. Intra-rater reliability correlation for age-appropriate phonological processes was moderate for both speech samples. Similarly, for disordered phonological processes intra-rater reliability was high based upon both correlation and percentage agreement. In contrast, Sample B shows superior intra-rater reliability for delayed phonological processes based on both correlation and percentage agreement.

Table 5.8. Intra-rater Reliability: correlation scores for each speech sample for Phonological Processes

	Sample A (5 video recordings)				**Sample B (5 video recordings)**			
	n	***r***	**Interpretation**	**% Agreement**	**n**	***r***	**Interpretation**	**% Agreement**
No phonological processes present	35	-0.029	negligible	94.3	35	0.435	moderate	88.6
Age Appropriate Phonological processes	35	0.443	moderate	71.4	35	0.511	moderate	80.0
Delayed Phonological processes	35	0.318	weak	65.7	35	0.614	moderate	85.7
Disordered Phonological processes	35	0.798	strong	91.4	35	0.770	strong	91.4

5.4.2.iv VPC-Rate

The full range of scalar points were used for both Sample A and Sample B. For inter-rater reliability, there were a total of 131 ratings of VPC-Rate for Sample A (excluding two 'unable to score') and 98 ratings for Sample B. ICC scores for inter-rater reliability were higher for Sample A than Sample B. The ICC score was .841 for Sample A and .822 for the matched Sample A recordings, both indicating very good reliability. In comparison, for Sample B the

ICC is .561, indicating moderate reliability. Majority agreement was similar, 68.4% for Sample A and 64.3% for Sample B. However, majority agreement was highest in the matched recordings for Sample A, 78.6%.

Intra-rater reliability scores were high for both speech samples, with strong correlations between ratings for both speech samples. The *r* value for Sample A was .879 and for Sample B .869. The results for both inter-rater and intra-rater reliability show that VPC-Rate can be reliably used as an outcome measure at age 3-years in the CP±L population, with Sample A having stronger reliability based on inter-rater reliability.

5.5 Comparison of reliability scores using ordinal and Visual Analogue Scale

Both VPC-Rate and hypernasality were analysed by the Listener SLTs using a visual analogue scale (VAS): Sample A, 31.6% of video recordings; Sample B, 21.4% of video recordings. A comparison of the ICC scores for VPC- Rate and Hypernasality when rated using ordinal scales or VAS is shown in Table 5.9. The results for matched recordings i.e. the same recordings scored using ordinal scales and VAS are also presented.

All ICC scores are higher for the ordinal ratings than for VAS for Sample A for both VPC-Rate and hypernasality. This is evident irrespective of whether all recordings were used or only those matched. For Sample B, there is less difference between the ordinal and VAS scores for both VPC-Rate and hypernasality, because reliability was only moderate for these parameters on the ordinal scale. However, when analysing the matched ordinal scores for Sample B, this suggests that VAS scores are more reliable i.e. moderate reliability for VPC-Rate using the VAS compared to poor on the ordinal scale, and fair reliability using VAS compared to poor on the ordinal scale.

Table 5.9. Sample A and Sample B: Comparison of Ordinal scale and VAS ratings for VPC-Rate and Hypernasality

	Ordinal Sample A			**Ordinal Sample B**		
	n	**ICC**	**Interpretation**	**n**	**ICC**	**Interpretation**
VPC-Rate	131	0.841	very good	98	0.489	moderate
Hypernasality	130	0.790	good	97	0.573	moderate
	Ordinal Sample A Matched			**Ordinal Sample B Matched**		
	n	**ICC**	**Interpretation**	**n**	**ICC**	**Interpretation**
VPC-Rate	41	0.818	very good	21	0.000	poor
Hypernasality	41	0.803	good	21	-0.056	poor
	VAS Sample A			**VAS Sample B**		
	n	**ICC**	**Interpretation**		**ICC**	**Interpretation**
VPC-Rate	42	0.586	moderate	21	0.476	moderate
Hypernasality	42	0.606	moderate	21	0.345	fair

5.6 Listener SLT Feedback

5.6.1 Multiple choice and Likert-scale responses

The Listener SLTs completed a questionnaire to gain information about how they viewed the acceptability and usability of the two speech samples. In total 6/7 of the listeners completed the questionnaire (85.7 %). Table 5.10 summarises the multiple-choice feedback relating to the speech sample, and Table 5.11 summarises all the Likert scale questions

The Listener SLTs reported that the 3-year old participants engaged with both speech samples. Whilst indicating that they thought the 3-year olds completed Sample A more fully, and that this speech sample (spontaneous speech and single word naming) most closely matched the speech sample they used in their clinical practice, the Listener SLTs indicated that they would be more likely to use Sample B or both speech samples in their clinical practice. Indeed, 66.7% of the Listener SLTs reported that they found Sample B easier to analyse. Examining the amount of time it took the Listener SLTs to analyse the speech samples Sample B was preferred. Analysis of their responses indicates that it is the opinion of the Listener SLTs that Sample A took too long to analyse, compared to Sample B which took an optimal time, or slightly too short a time.

Table 5.10 Summary of Listener SLT responses for multiple choice questions relating to the speech samples.

Question	Sample A		Sample B		Both		Neither	
	N	%	N	%	N	%	N	%
Which sample did you find it easier to analyse?	0	0	4	66.7	2	33.3	0	0
Which sample would you prefer to use when assessing children at age 3-years with cleft palate?	0	0	2	33.3	4	66.7	0	0
Which sample do you think the 3-year old children engaged with the most?	1	16.7	2	33.3	3	50.0	0	0
Which sample do you think the 3-year old children most fully completed?	3	50.0	2	33.3	1	16.7	0	0
Which speech sample most closely matches the speech sample you use in clinical practice when assessing speech in 3-year old children with cleft palate?	3	50.0	1	16.7	2	33.3	0	0
Which speech sample do you think you would be most likely to use in your future clinical practice if available?	1	16.7	2	33.3	3	50.0	0	0

Table 5.11. Summary of Listener SLT responses for Likert-scale questions.

Method 1= the separate analysis of resonance and nasal airflow errors
0 = most negative answer
10= the most positive answer
For the two questions relating to the amount of time it took to analyse the speech samples: 0 = “too long”, 10 = “too fast”, 5 = the ideal length of time

Question	**n= responses**										
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
How closely does **Method 1** match the methods you use to assess the speech characteristics associated with velopharyngeal function?									2	2	2
How closely does **VPC-Rate** match the methods you use to assess the speech characteristics associated with velopharyngeal function?					1	1	2		1		1
How acceptable to you is **Method 1** to assess the speech characteristics associated with velopharyngeal function in 3-year old children with cleft palate?						1		1	1	2	1
How acceptable to you is **VPC-Rate** to assess the speech characteristics associated with velopharyngeal function in 3-year old children with cleft palate?					1			1	2		2
How easy was it to form judgements using **Method 1**?							1	2	1	1	1

How easy was it to form judgements using **VPC-Rate**?							3	1	1		1
Describe the level of your previous experience using VAS to rate children's speech.	3	1							1		
Describe the level of your previous experience using ordinal scales to rate children's speech.		1				1					4
Describe the amount of time it took you to listen to and analyse **Sample A**.			1	2	2	1					
Describe the amount of time it took you to listen to and analyse **Sample B**.						3	2			1	
How appropriate were the parameters of speech that you analysed using the tool for the assessment of speech in 3-year old children with cleft palate?								1	3		2
How important it is to you that there is a valid and reliable framework to assess speech at age-3 years in the cleft palate population.										3	3
How likely is it you would be to use a valid and reliable assessment framework to measure speech outcomes at age three-years in the cleft palate population?										3	3

Method 1 referred to the separate analysis of resonance and nasal airflow errors associated with velopharyngeal function. The Listener SLTs reported that the established method of using separate parameters most closely matched their current assessment methods within their cleft team. Whilst both methods were judged to be acceptable by the Listener SLTs, their responses indicate that they found it easier to form judgements using the separate measures of resonance and nasal airflow than VPC-Rate. The Listener SLTs were asked to indicate their preference, with an equal split of 33.3% between separate parameters, VPC-Rate, and using both.

The questions also asked the Listener SLTs to compare their experiences and preferences for the use of both ordinal scales and VAS. The majority of the Listener SLTs ranked themselves as being very experienced using ordinal scales with a score of '10', compared with majority score of '0' for VAS. Similarly, only 16.7% (n=1) said they preferred to use VAS, compared to 50% (n=3) of the Listener SLTs who reported they preferred using ordinal scales, 33.3% (n=2) said they preferred to use both scales.

The Listener SLTs highly rated the appropriateness of the parameters of speech which they analysed in the study using higher scores (7, 8, 10) on the Likert scale. In addition, Table 5.12 shows the parameters of speech which the Listener SLTs indicated that they use in their routine clinical assessments of 3-year olds with CP±L. This list of parameters included those used in the Adapted CAPS-A, and also additional parameters either used by the Study SLT or identified in Phase 1 of the study.

All the cleft speech parameters included scored highly, with 83.3%-100% of the Listener SLTs also using these parameters in their clinical practice. Phonology, however, was only reported to be used by 66.6% of Listener SLTs in their clinical practice. In addition to the parameters of speech used in this study, three additional parameters were included in the list: voice, intelligibility, and phonetic inventory. These parameters were also identified in the scoping review and were all scored highly by the Listener SLTs (83.3%-100%).

Table 5.12 Parameters of speech used by the Listener SLTs in their routine clinical assessment of 3-year olds with CP±L.

Parameter of speech	N	%
Hypernasality	6	100
Hyponasality	5	83.3
Nasal Emission	5	83.3
Nasal Turbulence	5	83.3
Overall judgement of nasal airflow characteristics	6	100
Overall judgement of velopharyngeal function for speech	5	83.3
Cleft Speech Characteristics	6	100
Phonology	4	66.6
Voice	5	83.3
Intelligibility	6	100
Phonetic Inventory	5	83.3
Other	0	0

Table 5.13 Responses of Listener SLTs about the usability and acceptability of the Adapted CAPS-A.

Descriptor	N	%
Complicated	0	0
Unclear	0	0
Too lengthy	0	0
Clear	6	100
Concise	2	33.3
Easy to use	5	83.3
Usable in clinical practice	6	100
Usable in audit	4	66.6
Not usable in clinical practice	0	0
Not usable in audit	0	0
Age appropriate to use in the analysis of 3-year old's speech	5	83.3
Not age appropriate to use in the analysis of 3-year old's speech	0	0
Too simple for use in the analysis of 3-year old's speech	0	0
Too complex for use in the analysis of 3-year old's speech	0	0

Table 5.13 summarises the Listener SLT's responses relating to the usability and acceptability of the Adapted CAPS-A. The Listener SLT's responses are very positive, 100% of the responses indicate that this was clear and usable in clinical practice. It also scored highly with regards to being appropriate to use with 3-year olds (83.3%) and as usable in clinical audit (66.6%). The responses also highlight the importance with which the Listener SLTs view a valid and reliable framework to assess speech outcomes at age 3-years, and the likelihood they would use such a framework with all the responses either '9' or '10'.

5.6.2 Free-text responses

In addition to the multiple choice and Likert scale questions, optional free text questions allowed the Listener SLTs to expand upon their answers. These free text responses were not extensive and totalled 1586 words which were organised around emerging topics (Appendix BB). The following four topics emerged from the Listener SLT's responses: familiarity- which referred to assessment materials, methods of assessment and rating scales; practicality- which related to clinical assessments with 3-year olds and completing audit/outcome assessments; listening ease- which related to quality and length of the speech samples, and the clarity of the scales used; advantages of assessment- which related to assessments at 3-years-old, and the next steps for clinical audit.

5.6.2.i Familiarity

With regards to speech samples, a preference for Sample B was explained because, the Listener SLT was "used to listening to the GOS.SP.ASS sentences for CAPS-A listening". Another Listener SLT commented, "I am more used [sic] to listening to sentences for CAPS-A 5-year audits". Similarly, another SLT commented that, "I don't generally use a spontaneous speech sample". With reference to two approaches used to analyse the parameters of speech related to velopharyngeal function, the Listener SLTs again referred to their familiarity with one approach over another, with one listener explaining that measuring resonance and NAE separately is "how I am used to doing it." Similarity to established methods was also mentioned four times by the Listener SLTs when they discussed how separate judgements of resonance and NAE matched their clinical practice. Comments relating to VPC-Rate indicated that this was less familiar, "[I] Think it would become easier with more familiarity", "In my own head, I think I'm still using Method 1 (separate analysis of hypernasality, NAE)." Some comments reflected that VPC-Rate is different from the established methods used by the

Listener SLTs, “I do make an overall judgement whether I think there is VPI or a risk of this. I don't tend to 'rate' this”, “If possible, we try to obtain some additional information about articulation and cleft patterns and more detail in relation to resonance and NAE”.

Familiarity also featured in the Listener SLT’s explanations of why they preferred ordinal over VAS scales, “I think I am so used to using nominal scales, this feels more familiar. I have no experience of using VAS”, “My preference definitely relates to my experience and therefore confidence is using the scales.” One comment directly referenced the need for more experience with VAS, “I would need more experience to feel confident about VAS and more consensus practice”.

5.6.2.ii Practicality

The Listener SLT preferences often linked to the practical and pragmatic use of the two speech samples and assessment methods. Time was frequently referenced, particularly in relation to a preference for Sample B, with one Listener SLT commenting that they preferred to analyse the sentences because “it was quicker”, and another commented “it [Sample B] seemed to be quicker for clinical purposes and you can also listen to resonance too.”

Practical comments also related to the engagement of the 3-year olds with the assessment materials, and completion of the two types of speech samples. One Listener SLT commented on the picture stimuli used, “Pictures for both seemed equally engaging; perhaps lower demand from single word naming, but I didn't notice huge difference”.

Comments also referenced the practicality of using the speech samples in clinical practice, “[I] wonder if single words were easier for some children with reduced language skills? However, it felt like there were a lot of pictures for the single naming task, so it may be difficult to hold their attention”. Language skills were referenced again as a factor which would influence the SLT’s decision as to which speech sample to use, “If [the] child has delayed language skills then the single word picture naming would be more appropriate”, “[selection of the speech sample] varies depending on the language skills of the child”.

The variability of 3-year old children (“the speech of 3-year olds can be very inconsistent”) was also referenced by the Listener SLTs as a factor that would influence not only the speech samples they used, but also the parameters of speech which could be assessed. When discussing their preferred method of assessment, whether separately

analysing resonance and NAEs, or using an overall measure like VPC-Rate, one Listener SLT commented, “Sometimes the method we use depends on the child, if it is not appropriate, we use a method similar to method 2 [VPC-Rate]”. Another discussed that they would aim to assess these parameters separately “providing the sample was good enough”.

The Listener SLTs had differing views on the abilities of 3-year olds; one commented “many 3-year olds can sit and do quite a detailed speech assessment which allows us to make future decisions regarding therapy or further investigations of palate function”, whilst another commented that, “few children can complete the GOS.SP.ASS at 3.” One Listener SLT concluded that, “this study will reveal just how co-operative our 3-year olds are and how they can sit and do a detailed speech assessment”. Another Listener SLT discussed the need to consider what information they needed to pass onto community SLTs and how this practically impacted upon their assessments, “in order to pass on information to local SLTs, it is good to have more detail so you can discuss therapy ideas with them going forward.”

5.6.2.iii Listening Ease

Comments relating to listener ease also related to speech samples. One Listener SLT commented that they could get a “greater sense of resonance from [the] phrase level sample”. Another comment reflects the adjustment made to analyse more than one phoneme in sentence repetition, “I found it slightly easier but mainly because it was quicker which made it less tiring. It took a while to get used to which sounds I was listening for but after a few children this eased”. The length and complexity of the spontaneous speech sample was also mentioned, “the spontaneous speech sample can be frustrating/time consuming to assess”, “The length of the spontaneous speech made Sample A harder to focus on”.

The methods and scale descriptors used were also factors the Listener SLTs referenced, with “clear descriptions and parameters” for the measures of resonance and NAE. Some of the Listener SLTs discussed their preference for VPC-Rate because of ease of use, as separate measures “could be too specific for rating a 3-year old and not necessary to be so detailed”, and because there is “too much to rate with limited speech samples and inconsistency is tricky too”. However, another Listener SLT commented that they thought VPC-Rate was challenging to use in the absence of pressure consonants, “It can be difficult to make decisions about VP competency if the child is not using any oral pressure consonants or if they are quite inconsistent in productions which makes this scale more difficult to use”.

5.6.2.iv Advantages of the assessments/samples used

The Listener SLTs commented on what they believed to be the perceived advantages of assessments at age-three-years, "3 [years] is a clinically useful time point - allowing time for SLT intervention ahead of school start", "It would be invaluable to have a valid and reliable tool for 3-year olds to help with management decisions and comparisons across 3-year olds". In particular, the Listener SLTs referenced how the assessments at age-3 years "can inform management earlier on", "it would inform therapy plans and recommendations and help make decisions about whether diagnostic therapy is needed or further investigations".

In addition, to the advantages of 3-year assessments, the Listener SLTs also described the advantages of a framework for assessment to allow for clinical audit/outcome measures at this age, "It would be interesting to see comparisons between their speech at 3 and 5 years", "we have looked at 5 year outcomes for quite a long time now and it would be interesting to look at the speech of children with cleft palate at a younger age". "It would be beneficial to formalise an assessment framework and have a valid and reliable tool for assessing 3-year olds to allow for comparisons across centres and to look at trends over time".

5.7 Results Summary

In summary, the results show that most children at age-3 -years can complete detailed speech samples in a time frame which is suitable for this age group and stage of development. It took participants less time to complete Sample B than Sample A, and the Listener SLTs viewed this as an advantage. The results from the Control Group indicated that the assessment has good specificity, in that these children, without CP±L, were not found to have speech characteristics associated with cleft palate.

In the CP±L group, the results indicate that the speech samples used in this study are sufficient to reliably assess both resonance and NAE in 3-year old children with CP±L. In addition, VPC-Rate can be reliably used with either speech sample as an overall measure of velopharyngeal function, however the Listener SLTs were not in consensus as to how this measure should be used at age 3-years. Listener agreement was high for CSCs except for dentalisation/inter-dentalisation. The results for CSCs highlight the importance of considering both ICC and percentage agreement, as ICC scores were negatively impacted by limited variability in the data and for some CSCs a 'green' outcome was used frequently. The fully completed samples of Sample A had the highest levels of agreement for CSCs. When ordinal

ratings were compared to VAS scores, VAS scores showed lower levels of listener agreement for Sample A. The Listener SLTs reported a lack of familiarity with VAS.

Feedback from the Listener SLTs in the study is positive; listener preferences related to their familiarity with a speech sample or assessment method, practical/pragmatic implications, and the overall ease of the analyses. The Listener SLTs strongly supported the need for an assessment framework use with 3-year olds with reference to both clinical assessments, and audit/outcome measures.

Chapter 6. Discussion

The present study proposes a valid and reliable speech assessment framework to measure speech outcomes at age-3 years in the CP±L population (Section 6.6). The assessment framework can be used with either of the speech samples developed in the study. This was achieved through the development of two speech samples, specifically designed for 3-year olds, and a detailed examination of the specificity, sensitivity, and reliability of these speech samples and the Adapted CAPS-A outcome measure, with reference to listener reliability for resonance, NAE, CSCs, and phonology and consideration of different assessment methods and rating scales.

Routine care was not disrupted to conduct the study, however, the overall demographics of participants in the CP±L group indicate that the group was representative of children born with CP±L in 2015 and 2016 in the UK, supporting a broader generalisation of the findings of this study to the wider cleft population. The study cohort was representative both in terms of the frequency of cleft types (CP occurred most frequently in 50% of participants, compared to 43.8% of all 2015 births and 39.4% of all 2016 births; followed by UCLP which occurred in 35% of participants, compared to 22.2% of all 2015 births and 18.6% of all 2016 births; BCLP was the rarest form of cleft in 15% of participants, compared to 8.7% of all 2015 births and 9% of all 2016 births) and also in the higher number of males: females (65:35 in this study, 57:43 in 2016 data, and 57:43 in 2015 [CRANE 2016, 2017]). There was also a representative spread of ages in this study, with an age range between 3 years 0 months and 3 years 10 months, and an average age of 3 years 4 months, indicating that the results can be generalised more widely to 3-year olds with CP±L. There was also a spread of ages in the control group, from 3 years 1 month to 3 years 7 months. These demographics suggest that the findings of the present study are applicable to the intended population.

6.1 Exploring completion rates of the two speech samples by participants in the CP±L group and those in the Control Group

The current study is the first to specifically compare completion rates of two different speech samples by 3-year old children with CP±L. The results indicate that Speech Sample A (connected speech and single word naming) is an easier speech sample for 3-year olds with

CP±L to complete, given that more of the participants attempted this speech sample, even if they could not fully complete it, 95% compared to 70% for Sample B (sentence repetition). Based on completion rates it is recommended that Sample A is used as the primary speech sample at age 3-years, as more participants attempted this speech sample. None of the participants who partly completed Sample A could complete Sample B which also indicates that Sample A is an easier sample for 3-year old children. The Listener SLTs also reported that Sample A was an easier task and more suitable for children with delayed language skills.

A key finding from the study is that Listener SLTs were reliable in their judgements for Sample A for the parameters of hypernasality, hyponasality, NAE and VPC-Rate even when analysing partially completed speech samples. Reliability results were comparable for all the participants who partially or fully completed Sample A, compared to those who only fully completed this sample. For CSC outcomes full completion of the speech sample resulted in higher listener reliability. This is not an unexpected finding as a more complete speech sample would provide Listener SLTs with more information on which to base their judgements. Despite this, good levels of reliability were still evident when data from those participants who could not fully complete the speech sample was included, indicating that CSCs can still be reliably assessed even when the speech sample is not completed in full. This is a highly relevant finding at age-3, when children present with variable attention, listening and language levels and many cannot complete a speech sample in full. Indeed, these results suggest that children only part completing Sample A need not be routinely excluded in the reporting of speech outcomes at age-3 years.

In cleft care internationally, speech assessments and the reporting of outcome measures in non-syndromic cohorts at age 5-years have dominated practices (Butterworth et al. 2022; Sell et al. 2017; Chapman et al. 2016; Sell et al. 2015; Britton et al. 2014; Lohmander et al. 2009; Lohmander et al. 2005). One of the reasons why the time point of age-5-years may have been selected over earlier ages is the assumption that 5-years of age is the earliest age that most children can cooperate with audio-visual recordings using a standard speech sample. This study demonstrated that this is not the case, and that when speech samples and assessment materials were designed specifically for this age group, 70% of non-syndromic 3-year olds with CP±L and 100% of 3-year olds in the Control Group could engage and complete

detailed speech samples. This adds weight to the argument that the timepoint at which speech outcomes are first measured should be brought forward from age-5 to age-3 years.

Although more participants in the CP±L group attempted the sample of spontaneous speech and single word naming compared to short sentence repetition, 70% (n=14) fully completed both speech samples in the same assessment session. This is advantageous in clinical assessment, particularly when children present with disordered speech production as Bates & Titterington (2018) have recommended that speech is assessed across different samples and contexts. Thus, the ability to complete a detailed assessment in a single assessment session is beneficial not only for cleft teams in terms of managing waiting lists and the demand for appointments, but also because it may reduce the need for patients and their families to make multiple visits to the cleft team.

The high levels of engagement and completion of the speech samples in this study almost certainly reflected the specific design of the speech samples for 3-year olds. Detailed attention was given to word selection on both a phonetic and lexical basis to facilitate independent naming and every effort was made to balance the number of words needed to comprehensively assess speech in English, with an assessment of appropriate length for 3-year olds. In the development of Sample B, the established GOS.SP.ASS (Sell et al. 1999) sentences were shortened and alliteration was removed when possible to make the sentences more suitable for 3-year olds. Whilst the Klintö et al. (2011) study (which reported on listener reliability when analysing different speech samples at age-5 years) did not specifically report on completion rates, they reported that the 'drop out rate' was lower for single word naming and sentence repetition compared to narrative re-telling and spontaneous speech samples. In the current study, the spontaneous speech sample was most frequently attempted. To aid completion of the spontaneous speech sample in the present study, the 3-year olds had a selection of toys, books and games to talk about which were selected to be age appropriate, stimulating and appealing for children their age. This may have facilitated their engagement and subsequent completion of the speech sample. Input from the PPI group was essential in the selection of these toys, books and games and may have facilitated participant engagement and completion of the spontaneous speech sample.

In the Control Group 100% of participants completed all of both speech samples compared to only 70% in the CP±L group. Although children with complex medical needs (including those who were non-verbal), were excluded from the study, of the six participants in the CP±L group who either did not complete or partially completed the speech samples, three (50%) presented with delayed language skills, and one (16.7%) presented with delayed attention skills which impacted upon their ability to complete the speech samples. No hearing data was collected in the study and it is possible, given the association between CP±L and conductive hearing loss in the pre-school years (Fitzpatrick et al. 2021), that hearing difficulties may also have impacted the ability of the CP±L group to fully complete the speech samples. In addition, it took participants in the CP±L group almost three times as long to complete the spontaneous speech and single word naming sample than it did the short sentence repetition sample. In the CP±L group, more participants who partially completed Sample A were randomised to complete this sample first (4/5, 80%) which meant that they had already been working for longer when they started Sample B (the remaining participant was randomised to complete Sample B first but was unable to complete this). Fatigue could have been a significant factor for these four participants, impacting their ability to engage with the subsequent speech sample.

When examining completion rates in the Control Group it was also important to consider the possible impact of bias introduced both in relation to the sample size and recruitment of the Control Group and the subsequent impact this had on completion rates. Firstly, the sample size was much smaller in the Control Group. Had there been an equal number of Control Group participants it is possible that some of them may not have fully completed both speech samples. Caution therefore needs to be applied when extrapolating completion rates in this study to the wider non-cleft paediatric population, given the significant developments in relation to both language and attention skills which take place between age-3 and 4-years (Sharma & Cockerill 2014). A second source of bias specifically relates to selection bias, i.e. those parents of participants in the Control Group opted into the study because they thought their child would be able to complete the tasks in the study as set out in the participant information leaflet.

In summary, the results of the current study indicate that the 3-year olds with CP±L were able to complete speech samples specifically designed for this age group, and that both

the speech samples and the accompanying images and toys selected to elicit the speech samples were suitable for 3-year olds. Sample A is recommended for use with children with delayed language or attention skills as this was an easier task and encouraged participant engagement. The results of the current study indicate that partially completed speech samples should not be automatically excluded in the reporting of outcome measures at age-3 years, although consideration needs to be given that it may not be possible to judge extremely limited speech samples, and thus the outcome on the Adapted CAPS-A 'unable to score' is particularly relevant at this age.

6.2 Examining the validity of the speech samples and the Adapted CAPS-A

Evidence from Wren (2013) indicates that the validity of both the speech samples and assessments typically used by SLTs to assess speech at age-3 years in the UK is questionable given that many of the speech samples and assessments have not been designed specifically for this age group nor for the cleft population. It was therefore essential to develop speech samples to assess the speech of 3-year-olds with CP±L in sufficient detail; to develop an assessment framework with high content validity, assessing parameters of speech appropriate for both children aged-3 years and also with CP±L; and to ensure construct validity by developing an assessment framework with high specificity and sensitivity. In the current study the prospective nature of the data collection meant that no case selection was used and thus the speech samples produced by participants with CP±L are ecologically valid and reflect clinical data.

The CAPS-A and subsequent versions of this framework have been shown to be specific, sensitive, and valid for the cleft population when tested on older children (Ogata et al. 2022; Bruneel et al. 2020; Chapman et al. 2016; Sell et al. 2009; John et al. 2006). The results of this study show that the modifications which have led to the Adapted CAPS-A are sensitive and applicable for use with 3-year olds. For participants with CP±L, the Listener SLTs judged the cleft specific parameters of speech (hypernasality, NAE and CSCs) to be present in both speech samples. In addition, VPC-Rate (a measure not included in the original CAPS-A but which was included based on the scoping review [Fitzpatrick et al. 2020]), is a sensitive outcome measure at age-3 years. As would be anticipated, the full range of scores for VPC-Rate were used in the analysis of the speech in the CP±L group. The use of VPC-Rate at age-3

years is also validated by the adoption of VPC-Rate as the primary outcome measure for palatal function at age-3 years in the recent ToPS Trial (Persson et al. 2022), and because it is recommended internationally by ICHOM for use in the cleft palate population in older age groups (Allori et al. 2017b).

The parameters of speech included in the study are validated by the inclusion of these parameters in existing frameworks (albeit for older children) (SVANTE [Lohmander et al. 2017b; Lohmander et al. 2005]; CAPS-A [John et al. 2006]; CAPS-A-AM [Chapman et al. 2016]; ICHOM [Allori et al. 2017b]; CAPS-A Belgian Dutch [Bruneel et al. 2020], CAPS-A-Japan [Ogata et al. 2022]), research protocols (Scandcleft Trial: Lohmander et al. 2017a; Willadsen et al. 2017, ToPS Trial: Shaw et al. 2019) and by the high congruence of these parameters with those used by the Study SLT in the participant's clinical assessments and those reported to be used by the Listener SLTs in their clinical practice.

The scoping review also highlighted the importance of not only using a linguistic approach to assessments at age-3 years, but also to assess speech from a developmental perspective (Fitzpatrick et al. 2020). As such an assessment of phonological processes was included in the Adapted CAPS-A. However, phonology, was only reported to be used by 66% of the Listener SLTs in their clinical assessments. This is in contrast to the relative frequency with which either phonology or age-adjusted measures such as PCC-A were reported in the scoping review. In the current study, there were high levels of percentage agreement relating to the presence of phonological processes in both speech samples for both the CP±L and Control Groups, which supports the inclusion of this measure in the Adapted CAPS-A.

The Listener SLTs also reported assessing intelligibility, voice and recording a phonetic inventory in their clinical assessments, none of which were included in this framework. Speech intelligibility is described as the match between what a listener perceives, and what the speaker intended to say (Gnanavel & Pushpavathi 2012). Sell & Pereira (2015) report that intelligibility is difficult to use as a speech outcome measure because it is hard to define, is impacted by a number of factors including speaker characteristics which are not linked to the cleft palate (e.g. accent, rate of speech, voice quality), as well as the nature of the listener (e.g. familiarity with the speaker, trained listeners), and the type of speech sample (single words versus sentences) (Sell & Pereira 2015: 704). Challenges defining intelligibility are exemplified by the intelligibility outcome scale in the CAPS-A (John et al. 2006) which

collapsed intelligibility and acceptability into a single measure. However, Whitehill (2002) argued that intelligibility and acceptability are separate entities and can be impacted differently by hypernasality; as such the intelligibility scale is no longer used in outcome reporting at age 5-years in the UK, despite evidence that the measure is reliable (Sell et al. 2015). Consequently, this scale was not included in the Adapted CAPS-A nor is included in the final assessment framework proposed in this thesis (see Section 6.6).

Derivations of the CAPS-A have further adapted the original intelligibility scale. In the Japanese version of the CAPS-A, intelligibility and acceptability are measured separately (Ogata et al. 2022), and Bruneel et al. (2020) opted to assess understandability and acceptability as two separate measures. However, both studies continue to measure intelligibility from an impairment basis which does not consider evidence from McLeod et al. (2012: 649) in relation to the International Classification of Functioning, Disability and Health: Children and Youth Version (ICF-CY; WHO 2007). McLeod et al. (2012) report that intelligibility is influenced by both production factors (body functions) and contextual factors (environmental factors).

The challenge of who should judge intelligibility is well exemplified in the scoping review: Hodge & Gotzke (2007) utilised graduate students in speech & language pathology, Safaiean et al. (2017) used an SLT and two non-professional listeners and Dayashankara et al. (2011) used a surgeon, audiologist and SLT, illustrating the lack of agreement in this area. ICHOM (Allori et al. 2017b) proposed an alternative to an impairment-based measure of intelligibility and have advocated the use of a patient/parent reported outcome measure, The ICS (McLeod et al. 2012). The ICS is a valid and reliable parent or carer reported measure (developed for the non-cleft population) in which parents/carers report their child's intelligibility with different communication partners, validated and tested in 14 languages (McLeod et al. 2020). In the ongoing UK longitudinal cohort study, the Cleft Collective Speech & Language Study (Wren et al. 2018), the ICS has been administered at age-3 years and ICS norms for 3-year olds with CP±L have recently been published (Seifert et al. 2021). This research constitutes a significant step forward in our understanding of intelligibility outcomes at age-3 years, and the development of a valid outcome measure for intelligibility. As such it is recommended that the ICS is included as an additional speech outcome measure at age-3

years and has been included in the final speech outcome assessment framework recommended in the present study (Section 6.6).

Although voice was frequently reported to be used in clinical assessments by the Listener SLTs in the current study, and by cleft centres in the Wren (2013) survey, it was not included in the Adapted CAPS-A. Voice was infrequently reported when compared to other parameters in studies in the scoping review and no clear method was recommended as how to assess voice. Voice is judged very simply in the CAPS-A using a binary scale (the absence/presence of distinctive or abnormal voice quality) and was included to provide more information as to when listeners used 'unable to score' for resonance and NAE rather than to assess this parameter in sufficient detail for it to be considered as an outcome measure (John et al. 2006). As such there are no UK outcome standards relating to voice outcomes (Britton et al. 2014). Based on the scoping review, current standards for outcome reporting at age 5-years in the UK, and the international recommendations made by ICHOM (Allori et al. 2017b), voice was not included in the Adapted CAPS-A as a core outcome measure at age-3 years.

To ensure that conclusions made regarding listener reliability were valid it was essential to minimise factors other than the speech samples i.e. speech and picture naming (Sample A) versus short sentence repetition (Sample B), that could have impacted the reliability of listener judgements. To achieve this, the two types of speech samples were matched for the type of phonemes and the frequency that these phonemes were assessed. This reduced the impact of a sample containing later emerging, and thus more challenging, phonemes, or an imbalance in the frequency of sampling a particular phoneme between the samples. This was essential to the validity of recommendations made in the study.

Examining the specificity of the speech samples and the Adapted CAPS-A, participants in the Control Group were not identified by the Listener SLTs as having speech characteristics associated with cleft palate. This indicates that both the speech samples, and the Adapted CAPS-A have good specificity and high construct validity. An important finding to consider is that, in contrast to those speech characteristics associated with cleft palate, the Listener SLTs, frequently judged phonological processes as present in the speech of the Control Group (88.6% for Sample A and 97.0% for Sample B). This is an expected finding at age-3 years when typically developing children present with developmental phonological processes (McLeod 2009). However, it especially important in the current study because it demonstrates that the

Listener SLTs were able to distinguish between cleft specific and developmental speech immaturities, and only the latter were identified in the Control Group.

In the Control Group (when dentalisation/inter-dental realisations are excluded), normal speech outcomes for CSCs were evident in 385/385 of ratings (100%) for Sample A, and 371/374 of ratings (99.2%) for Sample B. This shows that the assessment protocol had good specificity for CSCs when either speech sample was used. Examining the Listener SLT judgements for dentalisation/inter-dentalisation, two Control Group participants were consistently rated by all Listener SLTs as having evidence of this speech process. Given that dentalisation/inter-dental realisations can occur at age 3-years developmentally (Sell et al. 1999; Smit 1993), this is not a surprising result and is why for this parameter all scalar points resulted in a green outcome on the traffic light scale, as it would not have been appropriate to judge dentalisation/inter-dental realisations as a moderate or severe speech impairment at this age. Klintö et al. (2015a) also reported a low frequency of CSCs in their control group at age-5 years. The current study provides evidence that the Adapted CAPS-A has high specificity for CSCs in a younger age group.

Other key parameters of speech relating to specificity were those assessing palatal function for speech i.e. hypernasality, NAE, Passive CSCs and VPC-Rate. The Adapted CAPS-A had good specificity for all of these parameters. Normal outcomes were seen for all judgements of NAE, Passive CSCs and VPC-Rate in the Control Group. This compares well to the norm scores reported for 3-year olds on the SVANTE (Lohmander et al. 2017), in which 94% were rated as having normal VP function for speech. Hypernasality was very frequently judged to be absent in the Control Group, 97.1% for hypernasality for Sample A, and 100% for Sample B. In the SVANTE normative data, mild hypernasality was also detected at age 3-years in children without cleft palate, but again this was very infrequent (Lohmander et al. 2017b). These results show that for parameters of speech associated with palatal function the Adapted CAPS-A had a high level of specificity using either speech sample, in that those children without CP±L were not identified as having these features in their speech.

There was a higher number of participants in the Control Group in this study than has been used in other studies validating the CAPS-A. It was of particular importance in this study to use a control group for specificity testing because the children in this study were younger than those previously used in other CAPS-A related studies. In the current study, ten speech

samples were completed by the Control Group, equating to a total of 630 ratings of cleft speech parameters for Sample A and 612 ratings for Sample B. Whilst Bruneel et al. (2020) also used ten speech samples from a control group in their validation of the Belgian Dutch CAPS-A, in the original CAPS-A study (John et al. 2006) only one speech sample from a control group was used and, in the CAPS-A Americleft (Chapman et al. 2016), no samples from a control group were used.

In addition to the specificity and the construct validity demonstrated by the results of the Control Group, the results of the CP±L group demonstrate sensitivity and high content and construct validity, relating to the construction of the speech samples, the parameters of speech included in the Adapted CAPS-A, and the presence of these cleft specific speech characteristics only in the speech of 3-year olds with CP±L. This supports the use of these speech samples and the Adapted CAPS-A with 3-year old children with CP±L.

6.3 Investigating the reliability of listener judgements using the speech samples and the Adapted CAPS-A

Kent (1996) reported that perceptual speech assessments are susceptible to error and bias. Sell & Pereira (2015) reported that assessing cleft speech presents significant challenges in relation to listener reliability because of the complex speech profiles associated with CP±L. The 2004 review by Lohmander & Olsson indicated that the reliability of listener judgements was not routinely examined or reported on in cleft palate literature. Almost two decades later, listener reliability is reported much more frequently, with recent studies by Bruneel et al. (2022), Willadsen et al. (2022), Morrison et al. (2021), Cleland et al. (2021) being examples. However, the scoping review indicated that reliability outcomes have not been universally reported at age-3 years, due to the use of retrospective case analysis (e.g. Rezaei et al. 2022; Peryer et al. 2021; El Ezzi et al. 2015; Hamming et al. 2009; Hattee et al. 2001). No other studies have examined the impact of different speech samples on the reliability of listener judgments at age-3 years, therefore, the current study makes a significant contribution to our understanding of the reliability of speech outcomes at age-3 years.

As in other studies which have developed cleft assessment protocols based on the CAPS-A (Ogata et al. 2022; Bruneel et al. 2020; Chapman et al. 2016; Sell et al. 2009; John et al. 2006) the analysis of listener reliability was a crucial component of present study and was

essential to determine if speech outcomes could be reliably assessed at age 3-years. However, as Chapman et al. (2016) previously described, it is challenging to compare reliability results across studies because different statistical approaches have been used. In studies using CAPS-A a range of statistical measures have been used including, ICC, weighted and unweighted kappas and percentage agreement (Ogata et al. 2022; Bruneel et al. 2020; Baillie & Sell 2020; Chapman et al. 2016; Sell et al. 2015; Sell et al. 2009; John et al. 2006). However, Chapman et al. (2016) pointed out that kappa scores can be interchangeable with single measure ICCs which were used in the current study. Studies not using the CAPS-A, such as the Scandcleft Trial (Lohamnder et al. 2017a; Willadsen et al. 2017) have often measured reliability as a point-by-point transcription agreement, capturing agreement for articulation as a whole, rather than for specific CSCs. This makes comparisons of reliability with the current study and other studies using derivations of the CAPS-A extremely challenging.

In addition, the number of listeners used to measure reliability varies considerably between studies that assessed 3-year olds, and also studies that reported speech outcomes using the CAPS-A (i.e Baillie & Sell 2020). This is important to consider because in studies using a higher number of listeners there is potential for greater inter-rater variation which may impact reliability ratings. A strength of the present study is that seven Listener SLTs were used, which is comparable to other studies validating the CAPS-A. The number of Listener SLTs used in the current study is also more representative of the number of SLTs working in larger cleft teams in the UK and thus the levels of inter-rater reliability are likely to reflect listener variation in clinical practice. In comparison, studies which have reported the reliability of listener judgements when assessing 3-year olds have used a much smaller number of listeners for example, two listeners were used in the studies by Chacon et al. (2017), Raud Westberg et al. (2017), Klinto et al. (2016), and four in the Willadsen et al. (2018) study. It is arguable that the higher number of listeners in the present study resulted in increased inter-rater variation, and whilst this may more accurately reflect listening practices within a cleft team, direct comparisons of reliability with studies of fewer listener is somewhat flawed.

It is also very difficult to directly compare the results of this study to those reporting on older age groups. Older children are less likely to have complex speech profiles, because developmental immaturities should have resolved (McLeod 2009) and children should have accessed therapy intervention. It is therefore arguable that the speech of older children is less

complex for listeners to analyse, which would result in more reliable listener judgements. This is exemplified by the longitudinal study by Persson et al. (2006) which showed that articulation outcomes improved with age in the CP±L population. This study included children from 3-10 years of age, but did not report separate reliability outcomes for the 3-year olds therefore making comparisons with reliability outcomes in the current study extremely difficult.

Whilst listener reliability in the current study is compared with the other studies using the CAPS-A, there are no published studies using the CAPS-A which assess children as young as those in the current study (Ogata et al. 2022; Bruneel et al. 2020; Baillie & Sell 2020; Chapman et al. 2016; Sell et al. 2015; Sell et al. 2009; John et al. 2006), so comparisons of reliability need to be interpreted with this understanding. Despite these challenges, the reliability results obtained here are discussed in the context of other studies using the CAPS-A, to specifically examine whether reliability using the Adapted CAPS-A at age-3 years is comparable to other studies using the CAPS-A. Comparisons with the existing literature have also been made to determine if reliability results in the current study are similar to other studies either investigating speech outcomes in 3-year olds, using VPC-Rate, or comparing VAS and ordinal scales. A summary of the reliability results of the studies used to make such comparisons is provided in Appendix CC and Appendix DD. However, caution was still applied when making direct comparisons given the number of factors previously outlined which could have influenced reliability scores.

6.3.1 Statistical Considerations

Inter-rater reliability was calculated using ICC, Krippendorf's alpha and percentage agreement. Intra-rater reliability was calculated using paired *t*-tests and percentage agreement. Percentage agreement was presented alongside the ICC due to a lack of variability in the range of severity within the data as not all scalar points were used. This occurred not just for specific speech videos recordings, but for individual speech parameters. For example, there were several parameters with a high prevalence of 'normal' outcomes, including hyponasality, double articulation, pharyngeal articulation, and double articulation with a glottal. The lack of variability resulted in low ICC scores, but high percentage agreement, a similar finding to other studies using the CAPS-A or tools derived from it (Ogata et al. 2022; Baillie & Sell. 2020; Bruneel et al. 2020; Chapman et al. 2016).

A potential solution would have been to specifically select cases for analysis to reflect the full spectrum of speech outcomes as per John et al. (2006), Sell et al. (2009), and Chapman et al. (2016), however due to the prospective collection of speech samples in this study it was not possible to pre-select the speech samples to provide a full spectrum of outcomes for CSCs, resonance, and nasal airflow. Interestingly, even when Bruneel et al. (2020) did case select, this still did not guarantee the use of every scalar point for every speech parameter because some parameters of cleft speech do not occur as frequently as others (CRANE 2019) e.g. pharyngeal articulation and double articulation with a glottal. Indeed, both of these CSCs also occurred infrequently in the present study. In the current study the best solution to this statistical artifact was to present ICC, Krippendorf alpha and paired *t*-test scores alongside percentage agreement in order to most accurately reflect listener agreement, and this study recommends that this approach is used in future studies using the CAPS-A.

6.3.2 Comparisons of Listener Reliability for Sample A and Sample B

6.3.2.i Resonance and NAE

The reliability results obtained in the current study indicate that either speech sample could be used to assess speech at age-3 years, as both samples resulted in good levels of listener inter and intra reliability for the majority of speech parameters assessed. For hypernasality, hyponasality, NAE and VPC-Rate inter-rater percentage agreement was consistent for both speech samples with percentage agreement >61% for hypernasality and VPC-Rate which indicates good levels of agreement. Indeed, for hyponasality and NAE percentage agreement was >81% indicating very high levels of agreement. Intra-rater agreement was also strong for both speech samples for hypernasality, and percentage agreement was good for both hyponasality and NAE. Whilst the Listener SLTs had not been trained to assess nasal emission and nasal turbulence in a single measure their results showed good levels of reliability. The comparison of inter-rater reliability results for hypernasality, hyponasality and NAE when Sample A is partially or fully completed, with those when Sample A is fully completed, and Sample B, indicates that partially completed speech samples can be included in the analysis of hypernasality, hyponasality and NAE without negatively impacting reliability scores.

The inter-rater reliability scores in this study for hypernasality are particularly promising because inter-rater reliability scores for hypernasality have been suboptimal in the

cleft population (Brunnegard & Lohmander 2007) including studies of 3-year olds (Lohmander & Persson 2008; Persson et al. 2006). For hypernasality Sample A outperformed Sample B on calculations of inter-rater reliability, with good reliability for Sample A (ICC= 0.790) compared to moderate reliability for Sample B (ICC= 0.489). On the other hand, percentage agreement was similar for both samples (73.7% for Sample A and 78.6% for Sample B).

Since hypernasality occurs across word boundaries, it could be hypothesised that sentence/phrase level repetition (Sample B) would be the more reliable sample on which to assess hypernasality. However, this was not found in the present study. There are two reasons why such good reliability was reported for hypernasality for Sample A. Firstly, hypernasality was rated across the entire speech sample i.e. both spontaneous speech and single-word naming, in order to increase the length of the sample analysed and facilitate the assessment of hypernasality across word boundaries. Secondly, Sample A was constructed to include high vowels to facilitate judgements of hypernasality (CLISPI 2017; Lohmander et al. 2009), which is likely to have facilitated judgements of borderline-mild hypernasality (John et al. 2006; Sell et al. 1999). Indeed, the combined use of spontaneous speech and single word naming to analyse hypernasality over single words containing high vowels alone is supported by evidence from Lohmander & Persson (2008) as percentage agreement was only 39% when only single words with high vowels were used. In comparison, in their assessment of 3-year olds, Raud Westberg et al. (2017) reported high listener agreement (80%) when analysing hypernasality using single words only. This finding is surprising given evidence from the Scandcleft Trial which reported suboptimal levels of reliability when hypernasality was analysed using a nine-word string formed by editing together recordings of single word naming (Lohmander et al. 2017a; Willadsen et al. 2017). In the current study, the combination of spontaneous speech and single word naming for judgements of hypernasality may have resulted in stronger inter-rater reliability in this study for Sample A compared to that reported by Lohmander & Persson (2008) for single words, and that reported in the Scandcleft trial for a nine-word string (Lohmander et al. 2017a; Willadsen et al. 2017), because longer, multiple samples of connected speech were used. As such it is recommended that if single words are used in the analysis of hypernasality, that this is used in conjunction with a sample of connected speech, or alternatively sentence repetition could be used.

Taking into consideration both ICC and percentage agreement, inter and intra-reliability results for hypernasality, hyponasality and NAE are also comparable to scores reported in the other studies developing the CAPS-A as an outcome tool with older age groups (Ogata et al. 2022; Bruneel et al. 2020; Chapman et al. 2016; John et al. 2006). Baillie and Sell (2020) and Sell et al. (2015), also using the CAPS-A with older age groups, both reported higher inter-rater percentage agreement for hypernasality than other studies using the CAPS-A, potentially because fewer listeners were used.

In the current study nasal emission and nasal turbulence were combined into a single measure with very high levels of percentage agreement. Whilst Baillie and Sell (2020) and Sell et al. (2015) reported high levels of listener agreement for separate measures of nasal emission and nasal turbulence, overall reliability outcomes for nasal emission and nasal turbulence are inconsistent across studies validating the CAPS-A, even when the two measures were combined (Ogata et al. 2022). The results of the current study support the combination of nasal emission and nasal turbulence into one outcome measure at age-3 years, this is also supported by other studies reporting on 3-year outcomes with similar levels of listener inter and intra-rater agreement using a combined measure of NAE (Larson et al. 2017; Raud Westberg et al. 2017; Lohmander & Persson et al. 2008; Persson et al. 2006). Whilst these studies, all originating from Sweden use an established protocol to combine nasal emission and nasal turbulence, the Listener SLTs in the current study were not trained to do so, but were still very reliable in their judgements, further supporting the use of a combined measure of NAE at age-3 years.

The Listener SLTs completed a practice listening session and had the opportunity to receive feedback on their ratings of hypernasality, hyponasality and NAE. This opportunity for calibration, although limited, may have also supported the levels of reliability observed in the study as listener training has shown to be effective in improving reliability for these parameters of speech (Persson et al. 2022; Oliveira et al. 2016; Chapman et a. 2016). Reliable judgements in the present study may be because the scales for hypernasality, hyponasality and NAE were unchanged from the routinely used scales in the CAPS-A, with the exception that nasal emission and turbulence were combined. Previous experience and training using the CAPS-A seems to have supported good reliability ratings for both speech samples.

In conclusion, both Speech Sample A and Sample B result in good levels of listener reliability for hypernasality, hyponasality and NAE, indicating that these parameters can be as reliably assessed using a derivation of the CAPS-A at age-3 years as they can be for older children, provided this assessment is completed by SLTs who have had training in the CAPS-A. For hypernasality reliability was higher in the current study then other studies reporting outcomes at 3-years (Lohmander & Persson 2008; Persson et al. 2006), which also recommends the future use of either Sample A or Sample B, in combination with the Adapted CAPS-A, to assess hypernasality at age 3-years.

6.3.2.ii Cleft Speech Characteristics

Both speech samples result in reliable judgments of CSCs at age 3-years indicating that either sample could be used to assess speech at this age. Whilst ICC scores were impacted by a lack of variability in the data for both Sample A and B, majority agreement ranged from good- very good for almost all CSCs (the exception being dentalisation/inter-dentalisation discussed below). Whilst Sample A had a higher number of CSCs with very high majority agreement (>81%), Sample B had seven CSCs with higher percentages for overall majority agreement. However, when matched recordings are considered i.e. those participants who fully completed Sample A and Sample B, the highest levels of percentage agreement are for Sample A. This is an important finding as it highlights that the highest levels of inter-rater agreement for CSCs were for the fully completed samples of Sample A.

Whilst ICC scores for intra-rater reliability were frequently impacted by a lack of variability in the use of the CAPS-A scale, percentage agreement was good-very good; for Sample A intra-rater agreement ranged from 73.5-100%, and for Sample B from 77.1-100%. Overall, intra-rater reliability as judged by percentage agreement is slightly higher for Sample B. This could be explained by the Listener SLT's familiarity with the analysis of sentences, which are frequently used with the GOS.SP.ASS clinical assessment (Sell et al. 1999) and routinely used with the CAPS-A (John et al. 2006). All the Listener SLTs had been specifically trained to analyse sentences in CAPS-A training (Sell et al. 2009), and thus may have developed consistent internal standards specifically relating to the analysis of sentence based speech samples.

Direct comparisons for either inter and intra-rater reliability for specific CSCs or even groups of CSCs are often not possible in the wider literature when point-by-point transcription

is used. In the Scandcleft Trial (Lohmander et al. 2017a, Willadsen et al. 2017) (assessing 5-year-olds), inter-rater point by point transcription agreement was between 83.9%-88.9%. In this study, inter-rater majority agreement across the CSCs varied between 63.2-100% for Sample A, and 64.3-100% for Sample B (when dentalisation/inter-dentalisation was excluded). It is of note that the Scandcleft Trial not only used point by point agreement, and assessed older children, but also that the trial was completed over many years, with intensive, repeated week-long listener training activities, which may account for the high levels of percentage agreement for articulation.

Examining those studies using a 3-year old cohort, Chacon et al. (2017) reported much higher rates of inter-rater reliability than was reported in this study, at 92.9%. However, the ranges of inter-rater percentage agreement reported by Larsson et al. (2017), Raud Westberg et al. (2017) and in studies by Klintö et al. (2016, 2015, 2014a, 2014b) are all comparable to the inter-rater majority agreement reported in the current study for CSCs. Thus, whilst comparisons are challenging, the reliability results for CSCs in the current study seem to be broadly in line with reliability reported for articulation in other studies of 3-year olds, indicating that the assessment of CSCs in this study is as reliable as previous studies of 3-year olds.

In the current study, listener reliability is reported for each CSC separately in order to provide more detail relating to listener reliability at age-3 years, and to understand the reliability of the separate CSCs. This contrasts with other studies using the CAPS-A that have reported reliability for each CSC group e.g. Anterior Oral CSCs, Posterior Oral CSCs, Non Oral CSCs and Passive CSCs. One benefit of the approach in the present study is that this specifically highlighted suboptimal levels of inter-rater reliability for dentalisation/inter-dentalisation. The ICC was fair for Sample A, poor for Sample B and majority agreement was markedly lower than for other CSCs, 32.2% for Sample A and 21.4% for Sample B. Suboptimal reliability for dentalisation/inter-dentalisation highlights one of the challenges of assessments with 3-year olds who have a developing speech sound system (McLeod 2009). Despite removing the label of dentalisation/inter-dentalisation as a CSC in the current study, it is highly likely that the Listener SLTs were inconsistent in how they classified errors of dentalisation/inter-dentalisation, either as a developmental immaturity or as a CSC, with a subsequent negative impact on reliability.

The challenges of classifying dentalisation/inter-dentalisation were considered in the design of the Adapted CAPS-A. Whilst the original CAPS-A descriptors were maintained (0 consonants affected; 1 or 2 consonants affected; 3 or more consonants affected), all of these outcomes were classified as 'normal' i.e. dark green outcomes on the traffic light scale (original outcomes were dark and light green) and this explanation was provided to the Listener SLTs. It is possible, that changing the established traffic light colours for this CSC led to some confusion, potentially biasing the listeners into judging this as a developmental speech immaturity rather than a CSC. Another possible reason for such inconsistencies may be because dentalisation/inter-dentalisation can result in an allophonic change (a variant in pronunciation) rather than a phonemic change (substitution of one phoneme for another which may change meaning) (Chapman et al. 2016). As such, because there was not a total phonemic change, this subtle change to the target phoneme may have been more difficult to perceive, leading to the inconsistent judgements by the Listener SLTs. Despite low reliability for this parameter, it has been retained in the final assessment framework proposed in this thesis (see Section 6.6). Whilst the current study indicates that further listener training is needed to establish more reliable judgements of dentalisation/inter-dentalisation, the ability to compare scores for dentalisation/inter-dentalisation at age-3 and 5-years would provide insights as to whether this articulation error should be considered developmental or cleft specific at age-3 years. If dentalisation/inter-dentalisation identified at age 3-years continues to persist at age 5-years it could be argued that this should be considered as a CSC as this has not resolved in the expected manner (Bowen 2015).

Other studies validating the CAPS-A have reported suboptimal inter-rater reliability for Anterior Oral CSCs (Bruneel et al. 2020; Chapman et al. 2016; Sell et al. 2009). In comparison, Baillie and Sell (2020), also using the CAPS-A, reported very high percentage agreement for both palatal/palatalisation and lateral/lateralisation. Such high reliability may be because Baillie and Sell (2020) only reported on these parameters and did not include an assessment of dentalisation/inter-dentalisation. It is possible that poor reliability for dentalisation/inter-dentalisation influenced the overall inter-rater reliability for Anterior Oral CSCs reported by Bruneel et al. (2020), Chapman et al. (2016) and Sell et al. (2009). Interestingly, in these studies (Bruneel et al. 2020; Chapman et al. 2016; Sell et al. 2009) intra-rater reliability for Anterior Oral CSCs was higher than inter-rater reliability, indicating that

although agreement between the listeners was suboptimal, the listeners did apply consistent internal standards (Keuning et al. 1999). This was also evident in the current study. Whilst intra-rater reliability for dentalisation/inter-dentalisation was lower than other CSCs it was much higher than that reported for inter-rater reliability (Sample A intra-rater reliability was 73.5% compared to 32.2% for inter-rater reliability, Sample B intra-rater reliability was 77.1% compared to 21.4% for inter-rater reliability). For intra-rater reliability the Listener SLTs internal standards were more consistent than those between the listeners; this suggests there is a future training need to reach consensus about judgements of dental/interdentalisation at age-3 years.

Although the argument for interpreting palatalisation/palatal as a developmental speech immaturity is less strong than that for dentalisation/inter-dentalisation, ICC scores were only fair for Sample A and were poor for Sample B, with majority percentage agreement amongst the lowest of all CSCs. It is possible that some Listener SLTs judged palatalisation/palatal as a developmental speech immaturity given evidence that palatalisation can still be resolving at age-3 years (James 2001).

Examining Posterior Oral CSCs, reliability results indicate that these CSCs can be reliably analysed using either Sample A or Sample B at age-3 years. Inter-rater reliability compares well to other CAPS-A validation studies, particularly as inter-rater reliability scores for this category of CSC were inconsistent in both the John et al. (2006) and Chapman et al. (2016) studies. Chapman et al. (2016) attributed such inconsistency to the misidentification of backing as palatal stops which were subsequently recorded as an Anterior Oral CSC. Listener inconsistency relating to the perception of palatal stops has also been previously reported by Santelmann et al. (1999). However, unlike the current study, Chapman et al. (2016) did not specifically report on separate CSCs, so it is not possible to substantiate their conclusion. It is, however, interesting that Persson et al. (2006) combined palatalisation/palatal articulation and backed to velar/uvular and reported this as ‘retracted oral articulation’. Sell et al. (2005) described the challenge of comparing CSC outcomes when differing definitions are used. Point by point transcription agreement for retracted oral articulation in the Persson et al. (2006) study was higher than that reported in the current study for backed to velar/uvular, and thus resulted in superior reliability in the Persson et al. (2006) study. However, had this approach been taken in the design of Adapted CAPS-A it

would have precluded comparisons with other studies using the CAPS-A, and may not have guaranteed superior reliability given evidence in the current study that the Listener SLTs may not have consistently judged palatalisation/palatal as a CSC at age-3 years. In addition, in English, palatal errors are often allophonic (Chapman et al. 2016; Santelmann et al. 1999) in comparison to backing to velar which results in phonemic change.

Taking into account both ICC scores and majority agreement, reliability scores for Non Oral CSCs and Passive CSCs were classified as good-very good for both speech samples in the current study. Reliability scores also compare well to studies using the CAPS-A. Evidence from the other CAPS-A studies indicates that Non Oral CSCs are rated more consistently than other classes of CSC (Bruneel et al. 2020; Chapman et al. 2016; Sell et al. 2015; Sell et al. 2009) and the current study indicated that this is also the case at age-3 years.

Overall, reliability ratings for CSCs were high in the current study and compare well both with other studies utilising or derived from the CAPS-A. This is important for two reasons, firstly it demonstrates that the Adapted CAPS-A at age-3 years resulted in as reliable ratings for CSCs as when used with older age groups, and secondly that the adjusted listening procedure for Sample B (i.e. that Listeners listened to the sentence first for one phoneme, and then again for another, a notable difference to the CAPS-A) did not impact upon listener reliability. The current study supports the separate reporting of reliability for each CSC rather than reporting reliability for whole classes of CSC. The approach used in the current study allowed for the identification of specific CSCs for which listener agreement was suboptimal and for which further listener calibration and training is required i.e. dentalisation/inter-dentalisation.

Using an adapted version of the CAPS-A, an outcome tool which the Listener SLTs in this study had been previously trained to use, appears to have supported good levels of reliability. In addition, the inclusion of a practice session within the listening methodology, in which each listener received specific feedback relating to their analysis of CSCs compared to other Listener SLTs may also have favourably supported inter-rater reliability. This type of listener calibration encouraged listeners to reflect on their scores and could have impacted their internal standards leading to more consistent ratings between listeners. Reliability results reported in this study support the use of either Sample A or Sample B to assess CSCs

at age-3 years, however, when only fully completed speech samples are considered Sample A has the highest levels of listener reliability.

6.3.2.iii Phonology

In the current study the Listener SLTs first judged whether phonological processes were present or absent, and then determined the type of phonological processes if present: age appropriate, delayed or disordered (these categories were not mutually exclusive). There were very high levels of agreement for the presence of phonological processes, 100% for Sample A and 92.9% for Sample B. This highlights the relevance of including phonology measures at age 3-years i.e. for Sample A there was 100% agreement that all the 3-year old participants with CP±L presented with phonological processes.

However, inter-rater percentage agreement for the categories of age-appropriate phonological process, delayed phonology and disordered phonology was highly variable for both speech samples, and it is not possible to conclude which speech sample was more reliable for these measures. It is possible that where a participant presented with significantly disordered phonology, and inconsistently used an age-appropriate phonological process, that some listeners rated this by selecting disordered phonology only, and others both disordered and age-appropriate phonology. It is also possible that some of the Listener SLTs felt that disordered phonology was captured in the judgement of CSCs and thus may not have selected this option for phonology. Ultimately such inconsistencies and poor reliability for the measure of phonology means that, at present, neither speech sample can be reliably used to measure phonology using the Adapted CAPSA.

There are three main factors which could have impacted upon reliability ratings for phonology. Firstly, although efforts were made to control for the application of different developmental norms by the Listener SLTs, by asking them to use the norms provided by Dodd et al (2003), the Listener SLTs probably still applied norms they had learned during their SLT training and which were subsequently ingrained through their clinical practice. Secondly, the three categories of age-appropriate phonological process, delayed phonology and disordered phonology were not mutually exclusive and as such the Listener SLTs could have combined the categories in differing ways resulting in differing judgements. Finally, phonology ratings were noted to vary in the practice listening. Despite feedback to encourage more consistent ratings with examples of how the categories could be applied, percentage agreement

remained variable. Furthermore, only 66% of the Listener SLTs reported that they assessed phonology at age-3 years, suggesting a lack of listener experience assessing phonology. As such future training and research is recommended to ensure that listeners apply the same standards when analysing phonology and to establish whether the Adapted CAPS-A can be used to measure phonology using the method described here.

Three-year olds are at a unique and complex stage of speech development (McLeod & Baker 2017: 202) which in the cleft population is further complicated by the presence of hypernasality, NAE and CSCs. Making judgements of phonology in complex speech profiles is a challenge at age-3 years in the CP±L population (Chapman & Willadsen 2011). Indeed, reliability scores for ratings of 'the presence of non-cleft speech immaturities' (which is the closest equivalent to phonology categories outcomes in the current study) in older age groups had suboptimal reliability (Ogata et al. 2022; Chapman et al. 2016; Sell et al. 2009). It could be hypothesised that judgments of developmental speech immaturities in older children would be more reliable as these developmental processes would occur more infrequently. In the CAPSA-AM study, poor reliability of non-cleft speech immaturities led to the abandonment of this measure in further reliability testing (Chapman et al. 2016). At age 3-years, it is contraindicated to exclude this rating, particularly when evidence from the scoping review highlights the importance of taking a developmental perspective when measuring speech outcomes at age-3, and because of the high frequency with which phonological processes were judged to be present in the current study.

In the present study, it is of note that intra-rater agreement for age-appropriate phonological processes, delayed phonological processes, and disordered processes was considerably higher than that for inter-rater agreement, in keeping with the studies by Ogata et al. (2022), Chapman et al. (2016), Sell et al. (2015), Sell et al. (2009) and John et al. (2006). This adds weight to the argument that listeners apply their own internal standards relating to phonology/non-cleft immaturities irrelevant of the age of the participant. As such, calibration and training are required before measures of phonology can be reliably assessed by different listeners at age-3 years. Teams using the CAPS-A with older age groups should also consider further calibration and training relating to measures of non-cleft speech immaturities.

Certainly, for a measure of phonology to be included in any future assessment protocol inter-rater reliability scores in relation to both Sample A and Sample B need to

improve, or an alternative method of assessment may be indicated. The first step could be to identify if there was a process issue in terms of how Listener SLTs recorded the phonological processes, by providing them with a sample transcription and asking them to record this on the Adapted CASP-A. This could potentially identify if the same phonological processes were being recorded in different ways, and any future listening instructions could be clarified with further exemplars. Any disagreements as to the presence or type of phonological processes, may indicate that further training specifically relating to phonology is required to support listener calibration, or alternative methods of assessment are required. A potential alternative could be to simplify the classification used to record the presence of typical phonological processes, atypical phonological processes, or both, as described by Howard, Heselwood & Harding-Bell (2019), or to complete a separate analysis of phonology using an established assessment tool such as the Phonological Screening Assessment (PSA) (Stevens & Isles 2011) or the Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al. 2002).

6.4 Examining the impact of different rating methods and scales on the reliability of listener judgements

6.4.1 Rating Methods

The current study introduced an additional scale VPC-Rate (assessing overall VP function) to the Adapted CAPS-A. At the study outset it was not possible to predict whether there would be good reliability for either speech sample for the parameters of hypernasality and NAE, thus VPC-Rate presented an alternative scale. Whilst the current study showed that resonance and NAE can both be reliably assessed at age-3 years, it has also highlighted the potential use of VPC-Rate with this age group. Majority inter-rater agreement for VPC-Rate was good for both speech samples, with the highest agreement evident when Sample A was fully completed. Whilst inter-rater percentage agreement was similar for all samples of Sample A and Sample B, ICC scores were very good for Sample A compared to moderate for Sample B. Intra-rater reliability was strong for both speech samples.

For Sample A, stronger reliability for VPC-Rate is reported than was previously reported by Lohmander et al. (2017c) in the development of the VPC-Rate scale. Lohmander et al. (2017c) reported moderate reliability for two listeners of the same language, similar to that of Sample B. Lohmander et al. (2017c) solely based the assessment of VPC-Rate on a

connected speech sample. It may be the length of Sample A, that facilitated reliable judgements between the listeners, because the listeners had a longer sample to analyse. Indeed, Sample A was almost three times longer than Sample B and is also longer than the connected speech sample used by Lohmander et al. (2017a).

More recently, Persson et al. (2022) reported on reliability of VPC-Rate in the ToPS Trial. In the ToPS study extensive training, calibration opportunities and listener testing took place over several years to enable reliable judgments of this parameter (Shaw et al. 2019). Persson et al. (2022) reported that absolute percentage agreement between the three SLTs was 80% and for majority agreement, when 2/3 SLTs agreed, agreement was 99%, and intra-rater reliability was 94%. These findings confirm that VPC-Rate is a reliable outcome measure for use with 3-year olds with CP±L, and especially shows the impact of specific and repeated training over several years to develop listener calibration. Despite good agreement in the current study, 68.4% majority agreement for Sample A and 64.3% for Sample B, agreement for VPC-Rate is lower than was reported by ToPS which is likely to reflect that VPC-Rate was a novel outcome measure for four of the Listener SLTs (three had been involved in the ToPS trial) and that the listener calibration which took place before the present study was extremely limited compared to the ToPS trial. However, the current study demonstrates that listeners inexperienced with the scale were still able to apply consistent internal standards despite a lack of experience, indicating that the scale has high construct validity.

Positive feedback from the Listener SLTs about the VPC-Rate outcome related to the relative ease of using this scale because separate measures "could be too specific for rating a 3-year old" and because there is "too much to rate with limited speech samples and inconsistency is tricky too." However, there was no consensus that VPC-Rate should replace established measures, suggesting that VPC-Rate should be used as an additional measure, as one Listener SLT described '[VPC-Rate] is the impression we want to achieve by the end of our assessment, whether the child's VP function is competent or not."

VPC-Rate fits very well if the objective of outcome measures at age-3 years is to identify those children with indicators of a poor speech outcome, in this case specifically linked to VP function. This is evidenced by the frequency with which overall measures of VP function were included at age-3 years in the scoping review (Fitzpatrick et al. 2020). The VPC-Rate scale could also aid clinical decision making, identifying those children who have a good

outcome following their primary palatal repair, those whose VP function requires monitoring, and those who are likely to require an objective assessment of their VP function and potentially secondary surgical intervention. VPC-Rate could be added into tools for clinical assessment e.g. GOS.SP.ASS (Sell et al. 1999).

6.4.2 Rating Scales

Whilst the CAPS-A and derivations of this (Ogata et al. 2022; Bruneel et al. 2020; Chapman et al. 2016; John et al. 2006) have been validated and tested for reliability using ordinal scales, the emerging evidence that parameters of speech associated with velopharyngeal function may be more validly measured using alternative scales could not be ignored (Yamashita et al. 2018; Bayliss et al. 2015). The current study is therefore the first to compare the impact of different speech samples on the reliability of listener ratings using different rating scales at age-3 years and is the first to investigate the reliability of VPC-Rate using a visual analogue scale (VAS).

Both VPC-Rate and hypernasality were analysed using both ordinal and VAS scales. For VPC-Rate, the ICC scores for both Sample A and Sample B decreased when the VAS was used. This was most notable for Sample A; the ICC decreased from 0.841, indicating very good inter-rater reliability on the ordinal scale, to 0.586 indicating moderate reliability on the VAS. For Sample B, there was no notable change to the ICC, which was 0.498 for the ordinal scale and 0.476 for the VAS, both of which are described as moderate reliability. However, when specifically examining only those recordings analysed using both an ordinal scale and VAS for Sample B, reliability was moderate for VAS compared to poor on the ordinal scale. It is recommended that an ordinal scale is used to analyse VPC-Rate when a speech sample comprised of spontaneous speech and single word picture naming is used. For phrase/sentence based speech samples, VAS may confer an advantage but this is unclear.

The reasons why Sample A had superior reliability for VPC-Rate on an ordinal scale rather than VAS should be explored, particularly as VPC-Rate was a new outcome to four of the Listener SLTs. One possible reason is that on the ordinal scale there were three clear descriptors to guide the Listener SLTs relating to competent, marginally incompetent and incompetent judgements, compared to the VAS in which only two of these descriptors were used at each end of the scale: competent and incompetent. Possibly it is those marginally

incompetent cases which the Listener SLTs were unclear where to position on the continuous VAS when analysing Sample A.

For hypernasality, Sample A showed good levels of inter-rater reliability on the ordinal scale compared to moderate levels of agreement for Sample B. On the VAS, ICC scores decreased for both Sample A and Sample B, this was most notable for Sample A, and the ICC decreased to 0.586 indicating moderate reliability. In comparison, Bettens et al. (2018) and Castick et al. (2017) reported comparable levels of reliability for hypernasality using ordinal scales and VAS, and Bayliss et al. (2015) reported superior reliability for hypernasality when using VAS when compared to ordinal scales.

One possible explanation as to why reliability for VAS was lower for Sample A in this study relates to the speech sample. For Sample A Listener SLTs were asked to rate hypernasality across the entire speech sample, two minutes of spontaneous speech and single-word naming. To date, no other studies have compared ordinal and VAS scales using this speech sample. In comparison, the speech samples used by Castick et al. (2017) and Bayliss et al. (2015) were similar to that of Sample B in this study, sentence repetition. Despite this, reliability was poorer for both VAS and ordinal scales for hypernasality in the current study, indicating that factors other than the speech sample also need to be considered to understand this difference.

Apart from the speech sample, several other factors could have impacted reliability ratings for the VAS. Unlike the listeners in the Baylis et al. (2015) study who, on average, indicated that they were comfortable using a VAS scale, the Listener SLTs in this study reported that they were very inexperienced in the use of VAS scales. In the practice listening for this study, the Listener SLTs used the VAS to rate both hypernasality and VPC-Rate and feedback was provided to support listener calibration. However, unlike Bettens et al. (2018), a specific training session on the two scales was not undertaken. This study used seven listeners, which is more than was used by Castick et al. (2017), Baylis et al. (2015) (five listeners) and Bettens et al. (2018) (four listeners). It is possible that more listeners in this study increased the level of listener variation and impacted the reliability ratings.

Another influencing factor is that unlike Castick et al. (2017) and Bettens et al. (2018), the speech samples analysed using VAS were randomly selected and were not specifically

selected to demonstrate a full spectrum of speech outcomes. This could have led to an imbalance in the use of the full scale (0-100) and could have impacted reliability scores in a similar way to ordinal ratings, in which a statistical artifact was observed. The decision not to specifically select recordings was made for practical and pragmatic reasons, firstly to avoid analysis of the same speech sample on multiple occasions (which may have led to listener recall). Secondly, this study was designed using prospective data collection and it was not possible to profile or case select the recordings in advance. Another factor to consider is the relatively small number of speech samples that were analysed using VAS in this study and thus it is not possible to predict whether the same trends would have been evident in a larger dataset. Based on this study, there is no evidence at age-3 years to specifically recommend using VAS over established ordinal scales, however, further research is required with a larger sample size to substantiate this.

6.5 Exploring the views of the Listener SLTs

Completion of the previous objectives has demonstrated that the speech samples developed in the study are suitable for 3-year olds, specific to the CP±L population, and have resulted in reliable listener judgements. However, it was important to consider the views of SLTs who had completed the listening in the study and who would use the proposed assessment framework. Thus, a questionnaire was designed to capture feedback from the Listener SLTs and their feedback was taken into account in the final assessment framework proposed in this thesis (Section 6.6). This is in contrast to other studies developing speech outcome tools. Whilst John et al. (2006) similarly considered listener acceptability in the design of the CAPS-A, others studies developing versions of the CAPS-A did not report on listener acceptability of the tools (Bruneel et al. 2020; Chapman et al. 2016). Acceptability of the proposed assessment tool was similarly not reported in the development of the SVANTE (Lohmander et al. 2017b) nor in the design of the Universal Parameters (Henningsson et al. 2008).

The feedback from the Listener SLTs validated the aim of this research, and they overwhelmingly supported the need for a valid and reliable assessment framework at age-3 years to measure speech outcomes. Listener SLTs welcomed an assessment framework to measure speech outcomes before 5 -years, with comments that this would highlight the need of children with CP±L to access therapeutic intervention before they start primary school. The

variation and lack of standardisation in clinical practice for assessments at age 3-years was highlighted by the Listener SLTs.

Of particular significance is feedback from the SLTs regarding the speech samples. The majority of Listener SLTs reported finding Sample B easier to analyse, associated with the shorter duration of this speech sample, thus being 'quicker' to analyse. It is interesting that whilst the SLTs reported Sample B as easier to analyse this was not reflected in inter-rater agreement, although intra-rater agreement for CSCs was slightly higher for Sample B. The majority of SLTs indicated that they would prefer to use both speech samples in the future, indicating that they felt the 3-year old participants engaged equally well with both speech samples. As such the combined use of both speech samples is considered in the final proposed framework.

Feedback from the Listener SLTs also confirmed the parameters of speech and the rating scales included in the final assessment framework. For example, 83.3% of the Listener SLTs reported that they would complete a phonetic inventory at age 3-years. The reliability of a phonetic inventory has not been investigated in the current study, and thus it cannot be recommended at age-3 years as an outcome measure. However, to aid clinical assessments, the assessment proforma has been adapted so that when the SLTs complete their transcription they can also record the phonetic inventory if they wish to do so. However, whilst the following speech sounds are assessed in both Sample A and Sample B /p, b, t, d, f, v, s, z, ʃ, tʃ, dʒ, k, g/, neither sample assesses all consonant sounds or vowels in English. Thus, SLTs may need to use additional resources or stimulability exercises in their clinical assessments in order to sample and record a comprehensive phonetic inventory for their patients.

The majority of the SLTs reported a preference for ordinal scales, or a combination of ordinal and VAS. As such, when reliability results were also taken into consideration, ordinal scales have been retained in the assessment framework in line with the original CAPS-A. The Listener SLTs had reported that the Adapted CAPS-A was clear, easy to use and appropriate for use to analyse the speech of 3-year olds with CP±L.

One objective in the current study was to measure the acceptability and usability of the speech assessment framework to the Listener SLTs and this has been achieved. Feedback

from the Listener SLTs shaped the final proposed framework, and represented the potential future users of the framework, this was essential if such a framework is to be adopted by the SLT cleft teams in the UK.

6.6 Final Proposed Assessment Framework to assess speech outcomes at age 3-years in the CP±L population

Whilst the aim of the present study was to propose a valid and reliable assessment framework to assess speech outcomes at age 3-years in the CP±L population, at the start of the study it was difficult to predict if this could be achieved, the SVANTE, originally developed in Swedish and now translated in Norwegian (Lohmander et al. 2017b; Lohmander et al. 2005) was and remains the only existing outcome tool which has been tested for reliability and validated to assess speech outcomes at age 3-years in the CP±L population. A wide variety of assessment tools and speech samples are currently used to assess speech at age 3-years across UK cleft teams, a reflection of the wide developmental profiles of 3-year olds. The scoping review (Fitzpatrick et al. 2020) highlighted that much fewer studies have investigated speech outcomes at age-3 years in comparison to age-5 years, probably because there may have been an assumption in cleft care internationally that 5-years of age is the earliest age that most children can cooperate and complete a standard speech sample and video recording.

However, the present study has highlighted the abilities of non-syndromic 3-year old children with CP±L. The results have shown that most 3-year olds can complete detailed speech assessments, generating speech samples which can be used both for clinical assessment and to measure speech outcomes and that such an assessment can be completed in a single appointment. Indeed, if clinically required to analyse complex or inconsistent speech profiles, the present study has demonstrated that more than one speech sample can often be completed in the same session. The current study has also highlighted the skills of SLTs specialised in cleft palate, who can reliably assess a number of speech parameters simultaneously in young children with CP±L when speech samples specifically designed for this age group are used.

The aim of the current study was the development of a speech outcome measure at age-3 years, a need specifically required in cleft care, rather than focussing on clinical

assessment. However, much consideration was given to pragmatic and practical considerations in the design of both speech samples to gain as much clinically relevant information within these speech samples to reduce the likelihood of patients needing to complete multiple additional clinical assessments, and thus have to return to the Cleft Centre for further appointments. Indeed, 70% of the participants in the CP± L group completed both speech samples in the same assessment session, and thus the proposed framework (Figures 6.1, 6.2 Tables 6.1, 6.2) recommends that for clinical assessment both Sample A, two minutes of spontaneous speech and single word picture naming, and Sample B, sentence repetition, are used as each has been shown to result in reliable listener judgements in this study. For clinical assessment it is recommended that Sample A is administered before Sample B given that more children attempted this speech sample.

The advantages of using both speech samples in clinical assessment are that speech can vary across the speech hierarchy, so that single words versus sentences or spontaneous speech may provide different information. The combined use of a spontaneous speech sample, single word naming and short sentence repetition meets many of the key requirements set out by the UK & Ireland Child Speech Disorder Research Network, Good Practice Guidelines (Bates & Titterington 2018). This states the need to assess consonant-vowel, consonant-vowel-consonant, mono and disyllabic structures, target consonants in different word positions, word levels as well as a connected speech assessment. In addition, Stoel-Gammon (2015) recommended that assessment of disordered speech in pre-schoolers should include "imitated, elicited and spontaneous contexts" of speech, as well as the assessment of spontaneous speech, single words and phrases (p.88), which would also be achieved through the combined use of both speech samples. However, as reported by Howard (2011), SLTs need to capture different aspects of speech production dependent on the individual patient's need in order to appropriately plan treatment. Thus, additional clinical assessments may be required in addition to Sample A and Sample B e.g. assessment of consonant and vowel sounds not sampled, stimulability, production of polysyllabic words, analysis of prosodic features and non-speech/oro-motor assessment (Bates & Titterington 2018). Whilst it is anticipated that the combination of both speech Sample A and Sample B will facilitate the clinical assessment of 3-year olds with CP±L, such assessment may not be wholly comprehensive and SLTs may need to use additional resources to complete a

comprehensive assessment based upon the presentation and needs of the individual patient in order to plan their treatment.

As discussed in Chapter 1, a speech outcome measure at age-3 years would facilitate the early identification of children at risk of poor speech outcomes, provide cleft surgeons with an earlier indication of the success of primary palate repairs, allow the comparison of outcomes at age-3 and age-5 years, and the reporting of longitudinal speech outcomes. The present study has demonstrated that each speech sample has good reliability for resonance, NAE and CSCs, and the combination of both Sample A and Sample B would assess speech across a wider hierarchy. However, it was not an objective of this study to assess reliability for both Sample A and Sample B in combination, not least given the average length of a combined speech sample, which would be over 23 minutes. Potentially there could be a fatigue effect for listeners analysing such lengthy speech samples, as well as a resource issue for cleft teams given the time required out of clinical practice for SLTs to analyse these speech samples. As such there is a need for speech outcome reporting to be effective and efficient, which could be achieved using either Sample A or B.

Several factors needed to be considered in the recommendation of one speech sample over another for use in clinical audit and outcome reporting at age-3 years. Firstly, both speech samples resulted in good listener reliability for resonance, NAE and CSCs. However, it was the fully completed samples of Sample A that had the highest levels of reliability for CSCs and VPC-Rate, which led to the recommendation that Sample A should be used as the primary speech sample for outcome reporting at age-3 years. Higher participant engagement and completion rates for Sample A also supported its selection, particularly because children with delayed language and attention skills still attempted this speech sample. When Sample A is partially completed it can still be reliably analysed, and as such partially completed speech samples can also be included in the reporting of speech outcomes at age-3 years. This is extremely relevant given that variable attention and language levels at age-3 may mean that more children are unable to complete speech samples in their entirety. This is supported by evidence that young children with CP±L are at higher risk of delayed attention (Khoshlab et al. 2021) and delayed language skills (Tillman et al. 2018). Sample A also contains the Restricted Word List (RWL) in English, which has all the advantages previously outlined in section 4.2.1. Feedback from the Listener SLTs relating to their perception that the 3-year olds

found Sample A easier to complete also supported the recommendation of Sample A as the primary speech sample.

However, the practical and pragmatic advantages of using Sample B should not be ignored. Sample B resulted in comparable levels of reliability for most parameters of speech assessed and it was the preferred sample of the Listener SLTs relating to both listener ease and the length of time required to analyse this speech sample. Sample B was completed considerably faster by children with CP±L than Sample A, in an average of 5 minutes 54 compared to 17 minutes 20 for Sample A. In terms of the demands placed on NHS cleft services, discussed in Chapter 1 of this thesis, the analysis of Sample B over Sample A would be more time efficient for cleft teams, which is likely to be an important consideration in the adoption of the proposed framework by UK cleft teams. It is also important to consider that sentence repetition is the recommended speech sample in the CAPS-A at 5-years.

Of interest, prior to the adoption of the CAPS-A by every cleft team in the UK, discussions took place at a national level in Lead Cleft SLT meetings with wider discussions in the national Cleft CEN. This study presents UK cleft teams with two valid and reliable speech samples for consideration, both of which can be used with the Adapted CAPS-A to report speech outcomes at age-3 years. Ultimately it will be the decision of UK cleft teams if either speech sample is adopted at a regional or national level.

The final version of the assessment framework uses the Adapted CAPS-A. This takes into consideration that the Listener SLTs reported that the Adapted CAPS-A was easy to use, clear and was appropriate to use with 3-year olds with CP±L. The Adapted CAPS-A remains characterised by ordinal scales, based not only on the superior reliability of the parameters of VPC-Rate and hypernasality using ordinal scales over VAS, but also because the Listener SLTs showed a preference for ordinal scales or a combination of ordinal scales and VAS. Use of the CAPS-A ordinal scales also supports the comparison of speech outcomes internationally, given the growing international use of the CAPS-A in the USA (Chapman et al. 2015), Belgium (Bruneel et al. 2020) and Japan (Ogata et al. 2022).

The measure VPC-Rate is included in the final version of the framework given the strong reliability observed for this measure, particularly for Sample A. The results of the present study also indicated that VPC-Rate can be reliably used even when Sample A is not

completed in full. Furthermore, six UK cleft teams participated in the ToPS trial and SLTs from each team were trained to use VPC-Rate (Shaw et al. 2019), therefore adopting this practice is likely to be well supported, especially given evidence from Persson et al. (2022) about the reliability of VPC-Rate.

As discussed in section 4.3.1, Bayliss et al. (2011) argued that nasal emission and nasal turbulence should be combined into a single measure and this is supported by the inter and intra-rater agreement scores presented in this study. In both the CAPS-A and the Adapted CAPS-A listeners count the instances of NAE, hence the use of a single measure of NAE at age 3-years would not preclude comparisons at age 5-years if nasal emission and nasal turbulence continue to be separately reported, because the same traffic light scale is used consistently across both separate and combined measures of NAE.

The modification to the traffic light scale for dentalisation/interdentalisation, outlined in section 4.3.2, is retained in the final proposed framework. A point of difference to the 5-year CAPS-A is the final recommendation that phonology outcomes should also be reported at age-3 years. Whilst listener inter-rater reliability results were suboptimal for phonology the high frequency with which phonological errors were identified at age-3 highlighted the importance of retaining a measure of phonology at age-3 years. However, further training and reliability testing is required before this outcome can be reported using the method described in the present study.

Listener familiarity and previous mandatory two-day training in the CAPS-A (Sell et al. 2009) contributed to the satisfactory levels of reliability in this study. As a minimum all SLTs using the Adapted CAPS-A should have completed CAPS-A training (www.caps-a.com). Additional training relating to the Adapted CAPS-A including the assessment of phonological processes, the different approach to the classification of dentalisation/interdentalisation and the VPC-Rate scale is recommended. Another advantage of using the Adapted CAPS-A is that the framework has also been designed in such a way that outcomes are presented using the CAPS-A traffic light system, which allows for the comparison of outcomes at age-3 and age-5 years and is highly familiar to UK cleft teams.

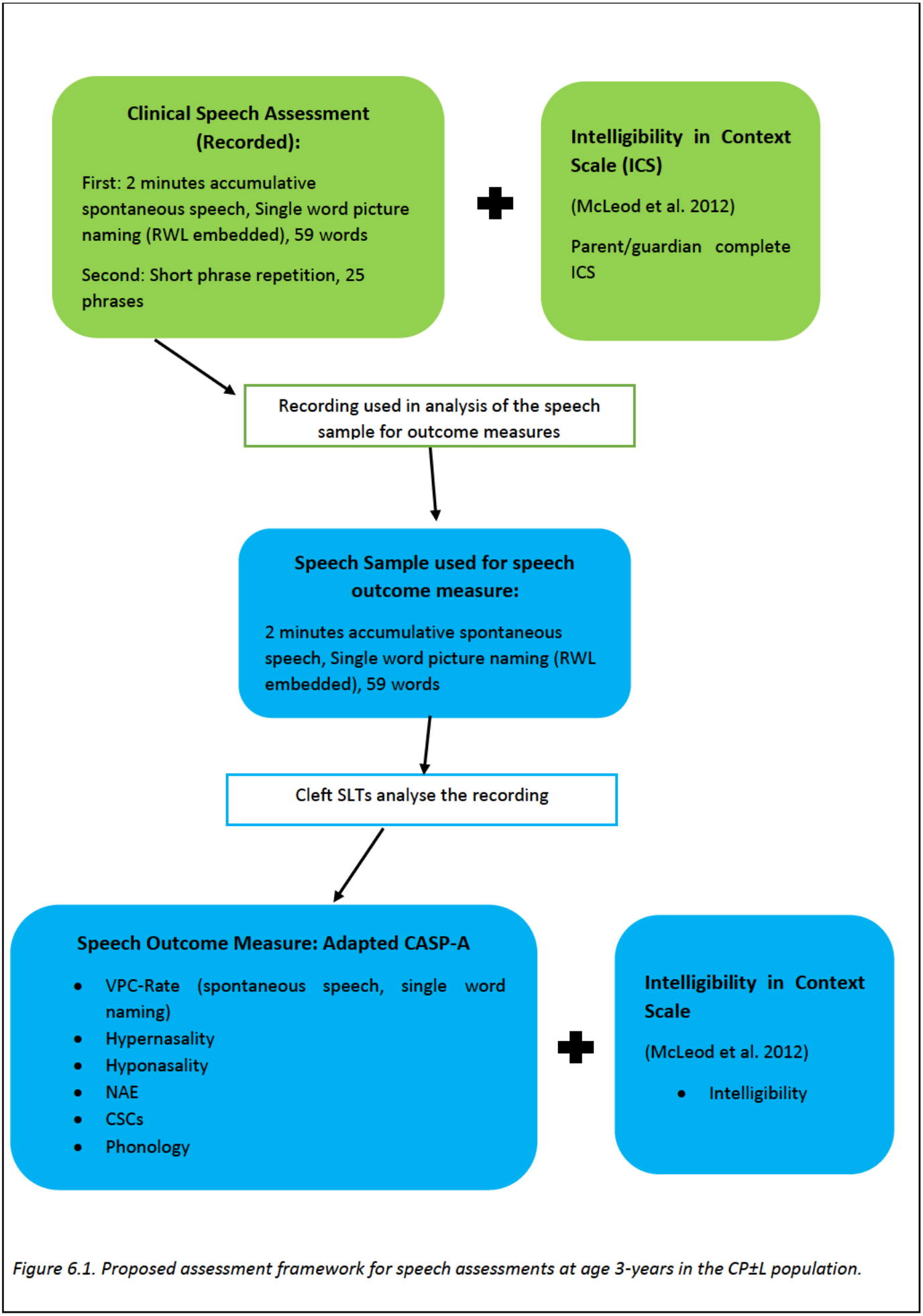


Figure 6.1. Proposed assessment framework for speech assessments at age 3-years in the CP±L population.

Table 6.1. The Combined Word List: Single word speech sample

Name:				ID Number:		
Date:			Age:		Cleft Type:	
		Vowel	Word	Structure	WI consonant	WF consonant
	Warm up		cat			
	Warm up		bike			
Restricted Word List	1	high	pooh	CV	p	
Restricted Word List	2		puppy	CVCV	p	
Restricted Word List	3	high	pea	CV	p	
Restricted Word List	4	second V high	baby	CVCV	b	
Restricted Word List	5	high	bee	CV	b	
Restricted Word List	6		ball	CVC	b	l
Restricted Word List	7		four	CV	f	
Restricted Word List	8	high	feet	CVC	f	t
Restricted Word List	9		five	CVC	f	v
Restricted Word List	10		Val	CVC	v	
Restricted Word List	11	second V high	very	CVCV	v	
Restricted Word List	12		vase	CVC	v	z
Restricted Word List	13		night	CVC	n	
Restricted Word List	14		nose	CVC	n	
Restricted Word List	15		knee	CV	n	
Restricted Word List	16	high	tea	CV	t	
Restricted Word List	17	high	two	CV	t	
Restricted Word List	18	high	teeth	CVC	t	
Restricted Word List	19		doll	CVC	d	l
Restricted Word List	20		door	CV	d	
Restricted Word List	21	second V high	daddy	CVCV	d	
Restricted Word List	22		house	CVC	h	s

Restricted Word List	23		horse	CVC	h	s
Restricted Word List	24		bus	CVC	b	s
Restricted Word List	25	high	kick	CVC	k	k
Restricted Word List	26	high	key	CV	k	
Restricted Word List	27	low	car	CV	k	
Restricted Word List	28		gate	CVC	g	t
Restricted Word List	29		girl	CVC	g	l
Restricted Word List	30	high	geese	CVC	g	
Combined Word List	31	high (accent dependent)	mum/mom	CVC	m	m
Combined Word List	32	high	shoe	CV	ʃ	
Combined Word List	33	high	juice	CVC	dʒ	
Combined Word List	34		chip		tʃ	p
Combined Word List	35	high	moo	CV	m	
Combined Word List	36	high	spoon	CCVC	sp	
Combined Word List	37	low	snail	CCVC	sn	
Combined Word List	38	high	sleep	CCVC	sl	
Combined Word List	39		mouse	CVC	m	
Combined Word List	40		web	CVC		b
Combined Word List	41	high	leaf	CVC	l	f
Combined Word List	42	high	sea	CV	s	
Combined Word List	43	low	hat	CVC	h	
Combined Word List	44		leg	CVC	l	g
Combined Word List	45	high	food	CVC	f	d
Combined Word List	46		zip	CVC	z	
Combined Word List	47	high	zoo	CV	z	
Combined Word List	48	high	sheep	CVC	ʃ	p

Combined Word List	49	high	**fish**	CVC		ʃ
Combined Word List	50	high	**chair**	CVC	tʃ	
Combined Word List	51		**shop**	CVC	ʃ	
Combined Word List	52		**beach**	CVC		tʃ
Combined Word List	53	high	**cheese**	CVC	tʃ	z
Combined Word List	54		**page**	CVC		dʒ
Combined Word List	55		**bridge**	CCVC		dʒ
Combined Word List	56		**sun**			n
Combined Word List	57		**jelly**	CVCV	dʒ	
Combined Word List	58	high	**ring**	CVC		ŋ
Combined Word List	59		**high**	CV	h	
Consonant Inventory						

Table 6.2. Sentence repetition speech sample

Name:					**ID Number:**				
Date:			**Age:**			**Cleft Type:**			
1.	**M**ary ca**me** **h**o**me**				13.	**Z**ebra li**v**es at the **z**oo			
Target	**m**	**m**	**m**		Target	**z**	**v**	**z**	
Transcription					Transcription				
2.	**P**u**pp**y has a **p**a**p**er				14.	**Fish** and **ch**ips			
Target	**p**	**p**	**p**		Target	**f**	**ʃ**	**tʃ**	
Transcription					Transcription				
3.	**B**ob i**s** a **b**a**b**y				15.	A **sn**ail **sh**ell			
Target	**b**	**z**	**b**	**b**	Target	**sn**		**ʃ**	
Transcription					Transcription				
4.	**Ph**o**ne** **f**ell o**ff**				16.	The **ch**ick is **h**at**ch**ing			
Target	**f**	**n**	**f**	**f**	Target	**tʃ**	**h**	**tʃ**	
Transcription					Transcription				
5.	**D**a**ve** dri**v**ing a **v**an				17.	**Sl**u**g** i**n** the **s**alad			
Target	**d**	**v**	**v**	**v**	Target	**sl**	**g**	**n**	**s**
Transcription					Transcription				
6.	**N**eil i**s** **t**e**n**				18.	**K**aren is **m**aking a **c**a**ke**			
Target	**n**	**z**	**t**	**n**	Target	**k**	**m**	**k**	**k**
Transcription					Transcription				
7.	I **l**i**ke** the **b**a**ll**				19.	The **ch**ildren **j**u**gg**le			
Target	**l**	**k**	**b**	**l**	Target	**tʃ**	**dʒ**	**g**	
Transcription					Transcription				
8.	**T**i**m** has a **h**a**t**				20.	**T**i**g**er in the **j**ungle			

Target	**t**	**m**	**h**	**t**
Transcription				
9.	**D**a**d** wants an oran**ge**			
Target	**d**	**d**	**dʒ**	
Transcription				
10.	**G**irl wa**sh**ing her **h**ands			
Target	**g**	**ʃ**	**h**	
Transcription				
11.	A **s**a**d** fa**ce**			
Target	**s**	**d**	**s**	
Transcription				
12.	**Sh**e is o**n** the bu**s**			
Target	**ʃ**	**n**	**s**	
Transcription				

Target	**t**	**g**	**dʒ**
Transcription			
21.	R**ng** the bell		
Target	**n**		
Transcription			
22.	A **sp**o**tt**y dog		
Target	**sp**	**t**	
Transcription			
24.	Wear your welly		
25.	Wow a yo-yo		
Additional transcriptions:			

The Adapted CAPS-A for 3-year old speech assessment				
Resonance				
Hypernasality				
Rating	Description		Unable to score	Score
0	Absent		8	
1	Borderline-minimal			
2	Mild-evident on close vowels e.g. z<u>oo</u>, thr<u>ee</u>, s<u>i</u>x / ũ, ĩ, ɪ̃/			
3	Moderate- evident on open and close vowels			
4	Severe- evident on vowels and voiced consonants			
Hyponasality				
Rating	Description		Unable to score	Score
0	Absent		8	
1	Mild- partial denasalisation of nasal consonants and adjacent vowels			
2	Marked- denasalisation of nasal consonants / m̃, ñ, ŋ̃/			
Nasal Airflow Errors				
Rating	Description		Unable to score	Score
0	Absent on pressure consonants		8	
1	Occasional: pressure consonants affected <10% of the sample			
2	Frequent: pressure consonants affected >10% of the sample (judged highly pervasive or highly distinctive)			
VPC-Rate				
Rating	Velopharyngeal Closure	Description	Unable to score	Score
0	Competent	Can include active nasal fricatives	8	
1	Marginally incompetent	Evidence of minor problems suggesting borderline closure		
2	Incompetent	Evidence of significant problems usually requiring surgical management		

Cleft Speech Characteristics Summary				
Speech Characteristics		**Absent 0**	**1 or 2 consonants affected**	**3 or more consonants affected**
	Anterior Oral Speech Characteristics			
1	Dentalisation/inter-dentalisation			
2	Lateralisation/ lateral			
3	Palatalisation/ palatal			
	Posterior Oral CSCs			
4	Double articulation			
5	Backed to velar/uvular			
	Non-oral CSCs			
6	Pharyngeal articulation			
7	Glottal articulation			
8	Active nasal fricatives			
9	Double articulation			
	Passive CSCs			
10	Weak and/or nasalised consonants			
11	Nasal realisation of plosives &/or suspected passive nasal fricative			
12	Gliding of fricatives/affricates			
	Phonology			

Phonological processes present	Age appropriate phonological processes	Delayed phonology	Disordered phonology

Figure 6.2 The Adapted CAPS-A

6.7 Study Limitations

Much deliberation was involved in the design of the speech samples, which underwent several iterations before the final versions were decided, including feedback from both the supervisory team and the PPI group. Following the study, some minor modifications were made to the final proposed speech samples (Section 6.6). In Sample B, the phrase 'Dad drinking orange juice' was used for assessment of the phoneme /dʒ/, assessed in word final position in 'orange' and word-initial position in 'juice.' However, the process of assimilation in connected speech led to the production of only one phoneme. In view of this, the sentence has been amended so that /dʒ/ is only assessed in word final position in 'orange' and has been revised to 'Dad wants an orange.' Word initial / dʒ/ is instead assessed in 'juggle' in another phrase.

This study used summary patterns to report speech outcomes for CSCs which allowed for direct comparisons with other studies validating or reporting speech outcomes using the CAPS-A. However, as point by point transcription agreement was not used in the study this precluded direct comparisons with many of the studies reporting articulation outcomes for 3-year olds as identified in the scoping review (Fitzpatrick et al. 2020). Had point by point transcription agreement been used in the current study this may also have provided useful information to confirm or refute the hypotheses made in relation to poor agreement for dental/dentalisation, palatal/palatalisation and phonological processes. For example, a potential explanation discussed in relation to all three of these areas in Section 6.3.2ii and 6.3.2.iii is that the Listener SLTs successfully identified the same errors/substitutions in their transcription, but subsequently classified these differently i.e. that some Listener SLTs recorded dental/dentalisation or palatal/palatalisation as a CSC and others as a developmental substitution, and that the Listener SLTs applied the phonology categories in different ways given that these were not mutually exclusive. Examination using point by point transcription agreement would have provided more detail to support these hypotheses than can be gained by using summary patterns alone.

A limitation which recurred for several parameters of speech is that not every scalar point of the ordinal scales was used in the Adapted CAPS-A due to the convenience sampling used, the need to ensure no additional visits to the cleft centre, and the prospective data

collection used in the study. Essentially this means that 'real world' data was used, however it was not possible to guarantee examples of a full spectrum of speech outcomes, which can also be the case even when speech samples are specifically selected. This is because of the relative infrequency with which some speech parameters occur, as observed in the Bruneel et al. (2020) study. In the current study, there were very high levels of 'green' traffic light outcomes for some parameters of speech and thus for these parameters it is only possible to conclude that the SLTs were reliable in their ability to judge the absence of a parameter of speech.

The lack of variability in the use of the Adapted CAPS-A scalar points impacted calculation of the ICC, Krippendorf's alpha and correlation, resulting in a statistical artifact. An alternative statistic would have been to use weighted kappa; however, Chapman et al. (2016) reported that this too was similarly affected by a lack of variability. This reflects a wider issue for reliability statistics used with ordinal scales. As such it is recommended that percentage agreement is reported alongside weighted kappa, Krippendorf's alpha and ICC scales to understand listener agreement.

Although an alternative scale, such as VAS may have supported the use of a wider range of inferential statistical tests, addressing some of the problems associated with reliability statistical tests and ordinal scales, this study did not find that VAS resulted in superior reliability for hypernasality and VPC-Rate, particularly for Sample A. This is the first study to report the use of VAS to measure VPC-Rate. However, the small number of speech samples analysed using the VAS limits the generalisability of the findings of the study. A solution would have been to have a larger sample size of 3-year olds. The sample size in this study equated to over 20% of all 3-year olds treated by the WMCLPS. Practically, a longer recruitment period, or a multi-site study would have been needed to facilitate a larger sample size.

A practice listening session took place to support listener calibration but only included one example of each speech sample in the practice listening session, which precluded feedback on intra-rater consistency. Although an alternative would have been to recruit multiple participants for use in practice analyses (Shaw et al. 2019), recruiting multiple participants for use in practice analyses involves a significant time investment and is likely to only be feasible in funded studies.

The current study used a process of constrained randomisation. The intention of using this method of randomisation was to achieve an equal proportion of participants randomised to Sample A and Sample B. However, when families rescheduled appointments or decided at the study appointment that they did not want their child to participate in the study this resulted in an imbalance between participants randomised to Sample A and Sample B. For example, 50% of the participants were intended for Sample A first, but in the final randomisation 60% of participants in the CP±L group were randomised to complete Sample A first. This imbalance did not impact the results of the study because, irrespective of randomisation, only children who fully completed one speech sample went on to fully complete the other. However, the limitation of using constrained randomisation in the study is acknowledged.

The Listener SLTs were not fully blinded as to whether the video recordings were of participants in the Control Group or participants in the CP±L group due to the physical evidence of a cleft lip repair for some participants. It is possible that this influenced the Listener SLT's judgments. An alternative would have been to only recruit children with isolated cleft palate, but this would have reduced the number of potential recruits and limited the generalisability of the results to the wider cleft population. The use of audio only recordings was considered, however both Sell et al. (2002) and Klintö & Lohmander (2017) reported that video recordings can result in more analytical listener judgements over audio only recordings. Ultimately the decision was made to use video with audio recordings although the potential impact of this on listener judgments is acknowledged.

As discussed in Section 6.1, there was a possible selection bias introduced in relation to the Control Group, with the potential for parents to opt their child into the study because they perceived that their child would be able to complete the speech assessments as described in the participant information leaflet. Such selection bias must be considered when making comparisons between the CP±L and Control Group in terms of completion rates for the speech samples. The impact of any bias should also be considered in terms of any conclusions made about the specificity of the assessment. Whilst there were 35 ratings of each speech parameter for Sample A, and 34 ratings for Sample B (560 total listening instances for Sample A, 544 listening instances for Sample B), and the maximum recruitment target of five participants in the Control Group was achieved, this is a relatively small number

of children. Whilst this is comparable or exceeds the number of control group participants used in other CAPS-A studies, in the studies specifically of 3-year old children by Klintö et al. (2016, 2015) and Willaden & Poulson (2011) a higher number of control group participants was used, 20 in the Klintö et al. (2016, 2015) studies and 14 in the Willaden & Poulson (2011) study. This may reflect that the unique stage of speech sound development of 3-year old children, which indeed is reflected in this study given that only 11.4% of Control Group Participants in Sample A, and 3.0% in Sample B were judged to have no evidence of phonological processes. Potentially, had a greater sample size been used in the Control Group there could have been further variation in the speech sound profiles analysed which could have influenced the results. Whilst the Klintö et al. (2016, 2015) and Willaden & Poulson (2011) studies also used a control group of children without any known speech difficulties, it is arguable that had a control group been recruited of non-cleft 3-year olds with delayed phonological development that this would have resulted in a more robust examination of specificity, examining if the Listener SLTs were able to consistently identify cleft specific patterns of articulation and phonology from delayed phonological development. The use of a control group of children with known speech delay at age-3 years should be considered in any subsequent validation studies.

6.8 Future Research

A recommendation for future research would be to include more substantial training on VAS; taking into consideration feedback from the Listener SLTs that they felt inexperienced using this approach. However, more recently the Borg c-M-scale has been proposed as an alternative to both VAS and ordinal scales (Yamashita et al. 2018). The Borg c-M scale allows ratings on a similar continuum to VAS (0-100), but listeners can also differentiate between degrees within the same category which could be useful for borderline/mild judgements. This scale appears to have more clinical relevance than VAS alone and should be considered in further research.

The current study highlighted the need for a speech outcome measure at age-3 years in the UK, and this was supported by the Listener SLTs. A dialogue across all UK cleft teams is indicated to discuss the clinical implementation of the findings of the present study. The collection of standardised speech samples at age-3 years across UK cleft teams would create

opportunities for multicentre research studies, and when used in combination with outcome measures at older ages would provide longitudinal data. The comparison of patient outcomes using the proposed framework at age 3-years and those at age 5-years using the original CAPS-A would provide information relating to the predictive validity of the Adapted CAPS-A and could be used to investigate the success of any therapy or surgical intervention between these time points.

Future training on the Adapted CAPS-A is also required, specifically relating to VPC-Rate as a new scale, and judgements of dentalisation/interdentalisation and phonology. The current scale for phonology was unreliable and needs further investigation. Such training could indicate whether any further adaptations to the scale and methods of judging phonological processes are required. The training could also be used as an opportunity to explore both the importance of including a phonetic inventory, as reported by the Listener SLTs, and the reliability of listener judgements for this measure. The present study has indicated that future training on the CAPS-A for the outcome of non-cleft developmental speech immaturities is required as there was a lack of consensus in the literature (Ogata et al. 2022; Sell et al. 2015; Chapman et al. 2016; Sell et al. 2009; John et al. 2006) as to how this parameter was analysed, indicating that further research specifically relating to training and calibration for this parameter is required.

The Covid-19 pandemic occurred within the time frame of this study. Given that cleft surgery was paused and infants had palate repairs later than usual (Arnaout et al. 2022) putting them at increased risk for poor speech outcomes (Russell et al. 2022) it has become more important than ever to reliably and validly measure speech outcomes at an earlier age than 5-years. In addition, audiological assessment and treatment was delayed nationally (Arnaout et al. 2022) and the impact of this on speech outcomes is not yet known. Waiting times to access community speech and language therapy have increased during and since the Covid-19 pandemic, and children are not able to access the same frequency or intensity of therapy from community SLT services, The Royal College of Speech and Language Therapists report that “children’s speech and language therapy services are facing significant challenges as they try to balance growing waiting lists with maintaining provision for children already on their caseload” (RCSLT 2022). Waiting lists at the WMCLPS are increasing as families want to access more input directly from the cleft centre. This has a subsequent impact on the

frequency and intensity of therapeutic intervention offered at the centre and the impact of this on speech outcomes is not currently known. The impact of the pandemic on speech outcomes will be informed by the standardised early assessment this study offers.

Chapter 7. Conclusion

This study highlights the abilities of 3-year olds with CP±L in which seventy percent of the non-syndromic 3-year olds were able to fully complete both speech samples within one assessment session, indicating that a comprehensive speech sample can be successfully gathered from the majority of 3-year olds in the nonsyndromic CP±L population. High completion rates in the study reflected the targeted design of the speech samples specifically for this age group, which considered not only the phonetic basis of the target words but also their lexical basis, the length of the phrases, the picture stimuli, and the toys used to generate the spontaneous speech sample. Using attractive assessment materials to elicit these speech samples resulted in high engagement and completion rates.

The results demonstrate that it is possible to reliably assess speech outcomes at age-3 years using the Adapted CAPS-A. Inter-rater majority percentage agreement was high for hypernsality, hyponasality, nasal airflow errors and overall agreement was also good for CSCs for each speech sample. This highlights the skills of the SLTs who were trained in the CAPS-A and were able to transfer these listening skills to the Adapted CAPS-A. Almost all speech samples were judged to have evidence of phonological processes emphasising the complexity of the speech profiles which were reliably analysed in the study.

No previous studies have specifically designed and recommended a speech sample for the measurement of speech outcomes at age-3 years, nor have compared how different speech samples may impact the reliability of listener ratings. The proposed speech sample and assessment framework constitutes a novel contribution to the reliable assessment of speech outcomes at age 3-years in the CP±L population. The proposed framework includes the measure VPC-Rate to provide an overall judgement of velopharyngeal function and combines audible nasal emission and nasal turbulence into a single measure based upon the very good levels of listener reliability reported in this study. Whilst the study has found evidence for the use of a sample of spontaneous speech and single word naming to assess speech outcomes at age-3 years, a national discussion across UK cleft teams is required to discuss how the findings in the study are implemented.

The need for this study has been only exacerbated by the Covid-19 pandemic with wider implications for speech outcomes relating to delayed palate surgeries and longer waiting lists to access speech therapy intervention in the NHS. It is intended that this framework will play a key part in ensuring that speech outcome measures of 3-year olds with CP±L are valid and reliable, and that multicentre outcome studies can be conducted in the preschool years for the benefit of children with cleft palate.

References

Ahl, R., Ahmad, T., Harding-Bell, A., Wharton, L., Jordan, A., and Hall, P. (2016) 'The Buccinator Mucomuscular Flap: An in-Depth Analysis and Evaluation of its Role in the Management of Velopharyngeal Dysfunction'. *The Cleft Palate-Craniofacial Journal* 53 (5), 177-184

Allori, A. C., Mulliken, J. B., Meara, J. G., Shusterman, S., and Marcus, J. R. (2017a) 'Classification of Cleft Lip/Palate: Then and Now'. *The Cleft Palate-Craniofacial Journal* 54 (2), 175-188

Allori, A. C., Kelley, T., Meara, J. G., Albert, A., Bonanthaya, K., Chapman, K., Cunningham, M., Daskalogiannakis, J., De Gier, H., and Heggie, A. A. (2017b) 'A Standard Set of Outcome Measures for the Comprehensive Appraisal of Cleft Care'. *The Cleft Palate-Craniofacial Journal* 54 (5), 540-554

Altman, D. G. (1991) *Practical Statistics for Medical Research.* London: Chapman and Hall/CRC press

Anandaciva, S. (2022) *What does the Spring Statement 2022 mean for the NHS?* [online] available from <https://www.kingsfund.org.uk/blog/2022/03/spring-statement-2022-nhs> [30 March 2022]

Apon, I., van Leeuwen, N., Allori, A. C., Rogers-Vizena, C. R., Koudstaal, M. J., Wolvius, E. B., Cano, S. J., Klassen, A. F., and Versnel, S. L. (2021) 'Rasch Analysis of Patient-and Parent-Reported Outcome Measures in the International Consortium for Health Outcomes Measurement Standard Set for Cleft Lip and Palate'. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 24 (3), 404-412

Arnaout, A., Martin, S., Hill, C., Hodgkinson, P., Thorburn, G., O'Neill, T. B., Moar, K., Treharne, L., Hall, P., and Miyagi, K. (2022) 'The Impact of COVID-19 on Cleft Services in Great Britain & Northern Ireland'. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 75 (5), 1689-1695

Atkinson, M. and Howard, S. (2011) 'Physical Structure and Function and Speech Production Associated with Cleft Palate'. in *Cleft Palate Speech, Assessment and Intervention.* ed. by Howard, S. and Lohmander, A. Oxford: Wiley-Blackwell, 5-22

Baigorri, M., Crowley, C., and Sommer, C. (2020) 'Addressing the Gap in Education for Cleft Palate: A Module Training Series for Craniofacial Assessment and Treatment'. *Perspectives of the ASHA Special Interest Groups* 5 (3), 662-668

Baillie, L. and Sell, D. (2020) 'Benchmarking Speech, Velopharyngeal Function Outcomes and Surgical Characteristics Following the Sommerlad Protocol and Palate Repair Technique'. *The Cleft Palate-Craniofacial Journal* 57 (10), 1197-1215

Basta, M. N., Silvestre, J., Stransky, C., Solot, C., Cohen, M., McDonald-McGinn, D., Zackai, E., Kirschner, R., Low, D. W., and Randall, P. (2014) 'A 35-Year Experience with Syndromic Cleft Palate Repair: Operative Outcomes and Long-Term Speech Function.'. *Annals of Plastic Surgery* 73, 130

Bates, S. and Titterington, J. (2021) *Good Practice Guidelines For the Analysis of Child Speech.* 2nd edn. [online] available from <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.nbt.nhs.uk/sites/default/files/document/Good%20practice%20guidelines%20for%20the%20analysis%20of%20child%20speech_2nd%20edition_2021.pdf> [14 June 2022]

Baylis, A., Chapman, K., and Whitehill, T. L. (2015) 'Validity and Reliability of Visual Analog Scaling for Assessment of Hypernasality and Audible Nasal Emission in Children with Repaired Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 52 (6), 660-670

Baylis, A. L., Munson, B., and Moller, K. T. (2011) 'Perceptions of Audible Nasal Emission in Speakers with Cleft Palate: A Comparative Study of Listener Judgments'. *The Cleft Palate-Craniofacial Journal* 48 (4), 399-411

Bearn, D., Mildinhall, S., Murphy, T., Murray, J. J., Sell, D., Shaw, W. C., Williams, A. C., and Sandy, J. R. (2001) 'Cleft Lip and Palate Care in the United Kingdom—the Clinical Standards Advisory Group (CSAG) Study. Part 4: Outcome Comparisons, Training, and Conclusions'. *The Cleft Palate-Craniofacial Journal* 38 (1), 38-43

Bell, J. C., Raynes-Greenow, C., Turner, R., Bower, C., Dodson, A., Nicholls, W., and Nassar, N. (2017) 'School Performance for Children with Cleft Lip and Palate: A Population-based Study'. *Child: Care, Health and Development* 43 (2), 222-231

Bessell, A., Sell, D., Whiting, P., Roulstone, S., Albery, L., Persson, M., Verhoeven, A., Burke, M., and Ness, A. R. (2013) 'Speech and Language Therapy Interventions for Children with Cleft Palate: A Systematic Review'. *The Cleft Palate-Craniofacial Journal* 50 (1), e1-e17

Bettens, K., Alighieri, C., Bruneel, L., De Meulemeester, L., and Van Lierde, K. (2020) 'Peer Attitudes Toward Children with Cleft (Lip and) Palate Related to Speech Intelligibility, Hypernasality and Articulation'. *Journal of Communication Disorders* 85, 105991

Bettens, K., Bruneel, L., Maryn, Y., De Bodt, M., Luyten, A., and Van Lierde, K. M. (2018) 'Perceptual Evaluation of Hypernasality, Audible Nasal Airflow and Speech Understandability using Ordinal and Visual Analogue Scaling and their Relation with Nasalance Scores'. *Journal of Communication Disorders* 76, 11-20

Bettens, K., Wuyts, F. L., and Van Lierde, K. M. (2014) 'Instrumental Assessment of Velopharyngeal Function and Resonance: A Review'. *Journal of Communication Disorders* 52, 170-183

Birch, A. L., Jordan, Z. V., Ferguson, L. M., Kelly, C. B., and Boorman, J. G. (2022) 'Speech Outcomes Following Orticochea Pharyngoplasty in Patients with History of Cleft Palate and Noncleft Velopharyngeal Dysfunction'. *The Cleft Palate-Craniofacial Journal* 59 (3), 277-290

Birmingham Community Safety Partnership (BCSP) (2013) report by Sharon Rowe, Assistant Chief Constable of West Midlands Police

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., and Basboll, H. (2008) 'Early Vocabulary Development in Danish and Other Languages: A CDI-Based Comparison'. *Journal of Child Language* 35 (3), 619-650

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S., Pascual, L., Pêcheux, M., Ruel, J., Venuti, P., and Vyt, A. (2004) 'Cross-linguistic Analysis of Vocabulary in Young Children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English'. *Child Development* 75 (4), 1115-1139

Bowen, C. (1998) *Ages and Stages Summary- Language Development 0-5 years* [online] available from <https://www.speech-language-therapy.com/index.php?option=com_content&view=article&id=34:ages&catid> [23 November 2022]

Brierley, N., Sainsbury, D., and Mehendale, F. (2022) 'COVID Cleft: A National Evaluation of Cleft Care Operating through the COVID-19 Pandemic in the United Kingdom'. *Cleft Palate-Craniofacial Journal*, 58

Britton L., Albery, L., Bowden, M., Harding-Bel,l A., Phippen, G., and Sell D (2016) National (UK) standards for speech for children born with cleft palate (+/-cleft lip /alveolus).

Britton, L., Albery, L., Bowden, M., Harding-Bell, A., Phippen, G., and Sell, D. (2014) 'A Cross-Sectional Cohort Study of Speech in Five-Year-Olds with Cleft Palate±lip to Support Development of National Audit Standards: Benchmarking Speech Standards in the United Kingdom'. *The Cleft Palate-Craniofacial Journal* 51 (4), 431-451

Bruneel, L., Bettens, K., De Bodt, M., D'haeseleer, E., Thijs, Z., Roche, N., and Van Lierde, K. (2020) 'Stages in the Development and Validation of a Belgian Dutch Outcome Tool for the Perceptual Evaluation of Speech in Patients with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 57 (1), 43-54

Brunnegård, K. and van Doorn, J. (2009) 'Normative Data on Nasalance Scores for Swedish as Measured on the Nasometer: Influence of Dialect, Gender, and Age'. *Clinical Linguistics & Phonetics* 23 (1), 58-69

Brunnegård, K. and Lohmander, A. (2007) 'A Cross-Sectional Study of Speech in 10-Year-Old Children with Cleft Palate: Results and Issues of Rater Reliability'. *The Cleft Palate-Craniofacial Journal* 44 (1), 33-44

Buckley, B. (2012) *Children's Communication Skills: From Birth to Five Years*. London:Routledge

Burg, M. L., Chai, Y., Yao, C. A., Magee III, W., and Figueiredo, J. C. (2016) 'Epidemiology, Etiology, and Treatment of Isolated Cleft Palate'. *Frontiers in Physiology* 7, 67

Butterworth, S., Rivers, C., Fullarton, M., Murphy, C., Beale, V., Neil-Dwyer, J., Van Eeden, S., Van Eeden, S., Hodgkinson, P. D., and Smyth, A. (2022) 'A Closer Look at Delayed Primary Cleft Surgery and Unrepaired Cleft Lip and/Or Palate in 5 UK Cleft Centers'. *The Cleft Palate-Craniofacial Journal* 59 (6), 724-731

Calzolari, E., Pierini, A., Astolfi, G., Bianchi, F., Neville, A. J., Rivieri, F., and EUROCAT Working Group (2007) 'Associated Anomalies in Multi-malformed Infants with Cleft Lip and Palate: An Epidemiologic Study of nearly 6 Million Births in 23 EUROCAT Registries'. *American Journal of Medical Genetics Part A* 143 (6), 528-537

CAPS-A (2022) *CAPS-A Training* [online] available from <https://www.caps-a.com/> [29 November 2022]

Castick, S., Knight, R., and Sell, D. (2017) 'Perceptual Judgments of Resonance, Nasal Airflow, Understandability, and Acceptability in Speakers with Cleft Palate: Ordinal Versus Visual Analogue Scaling'. *The Cleft Palate-Craniofacial Journal* 54 (1), 19-31

Chacon, A., Parkin, M., Broome, K., and Purcell, A. (2017) 'Australian Children with Cleft Palate Achieve Age-Appropriate Speech by 5 Years-of-Age'. *International Journal of Pediatric Otorhinolaryngology* 103, 93-102

Chapman, K. L., Baylis, A., Trost-Cardamone, J., Cordero, K. N., Dixon, A., Dobbelsteyn, C., Thurmes, A., Wilson, K., Harding-Bell, A., and Sweeney, T. (2016) 'The Americleft Speech Project: A Training and Reliability Study'. *The Cleft Palate-Craniofacial Journal* 53 (1), 93-108

Chapman, K. L. (2011) 'The Relationship between Early Reading Skills and Speech and Language Performance in Young Children with Cleft Lip and Palate'. *Cleft Palate–Craniofacial Journal* 48 (3), 301-311

Chapman, K.L. and Willadsen, E. (2011) 'The Development of Speech in Children with Cleft Palate'. in *Cleft Palate Speech, Assessment and Intervention.* ed. by Howard, S. and Lohmander, A. Oxford: Wiley-Blackwell, 23-40

Chapman, K. L., Hardin-Jones, M. A., Goldstein, J. A., Halter, K. A., Havlik, R. J., and Schulte, J. (2008) 'Timing of Palatal Surgery and Speech Outcome'. *The Cleft Palate-Craniofacial Journal* 45 (3), 297-308

Chapman, K. L., Hardin-Jones, M., and Halter, K. A. (2003) 'The Relationship between Early Speech and Later Speech and Language Performance for Children with Cleft Lip and Palate'. *Clinical Linguistics & Phonetics* 17 (3), 173-197

Choa, R. M., Slator, R., Jeremy, A., Robinson, S., Franklin, D., Roberts, A., Winter, R., Extence, H., Collard, M., and Clark, V. (2014) 'Identifying the Effect of Cleft Type, Deprivation and Ethnicity on Speech and Dental Outcomes in UK Cleft Patients: A Multi-Centred Study'. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 67 (12), 1637-1643

CLEAR RESOURCES. (2006) *CLEAR Phonology Screening Assessment*. Spilsby: CLEAR Resources

Cleland, J., Crampin, L., Campbell, L., and Dokovova, M. (2021) 'Protocol for Sonospeech Cleft Pilot: A Pilot Randomised Control Trial of Ultrasound Visual Biofeedback Versus Standard Intervention for Children with Cleft Lip and Palate.'

Cleland, J., Lloyd, S., Campbell, L., Crampin, L., Palo, J., Sugden, E., Wrench, A., and Zharkova, N. (2020) 'The Impact of Real-Time Articulatory Information on Phonetic Transcription: Ultrasound-Aided Transcription in Cleft Lip and Palate Speech'. *Folia Phoniatrica Et Logopaedica* 72 (2), 120-130

CLISPI (2017) *Good Practice* [online] available from <https://clispi.com/clispi.php?id=265&mid=265> [16 June 2022]

Cornefjord, M., Jakobsson, A., Stiernman, M., Blumhe, J., Mamedov, T., Klintö.K, Svensson, H., Whitling, S. and Becker, M. (2022) 'Using Artificial Intelligence to Evaluate Velopharyngeal Competence in Cleft Lip and Palate Children'. in Mehendale, F. V. (ed.) *Book of Abstracts,* '14th International Congress of Cleft Lip and Palate and Related Craniofacial Anomalies'. held 11-15 July 2022 at Edinburgh International Conference Centre, Edinburgh, UK

Counihan, D. T. and Cullinan, W. L. (1970) 'Reliability and Dispersion of Nasality Ratings'. *The Cleft Palate Journal* 7 (1), 261-270

CRANE (2021) *2021 Annual Report* [online] available from <https://www.crane-database.org.uk/reports/crane-database-2021-annual-report> [28 September 2022]

CRANE (2020) *2020 Annual Report* [online] available from < https://www.crane-database.org.uk/reports/crane-database-2020-annual-report/> [14 September 2022]

CRANE (2019) *2019 Annual Report* [online] available from < https://www.crane-database.org.uk/reports/crane-database-2019-annual-report/> [13 January 2021]

CRANE (2017) *Annual Report on Cleft Lip and/or Palate 2017* [online] available from < chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.crane-database.org.uk/content/uploads/2020/08/CRANE-Annual-Report-2017-v1.1.pdf> [22 October 2021]

CRANE (2016) *Annual Report on Cleft Lip and/or Palate 2016* [online] available from <https://www.crane-database.org.uk/reports/crane-database-2016-annual-report/> [22 October 2021]

Cronin, A. and McLeod, S. (2019) 'Craniofacial Anomalies'. in *The SAGE Encyclopedia of Human Communication Sciences and Disorders*. ed. by Ball, M. J. and Damico, J. London: SAGE Publications Ltd, 515-519

Croot, K., Au, C., and Harper, A. (2010) 'Prosodic Structure and Tongue Twister Errors'. *Papers in Laboratory Phonology* 10, 433-459

Crystal, D. (2002) *The Cambridge Encyclopaedia of Language Second Edition.* Cambridge: Cambridge University Press

D'Antonio, L. L., Scherer, N. J., Miller, L. L., Kalbfleisch, J. H., and Bartley, J. A. (2001) 'Analysis of Speech Characteristics in Children with Velocardiofacial Syndrome (VCFS) and Children with Phenotypic Overlap without VCFS'. *The Cleft Palate-Craniofacial Journal* 38 (5), 455-467

Day, S. J. and Altman, D. G. (2000) 'Blinding in Clinical Trials and Other Studies'. *BMJ* 321 (7259), 504

Dayashankara, R. J., Singh, S., Suma, G. N., Gokkula, K., Kumar, V., and Sursh, B. S. (2011) 'Comparative Study of 2 Palatoplasty Techniques to Assess Speech and Fistula in Primary Cleft Palate Patients'. *Journal of Dentistry for Children* 78 (1), 13-18

Dixon, M. J., Marazita, M. L., Beaty, T. H., and Murray, J. C. (2011) 'Cleft Lip and Palate: Understanding Genetic and Environmental Influences'. *Nature Reviews Genetics* 12 (3), 167-178

Dodd, B., Holm, A., Hua, Z., and Crosbie, S. (2003) 'Phonological Development: A Normative Study of British English-speaking Children'. *Clinical Linguistics & Phonetics* 17 (8), 617-643

Dodd, B., Zhy, H., Crosbie S., Holm, A. and Ozanne, A.(2002) 'Diagnostic evaluation of articulation and phonology (DEAP)'. *Psychology Corporation*

Dubey, A. K., Prasanna, S. M., and Dandapat, S. (2018) *Pitch-Adaptive Front-End Feature for Hypernasality Detection.* 'Interspeech 2018'. held 2-6 September at Hyderabad. ISCA

El Ezzi, O., Herzog, G., Broome, M., Trichet-Zbinden, C., Hohlfeld, J., Cherpillod, J., and de Buys Roessingh, A. S. (2015) 'Grommets and Speech at Three and Six Years in Children Born with Total Cleft Or Cleft Palate'. *International Journal of Pediatric Otorhinolaryngology* 79 (12), 2243-2247

El-Shazly, M., Helmy, Y., Abdelsalam, L., and Ali, T. (2022) 'Global Incidence of Cleft Palate'. in *Surgical Atlas of Cleft Palate and Palatal Fistulae*. ed. by Fayyaz, G.Q. Berlin: Springer, 29-34

Feragen, K. B., Aukner, R., Saervold, T. K., and Hide, Ø (2017) 'Speech, Language, and Reading Skills in 10-Year-Old Children with Palatal Clefts: The Impact of Additional Conditions'. *Journal of Communication Disorders* 66, 1-12

Feragen, K. B. and Stock, N. M. (2014) 'When there is More than a Cleft: Psychological Adjustment when a Cleft is Associated with an Additional Condition'. *The Cleft Palate-Craniofacial Journal* 51 (1), 5-14

Fitzpatrick, B., Panagamuwa, C., Levy, L. M., and Rihtman, T. (2021) 'The Impact of Hearing Loss on Speech Outcomes in 5-Year-Old Children with Cleft Palate±lip: A Longitudinal Cohort Study'. *International Journal of Pediatric Otorhinolaryngology* 149, 110870

Fitzpatrick, B., Coad, J., Sell, D., and Rihtman, T. (2020) 'Assessing Speech at Three Years of Age in the Cleft Palate Population: A Scoping Review of Assessment Practices'. *International Journal of Language & Communication Disorders* 55 (2), 165-187

Flynn, T. and Lohmander, A. (2014) 'A Longitudinal Study of Hearing and Middle Ear Status in Individuals with UCLP'. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society and European Academy of Otology and Neurotology* 35 (6), 989-996

Flynn, T., Persson, C., Moller, C., Lohmander, A., and Magnusson, L. (2014) 'A Longitudinal Study of Hearing and Middle Ear Status of Individuals with Cleft Palate with and without Additional Malformations/Syndromes'. *The Cleft Palate-Craniofacial Journal* 51 (5), e94-e101

Flynn, T. (2013) *Hearing and middle ear status in children and young adults with cleft palate* [online] available from <https://gupea.ub.gu.se/bitstream/2077/31719/2/gupea_2077_31719_2.pdf> [04 April 2018]

Frederickson, M. S., Chapman, K. L., and Hardin-Jones, M. (2006) 'Conversational Skills of Children with Cleft Lip and Palate: A Replication and Extension'. *The Cleft Palate-Craniofacial Journal* 43 (2), 179-188

Frey, J. R., Kaiser, A. P., and Scherer, N. J. (2018) 'The Influences of Child Intelligibility and Rate on Caregiver Responses to Toddlers with and without Cleft Palate'. *The Cleft Palate-Craniofacial Journal*, 1055665617727000

Gallagher, E. R. and Collett, B. R. (2019) 'Neurodevelopmental and Academic Outcomes in Children with Orofacial Clefts: A Systematic Review'. *Pediatrics* 144 (1), e20184027

Gibbon, F. E., Ellis, L., and Crampin, L. (2004) 'Articulatory Placement for/T/,/D/,/K/and/G/Targets in School Age Children with Speech Disorders Associated with Cleft Palate'. *Clinical Linguistics & Phonetics* 18 (6-8), 391-404

Gibbon, F. E. and Crampin, L. B. (2001) 'An Electropalatographic Investigation of Middorsum Palatal Stops in an Adult with Repaired Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 38 (2), 96-105

Gnanavel, K. and Pushpavathi, M. (2012) 'Effect of Type of Cleft and Speech Stimuli on Speech Intelligibility in Individuals with Cleft Lip and Palate.'. *Journal of the all India Institute of Speech & Hearing* 31

Gooch, J. L., Hardin-Jones, M., Chapman, K. L., Trost-Cardamone, J. E., and Sussman, J. (2001) 'Reliability of Listener Transcriptions of Compensatory Articulations'. *The Cleft Palate-Craniofacial Journal* 38 (1), 59-67

Gray, S., Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., and Cowan, N. (2019) 'Working Memory Profiles of Children with Dyslexia, Developmental Language Disorder, Or both.'. *Journal of Speech, Language, and Hearing Research: JSLHR* 62 (6), 1839-1858

Grunwell, P. and Sell, D. (2001a) 'Speech Assessment and Therapy'. in *Management of Cleft Lip and Palate*. ed. by Watson, A.C.H., Sell.D.A. and Grunwell. P. London: Whurr Publishers, 227-257

Grunwell, P. and Sell, D. (2001b) 'Speech and Cleft Palate/Velopharyngeal Anomalies'. in *Management of Cleft Lip and Palate*. ed. by Watson, A.C.H., Sell.D.A. and Grunwell. P. London: Whurr Publishers, 68-86

Grunwell P, Harding A. (1985) *PACS Toys Screening Assessment* Nfer-Nelson

Gugsch, C., Dannhauer, K., and Fuchs, M. (2008) 'Evaluation of the Progress of Therapy in Patients with Cleft Lip, Jaw and Palate, using Voice Analysis a Pilot Study'. *Journal of Orofacial Orthopedics/Fortschritte Der KieferorthopÃ¤die* 69 (4), 257-267

Gunther, E., Wisser, J. R., Cohen, M. A., and Brown, A. S. (1998) 'Palatoplasty: Furlow's Double Reversing Z-Plasty Versus Intravelar Veloplasty'. *The Cleft Palate-Craniofacial Journal* 35 (6), 546-549

Ha, S. and Oller, K. D. (2022) 'Longitudinal Study of Vocal Development and Language Environments in Infants with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 59 (10), 1286-1298

Hamming, K. K., Finkelstein, M., and Sidman, J. D. (2009) 'Hoarseness in Children with Cleft Palate'. *Otolaryngology—Head and Neck Surgery* 140 (6), 902-906

Handžić-Ćuk, J., Ćuk, V., Rišavi, R., Katušić, D., and Štajner-Katušić, S. (1996) 'Hearing Levels and Age in Cleft Palate Patients'. *International Journal of Pediatric Otorhinolaryngology* 37 (3), 227-242

Hardin-Jones, M. A. and Chapman, K. L. (2022) 'Non-Oral Compensatory Misarticulations Revisited'. *The Cleft Palate-Craniofacial Journal* 59 (8), 976-983

Hardin-Jones, M. and Chapman, K. L. (2014) 'Early Lexical Characteristics of Toddlers with Cleft Lip and Palate'. *The Cleft Palate-Craniofacial Journal* 51 (6), 622-631

Hardin-Jones, M. and Chapman, K. L. (2011) 'Cognitive and Language Issues Associated with Cleft Lip and Palate'. *Seminars in Speech and Language* 32 (2), 127-140

Hardin-Jones, M. A. and Jones, D. L. (2005) 'Speech Production of Preschoolers with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 42 (1), 7-13

Harding-Bell, A. (2019) 'Differential diagnosis of velopharyngeal dysfunction'. in *Case Studies in Cleft Palate Speech, Data analysis and principled intervention.* ed. Harding-Bell, A. Croydon: J&R Press, 157-188

Harding, A. and Grunwell, P. (1998) 'Active Versus Passive Cleft-type Speech Characteristics'. *International Journal of Language & Communication Disorders* 33 (3), 329-352

Hardwicke, J. T., Richards, H., Cafferky, L., Underwood, I., ter Horst, B., and Slator, R. (2016) 'Outcomes of Cleft Palate Repair in Patients with Pierre Robin Sequence: A Matched Case-Control Study'. *Plastic and Reconstructive Surgery* 137 (3), 927-935

Harman, N. L., Bruce, I. A., Kirkham, J. J., Tierney, S., Callery, P., O'Brien, K., Bennett, A. M., Chorbachi, R., Hall, P. N., and Harding-Bell, A. (2015) 'The Importance of Integration of Stakeholder Views in Core Outcome Set Development: Otitis Media with Effusion in Children with Cleft Palate'. *PLoS One* 10 (6), e0129514

Hattee, C., Farrow, K., Harland, K., Sommerlad, B., and Walsh, M. (2001) 'Are we Ready to Predict Speech Development from Babble in Cleft Lip and Palate Children?'. *International Journal of Language & Communication Disorders* 36 (S1), 115-120

Henningsson, G., Kuehn, D. P., Sell, D., Sweeney, T., Trost-Cardamone, J. E., and Whitehill, T. L. (2008) 'Universal Parameters for Reporting Speech Outcomes in Individuals with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 45 (1), 1-17

Hens, G., Sell, D., Pinkstone, M., Birch, M. J., Hay, N., Sommerlad, B. C., and Kangesu, L. (2013) 'Palate Lengthening by Buccinator Myomucosal Flaps for Velopharyngeal Insufficiency'. *The Cleft Palate-Craniofacial Journal* 50 (5), 84-91

Hodge, M. and Gotzke, C. L. (2007) 'Preliminary Results of an Intelligibility Measure for English-Speaking Children with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 44 (2), 163-174

Hoghoughi, M. A., Kazemi, T., Khojasteh, A., Habibagahi, R., Kalkate, Z., Zarei, Z., Hosseinpour, H., and Salimi, M. (2021) 'The Effect of Intervelar Veloplasty Under Magnification (Sommerlad's Technique) without Tympanostomy on Middle Ear Effusion in Cleft Palate Patients'. *BMC Pediatrics* 21 (1), 1-6

Howard, S., Heselwood, B. and Harding-Bell, A. (2019) ‘The nature of speech associated with cleft palate’. in *Case Studies in Cleft Palate Speech, Data analysis and principled intervention.* ed. Harding-Bell, A. Croydon: J&R Press, 23-50

Howard, S. (2013) 'A Phonetic Investigation of Single Word Versus Connected Speech Production in Children with Persisting Speech Difficulties Relating to Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 50 (2), 207-223

Howard, S. (2011) ‘Phonetic Transcription for Speech Related to Cleft Palate’. in *Cleft Palate Speech, Assessment and Intervention.* ed. by Howard, S. and Lohmander, A. Oxford: Wiley-Blackwell, 127-144

Howard, S. J. and Heselwood, B. C. (2002) 'Learning and Teaching Phonetic Transcription for Clinical Purposes'. *Clinical Linguistics & Phonetics* 16 (5), 371-401

Hulme, C. and Snowling, M. J. (2014) 'The Interface between Spoken and Written Language: Developmental Disorders'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1634), 20120395

Hutters, B., Bau, A., and Brondsted, K. (2001) 'A Longitudinal Group Study of Speech Development in Danish Children Born with and without Cleft Lip and Palate'. *International Journal of Language & Communication Disorders* 36 (4), 447-470

Jabal, J. (2018) *Approaches to better value: improving quality and cost* [online] available from < https://www.kingsfund.org.uk/publications/approaches-better-value> [04 November 2018]

James, D. G. H. (2001) ‘The use of phonological processes in Australian children aged 2 to 7:11 years’. *Advances in Speech-Language Pathology* 3 109-128

James Lind Alliance (2012) *Cleft Lip and Palate Top 10 Priorities* [online] available from < https://www.jla.nihr.ac.uk/priority-setting-partnerships/cleft-lip-and-palate/top-10-priorities/> [16 March 2019]

Jiang, R., Bush, J. O., and Lidral, A. C. (2006) 'Development of the Upper Lip: Morphogenetic and Molecular Mechanisms'. *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 235 (5), 1152-1166

John, A., Sell, D., Sweeney, T., Harding-Bell, A., and Williams, A. (2006) 'The Cleft Audit Protocol for Speech-Augmented: A Validated and Reliable Measure for Auditing Cleft Speech'. *The Cleft Palate-Craniofacial Journal* 43 (3), 272-288

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., Brownlie, E. B., Douglas, L., Taback, N., and Lam, I. (1999) 'Fourteen-Year Follow-Up of Children with and

without Speech/Language Impairments Speech/Language Stability and Outcomes'. *Journal of Speech, Language, and Hearing Research* 42 (3), 744-760

Jørgensen, L. D. and Willadsen, E. (2020) 'Longitudinal Study of the Development of Obstruent Correctness from Ages 3 to 5 Years in 108 Danish Children with Unilateral Cleft Lip and Palate: A Sub-study within a Multicentre Randomized Controlled Trial'. *International Journal of Language & Communication Disorders* 55 (1), 121-135

Kaptein, M. C., Nass, C., and Markopoulos, P. (2010) 'Powerful and Consistent Analysis of Likert-Type Rating Scales'. in Jones, M. (ed.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. held 10- 15 April in Atlanta. New York: Association for Computing Machinery

Karnell, M. P. (ed.) (2011) 'Instrumental Assessment of Velopharyngeal Closure for Speech'. *Seminars in Speech and Language* 32 (2), 168-178

Kay Elemetrics, C. (2001). NasometerTM II Models 6400/6402: Installation/Instruction manual. Lincoln Park, New Jersey: Kay Elemetrics.

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., and Duffy, J. R. (1999) 'Acoustic Studies of Dysarthric Speech: Methods, Progress, and Potential'. *Journal of Communication Disorders* 32 (3), 141-186

Kent, R. D. (1996) 'Hearing and Believing: Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders'. *American Journal of Speech-Language Pathology* 5 (3), 7-23

Kilcoyne, S., Luscombe, C., Scully, P., Overton, S., Brockbank, S., Swan, M. C., Johnson, D., Wall, S., and Wilkie, A. O. (2022) 'Hearing, Speech, Language, and Communicative Participation in Patients with Apert Syndrome: Analysis of Correlation with Fibroblast Growth Factor Receptor 2 Mutation'. *Journal of Craniofacial Surgery* 33 (1), 243-250

Kuehn, D. P. and Moller, K. T. (2000) 'Speech and Language Issues in the Cleft Palate Population: The State of the Art'. *The Cleft Palate-Craniofacial Journal* 37 (4), 1-35

Keuning, K. H., Wieneke, G. H., Van Wijngaarden, H. A., and Dejonckere, P. H. (2002) 'The Correlation between Nasalance and a Differentiated Perceptual Rating of Speech in Dutch Patients with Velopharyngeal Insufficiency'. *The Cleft Palate-Craniofacial Journal* 39 (3), 277-284

Keuning, K. H., Wieneke, G. H., and Dejonckere, P. H. (1999) 'The Intrajudge Reliability of the Perceptual Rating of Cleft Palate Speech before and After Pharyngeal Flap Surgery: The Effect of Judges and Speech Samples'. *The Cleft Palate-Craniofacial Journal* 36 (4), 328-333

Khoshab, N., Kanack, M., Kadakia, N., Vyas, R. M., and Chase, L. (2021) 'Newly Identified Developmental Delays in a Large Population of Children with Nonsyndromic Cleft Lip and/Or Palate'. *Plastic and Reconstructive Surgery–Global Open* 9 (10S), 50-51

Kilcoyne, S., Potter, K. R., Gordon, Z., Overton, S., Brockbank, S., Jayamohan, J., Magdum, S., Smith, M., Johnson, D., and Wall, S. (2021) 'Feeding, Communication, Hydrocephalus, and Intracranial Hypertension in Patients with Severe FGFR2-Associated Pfeiffer Syndrome'. *Journal of Craniofacial Surgery* 32 (1), 134-140

Klintö, K. and Lohmander, A. (2017) 'Does the Recording Medium Influence Phonetic Transcription of Cleft Palate Speech?'. *International Journal of Language & Communication Disorders* 52 (4), 440-449

Klintö, K., Salameh, E., and Lohmander, A. (2016) 'Phonology in Swedish-Speaking 5-Year-Olds Born with Unilateral Cleft Lip and Palate and the Relationship with Consonant Production at 3 Years of Age'. *International Journal of Speech-Language Pathology* 18 (2), 147-156

Klintö, K., Salameh, E., and Lohmander, A. (2015) 'Verbal Competence in Narrative Retelling in 5-year olds with Unilateral Cleft Lip and Palate'. *International Journal of Language & Communication Disorders* 50 (1), 119-128

Klintö, K., Olsson, M., Flynn, T., Svensson, H., and Lohmander, A. (2014a) 'Phonology in Swedish-speaking 3-year-olds Born with Cleft Lip and Palate and the Relationship with Consonant Production at 18 Months'. *International Journal of Language & Communication Disorders* 49 (2), 240-254

Klintö, K., Svensson, H., Elander, A., and Lohmander, A. (2014b) 'Speech and Phonology in Swedish-Speaking 3-Year-Olds with Unilateral Complete Cleft Lip and Palate Following Different Methods for Primary Palatal Surgery'. *The Cleft Palate-Craniofacial Journal* 51 (3), 274-282

Klintö, K., Salameh, E., Svensson, H., and Lohmander, A. (2011) 'The Impact of Speech Material on Speech Judgement in Children with and without Cleft Palate'. *International Journal of Language & Communication Disorders*, 1-13

Konst, E. M., Rietveld, T., Peters, H. F., and Prahl-Andersen, B. (2003) 'Phonological Development of Toddlers with Unilateral Cleft Lip and Palate Who were Treated with and without Infant Orthopedics: A Randomized Clinical Trial'. *The Cleft Palate-Craniofacial Journal* 40 (1), 32-39

Kotlarek, K. J., Sitzman, T. J., Williams, J. L., and Perry, J. L. (2021) 'Nonsedated Magnetic Resonance Imaging for Visualization of the Velopharynx in the Pediatric Population'. *The Cleft Palate-Craniofacial Journal*, 10556656211057361

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (1993) 'Perceptual Evaluation of Voice Quality: Review, Tutorial, and a Framework for Future Research'. *Journal of Speech, Language, and Hearing Research* 36 (1), 21-40

Kuehn, D. P. and Moller, K. T. (2000) 'Speech and Language Issues in the Cleft Palate Population: The State of the Art'. *The Cleft Palate-Craniofacial Journal* 37 (4), 1-35

Kummer, A. W. (2021) 'Speech and Resonance Disorders Secondary to Cleft Lip/Palate (CLP)'. in *Global Cleft Care in Low-Resource Settings*. ed. by Swanson, J.W. and Low, D. W. Berlin: Springer, 405-418

Kummer, A. W., Clark, S. L., Redle, E. E., Thomsen, L. L., and Billmire, D. A. (2012) 'Current Practice in Assessing and Reporting Speech Outcomes of Cleft Palate and Velopharyngeal Surgery: A Survey of Cleft Palate/Craniofacial Professionals'. *The Cleft Palate-Craniofacial Journal* 49 (2), 146-152

Lancaster, H. S., Lien, K. M., Chow, J. C., Frey, J. R., Scherer, N. J., and Kaiser, A. P. (2020) 'Early Speech and Language Development in Children with Nonsyndromic Cleft Lip and/Or Palate: A Meta-Analysis'. *Journal of Speech, Language, and Hearing Research* 63 (1), 14-31

Landis, J. R. and Koch, G. G. (1977) 'The Measurement of Observer Agreement for Categorical Data'. *Biometrics*, 159-174

Larossa, D. (2000) 'The State of the Art in Cleft Palate Surgery'. *The Cleft Palate-Craniofacial Journal* 37 (3), 225-228

Larsson, A., Schölin, J., Mark, H., Jönsson, R., and Persson, C. (2017) 'Speech Production in 3-year-old Internationally Adopted Children with Unilateral Cleft Lip and Palate'. *International Journal of Language & Communication Disorders* 52 (5), 626-636

Law, J., Boyle, J., Harris, F., Harkness, A., and Nye, C. (2000) 'Prevalence and Natural History of Primary Speech and Language Delay: Findings from a Systematic Review of the Literature'. *International Journal of Language and Communication Disorders* 35, 165-188

Lee, A., Liker, M., Fujiwara, Y., Yamamoto, I., Takei, Y., and Gibbon, F. (2022) 'EPG Research and Therapy: Further Developments'. *Clinical Linguistics & Phonetics*, 1-21

Lee, A. (2021) 'Electropalatography'. in *Manual of Clinical Phonetics*. ed. by Ball, M.J. New York: Routledge, 339-355

Lee, A., Bessell, N., and Gibbon, F. E. (2019) 'Tongue-Palate Contact for Nasal Versus Oral Stops in Speakers with Repaired Cleft Palate'. *Clinical Linguistics & Phonetics* 33 (9), 831-853

Lee, A., Whitehill, T. L., and Ciocca, V. (2009) 'Effect of Listener Training on Perceptual Judgement of Hypernasality'. *Clinical Linguistics & Phonetics* 23 (5), 319-334

Lethaus, B., Grau, E., Kloss-Brandstätter, A., Brauer, L., Zimmerer, R., Bartella, A. K., Hahnel, S., and Sander, A. K. (2021) 'Clinical Follow-Up in Orofacial Clefts—Why Multidisciplinary Care is the Key'. *Journal of Clinical Medicine* 10 (4), 842

Lewis, K. E., Watterson, T. L., and Houghton, S. M. (2003) 'The Influence of Listener Experience and Academic Training on Ratings of Nasality'. *Journal of Communication Disorders* 36 (1), 49-58

Lewis, K. E., Watterson, T., and Quint, T. (2000) 'The Effect of Vowels on Nasalance Scores'. *The Cleft Palate-Craniofacial Journal* 37 (6), 584-589

Lohmander, A., Persson, C., Willadsen, E., Lundeborg, I., Alaluusua, S., Aukner, R., Bau, A., Boers, M., Bowden, M., and Davies, J. (2017a) 'Scandcleft Randomised Trials of Primary Surgery for Unilateral Cleft Lip and Palate: 4. Speech Outcomes in 5-Year-Olds-Velopharyngeal Competency and Hypernasality.'. *Journal of Plastic Surgery and Hand Surgery* 51 (1), 27-37

Lohmander, A., Lundeborg, I., and Persson, C. (2017b) 'SVANTE–The Swedish Articulation and Nasality Test–Normative Data and a Minimum Standard Set for Cross-Linguistic Comparison'. *Clinical Linguistics & Phonetics* 31 (2), 137-154

Lohmander, A., Hagberg, E., Persson, C., Willadsen, E., Lundeborg, I., Davies, J., Havstam, C., Boers, M., Kisling-Møller, M., and Alaluusua, S. (2017c) 'Validity of Auditory Perceptual Assessment of Velopharyngeal Function and Dysfunction–the VPC-Sum and the VPC-Rate'. *Clinical Linguistics & Phonetics* 31 (7-9), 589-597

Lohmander, A. and Howard, S. (2011) 'Speech Assessment and Intervention'. in *Cleft Palate Speech, Assessment and Intervention.* ed. by Howard, S. and Lohmander, A. Oxford: Wiley-Blackwell, 123-126

Lohmander, A., Willadsen, E., Persson, C., Henningsson, G., Bowden, M., and Hutters, B. (2009) 'Methodology for Speech Assessment in the Scandcleft Project—an International Randomized Clinical Trial on Palatal Surgery: Experiences from a Pilot Study'. *The Cleft Palate-Craniofacial Journal* 46 (4), 347-362

Lohmander, A. and Persson, C. (2008) 'A Longitudinal Study of Speech Production in Swedish Children with Unilateral Cleft Lip and Palate and Two-Stage Palatal Repair'. *The Cleft Palate-Craniofacial Journal* 45 (1), 32-41

Lohmander, A., Friede, H., Elander, A., Persson, C., and Lilja, J. (2006) 'Speech Development in Patients with Unilateral Cleft Lip and Palate Treated with Different Delays in Closure of the Hard Palate After Early Velar Repair: A Longitudinal Perspective'. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery* 40 (5), 267-274

Lohmander-Agerskov, A. and Forss, M. H. (2005) *SVANTE: Svenskt Artikulations-Och Nasalitets-Test, manual* [In Swedish] 2nd edn. Lund: Pedagogisk design

Lohmander, A. and Olsson, M. (2004) 'Methodology for Perceptual Assessment of Speech in Patients with Cleft Palate: A Critical Review of the Literature'. *The Cleft Palate-Craniofacial Journal* 41 (1), 64-70

Lohmander-Agerskov, A. (1998) 'Speech Outcome After Cleft Palate Surgery with the G Teborg Regimen Including Delayed Hard Palate Closure'. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery* 32 (1), 63-80

Lohmander-Agerskov, A., Söderpalm, E., Friede, H., and Lilja, J. (1998) 'A Comparison of Babbling and Speech at Pre-Speech Level, 3, and 5 Years of Age in Children with Cleft Lip and Palate Treated with Delayed Hard Palate Closure'. *Folia Phoniatrica Et Logopaedica* 50 (6), 320-334

Mandrekar, J. N. (2011) 'Measures of Interrater Agreement'. *Journal of Thoracic Oncology* 6 (1), 6-7

Mani, M., Morén, S., Thorvardsson, O., Jakobsson, O., Skoog, V., and Holmström, M. (2010) 'Objective Assessment of the Nasal Airway in Unilateral Cleft Lip and Palate—a Long-Term Study'. *The Cleft Palate-Craniofacial Journal* 47 (3), 217-224

Mason, K. N. (2022) 'Magnetic Resonance Imaging for Assessing Velopharyngeal Function: Current Applications, Barriers, and Potential for Future Clinical Translation in the United States'. *The Cleft Palate-Craniofacial Journal*, DOI: 10556656221123916

McCormack, J., McLeod, S., Harrison, L. J., and McAllister, L. (2010) 'The Impact of Speech Impairment in Early Childhood: Investigating Parents' and Speech-Language Pathologists' Perspectives using the ICF-CY'. *Journal of Communication Disorders* 43 (5), 378-396

McGurk, H. and MacDonald, J. (1976) 'Hearing Lips and Seeing Voices'. *Nature* 264 (5588), 746

McLeod, S. (2020) 'Intelligibility in Context Scale: Cross-Linguistic use, Validity, and Reliability'. *Speech, Language and Hearing* 23 (1), 9-16

McLeod, S. and Baker, E. (2017) *Children's Speech, An Evidence-Based Approach to Assessment and Intervention.* Boston: Pearson

McLeod, S., Harrison, L. J., and McCormack, J. (2012) 'The Intelligibility in Context Scale: Validity and Reliability of a Subjective Rating Measure'. *Journal of Speech, Language, andHearing Research* 55(2), 648–656.

McLeod, S. (2009) *Children's Speech Acquisition* [online] available at <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cdn.csu.edu.au/__data/assets/pdf_file/0006/227652/Speech-acquisition-summary.pdf> [14 January 2021]

McLeod, S., Van Doorn, J., and Reed, V. A. (2001a) 'Normal Acquisition of Consonant Clusters'. *American Journal of Speech-Language Pathology* 10 (2), 99

McLeod, S., Van Doorn, J., and Reed, V. (2001b) 'Consonant Cluster Development in Two-Year-Olds: General Trends and Individual Difference'. *Journal of Speech, Language, and Hearing Research* 44 (5), 1144-1171

Meng, L., Bian, Z., Torensma, R., and Von den Hoff, J. W. (2009) 'Biological Mechanisms in Palatogenesis and Cleft Palate'. *Journal of Dental Research* 88 (1), 22-33

Microtronics Corporation (1994) *PERCI SARS system manual.* Chapel Hill: Microtronics Corporation

Mills, J. and Birks, M. (2014) 'Qualitative Methodology: A practical guide'. in *Introducing Qualitative Research.* ed. by Mills, J. and Birks, M. London: Sage, 3-16

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Prisma Group (2009) 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement'. *PLoS Medicine* 6 (7), e1000097

Morris, H. and Ozanne, A. (2003) 'Phonetic, Phonological, and Language Skills of Children with a Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 40 (5), 460-470

Morrison, M. M., Mason, N. T., Forde, B. L., Stone, P. R., Fowler, P. V., and Thompson, J. M. (2022) 'Speech Outcomes of a National Cohort of Children with Orofacial Cleft at 5 and 10 Years of Age'. *The Cleft Palate-Craniofacial Journal* 59 (11), 1400-1412

Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., and Shaw, W. C. (2009) 'Cleft Lip and Palate'. *The Lancet* 374 (9703), 1773-1785

Nachmani, A., Masalha, M., and Kassem, F. (2021) 'Phonological Profile of Patients with Velopharyngeal Dysfunction and Palatal Anomalies'. *Journal of Speech, Language, and Hearing Research* 64 (12), 4649-4663

NHS England (2022) *Cleft Lip and/or Palate Services including Non-Cleft Veloparyngeal Dysfunction (VPD) (All Ages)* [online] available from < chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.england.nhs.uk/wp-content/uploads/2022/04/Cleft-Lip-and-or-Palate-Services-including-Non-Cleft-Velopharyngeal-Dysfunction-VPD-All-Ages.pdf> [22 October 2022]

NHS Choices (2019) *Overview, Cleft Lip and Palate* [online] available from <https://www.nhs.uk/conditions/cleft-lip-and-palate> [20 July 2021]

NHS England (2018) *Cleft Lip and/or Palate Quality Dashboard 2018/19* [online] available from < chrome-

extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.england.nhs.uk/wp-content/uploads/2018/03/clp-metric-definitions-2018-19.pdf> [22 November 2022]

NHS (2014) *Five Year Forward View* [online] available from < chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf> [10 July 2019]

NHS England (2013) *NHS Standard Contract for Cleft Lip and/or Palate Services including Non-Cleft Velopharyngeal Dysfunction (VPD) (All ages)* [online] available from <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.england.nhs.uk/wp-content/uploads/2013/06/d07-cleft-lip.pdf> [13 December 2019]

NIHR CRN (2013) *Introduction to Good Clinical Practise (GCP): A practical guide to ethical and scientific quality standards in clinical research*. Leeds: NIHR

Nyberg, J., Peterson, P., and Lohmander, A. (2014) 'Speech Outcomes at Age 5 and 10 Years in Unilateral Cleft Lip and Palate After One-Stage Palatal Repair with Minimal Incision technique–A Longitudinal Perspective'*. International Journal of Pediatric Otorhinolaryngology* 78 (10), 1662-1670

Ogata, Y., Fujiwara, Y., Suzuki, K., Yamashita, Y., Imai, S., Tezuka.M, Sugiyama, C., Imamura., A., Sato, A., Hayakawa, T., Hasegawa, S., and Mizuto, Y. (2022) 'The development and validation of a Japanese outcome tool for the perceptual assessment of speech in cleft palate patients'. in Mehendale, F. V. (ed.) *Book of Abstracts,* '14th International Congress of Cleft Lip and Palate and Related Craniofacial Anomalies'. held 11-15 July 2022 at Edinburgh International Conference Centre, Edinburgh, UK

Oliveira, A. C., Scarmagnani, R. H., Fukushiro, A. P., and Yamashita, R. P. (eds.) (2016) 'The Influence of Listener Training on the Perceptual Assessment of Hypernasality.' *Codas* 28 (2), 141-148

Papaeliou, C. F. and Rescorla, L. A. (2011) 'Vocabulary Development in Greek Children: A Cross-Linguistic Comparison using the Language Development Survey'*. Journal of Child Language* 38 (4), 861-887

Pereira, V. J., Tuomainen, J., Lee, K. Y., Tong, M. C., and Sell, D. A. (2021) 'A Perceptual Outcome Measure of Velopharyngeal Function Based on the Cleft Audit Protocol for Speech—Augmented (CAPS-A VPC-Sum): Validation through a Speech Osteotomy Study'*. International Journal of Language & Communication Disorders* 56 (4), 754-767

Pereira, V. J., Tuomainen, J., Hay, N., Mars, M., Suchak, A., and Sell, D. A. (2020) 'Effect of Maxillary Osteotomy on Speech in Cleft Lip and Palate: Instrumental Outcomes of Velopharyngeal Function'*. The Cleft Palate-Craniofacial Journal* 57 (11), 1320-1331

Perry, J. L., Kuehn, D. P., Sutton, B. P., and Fang, X. (2017) 'Velopharyngeal Structural and Functional Assessment of Speech in Young Children using Dynamic Magnetic Resonance Imaging'. *The Cleft Palate-Craniofacial Journal* 54 (4), 408-422

Perry, J. L., Sutton, B. P., Kuehn, D. P., and Gamage, J. K. (2014) 'Using MRI for Assessing Velopharyngeal Structures and Function'. *The Cleft Palate-Craniofacial Journal* 51 (4), 476-486

Persson, C., Cooper, R., Conroy, E.J., Gamble., C., Walsh, T., Munro, K., Shaw, W., Willadsen, E. (2022) 'Inter- and intrarater agreement on cross-linguistic perceptual overall rating of velopharyngeal function using the short ordinal scale VPC-Rate'. in Mehendale, F. V. (ed.) *Book of Abstracts,* '14th International Congress of Cleft Lip and Palate and Related Craniofacial Anomalies'. held 11-15 July 2022 at Edinburgh International Conference Centre, Edinburgh, UK

Persson, M., Sandy, J. R., Waylen, A., Wills, A. K., Al-Ghatam, R., Ireland, A. J., Hall, A. J., Hollingworth, W., Jones, T., and Peters, T. J. (2015) 'A Cross-sectional Survey of 5-year-old Children with Non-syndromic Unilateral Cleft Lip and Palate: The Cleft Care UK Study. Part 1: Background and Methodology'. *Orthodontics & Craniofacial Research* 18, 1-13

Persson, C., Lohmander, A., and Elander, A. (2006) 'Speech in Children with an Isolated Cleft Palate: A Longitudinal Perspective'. *The Cleft Palate-Craniofacial Journal* 43 (3), 295-309

Persson, C., Elander, A., Lohmander-Agerskov, A., and Söderpalm, E. (2002) 'Speech Outcomes in Isolated Cleft Palate: Impact of Cleft Extent and Additional Malformations'. *The Cleft Palate-Craniofacial Journal* 39 (4), 397-408

Peryer, H., Slator, R., Thomson, J., and Richard, B. (2021) 'The Method of Surgical Lip Repair Affects Speech Outcomes in Children with Bilateral Cleft Lip and Palate'. *The Cleft Palate-Craniofacial Journal* 58 (4), 419-428

Peters, M. D., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., and Soares, C. B. (2015) 'Guidance for Conducting Systematic Scoping Reviews'. *International Journal of Evidence-Based Healthcare* 13 (3), 141-146

Peterson-Falzone, S.J. (2017) 'Cleft Lip and Palate'. in *The Clinicians Guide to Treating Cleft Palate Speech 2nd Edition.* ed. by Peterson-Falzone, S.J., Trost-Cardamone, J., Karnell, M.P., and Hardin-Jones, M.A. St. Louis: Elsevier, 1-19

Peterson-Falzone, S.J., Hardin-Jones, M.A. and Karnell, M. P. (2009) *Cleft Palate Speech 4th Edition.* St. Louis: Elsevier

Pulkkinen, J., Haapanen, M., Paaso, M., Laitinen, J., and Ranta, R. (2001) 'Velopharyngeal Function from the Age of Three to Eight Years in Cleft Palate Patients'. *Folia Phoniatrica Et Logopaedica* 53 (2), 93-98

Raud Westberg, L., Höglund Santamarta, L., Karlsson, J., Nyberg, J., Neovius, E., and Lohmander, A. (2017) 'Speech Outcome in Young Children Born with Unilateral Cleft Lip and Palate Treated with One-Or Two-Stage Palatal Repair and the Impact of Early Intervention'. *Logopedics Phoniatrics Vocology*, 1-9

RCLST (2022) *Responding to the impact of COVID-19 on the recovery and restoration of children's speech and language therapy services* [online] available from < chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.rcslt.org/wp-content/uploads/2022/03/Responding-to-the-impact-of-COVID-19-on-recovery-and-restoration-of-childrens-SLT-services_FINAL.pdf> [23 November 2022]

RCSLT (2018) *Bercow: 10 years on* [online] available from < chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.rcslt.org/wp-content/uploads/media/Project/RCSLT/bercow-ten-years-on-report.pdf> [23 November 2022]

RCSLT (2009) *RCLST Resource Manual for Commissioning and Planning Services for SLCN* [online] available from <https://www.rcslt.org/speech_and_language_therapy/comissioning/cleft_lip_and_palate_plus_intro> [04 December 2017]

Rescorla, L., Constants, H., Białecka-Pikul, M., Stępień-Nycz, M., and Ochał, A. (2017) 'Polish Vocabulary Development in 2-Year-Olds: Comparisons with English using the Language Development Survey'. *Journal of Speech, Language, and Hearing Research* 60 (4), 1029-1035

Rescorla, L., Frigerio, A., Sali, M. E., Spataro, P., and Longobardi, E. (2014) 'Typical and Delayed Lexical Development in Italian'. *Journal of Speech, Language, and Hearing Research* 57 (5), 1792-1803

Rescorla, L., Lee, Y. M., Kim, Y. A., and Oh, K. Y. (2013) 'Examining Noun Bias in the Lexical Development of Korean Toddlers'. *Journal of Speech, Language, and Hearing Research* 56, 735-747

Rezaei, P., Poorjavad, M., and Abdali, H. (2022) 'Speech Outcomes After Palatal Closure in 3–7-Year-Old Children'. *Brazilian Journal of Otorhinolaryngology* 88, 594-601

Russell, C., Fitzsimons, K., Britton. L., Van Eeden, S., Butterworth, S., Wahedally, H., Park, M.H., Medina, J. and van der Meulen, J. (2022) 'What Is the Optimal Timing of Palate Repair for Speech Outcomes? Analysis of Two Linked National Databases in England'. in Mehendale, F. V. (ed.) *Book of Abstracts,* '14th International Congress of Cleft Lip and Palate

and Related Craniofacial Anomalies'. held 11-15 July 2022 at Edinburgh International Conference Centre, Edinburgh, UK

Russell, J. and Albery, L. (2017) *Practical Intervention for Cleft Palate Speech*. 2nd edn. New York: Routledge

Safaiean, A., Jalilevand, N., Ebrahimipour, M., Asleshirin, E., and Hiradfar, M. (2017) 'Speech Intelligibility After Repair of Cleft Lip and Palate'. *Medical Journal of the Islamic Republic of Iran (MJIRI)* 31 (1), 500-504

Sand, A., Hagberg, E., and Lohmander, A. (2022) 'On the Benefits of Speech-Language Therapy for Individuals Born with Cleft Palate: A Systematic Review and Meta-Analysis of Individual Participant Data'. *Journal of Speech, Language, and Hearing Research* 65 (2), 555-573

Sandy, J., Williams, A., Mildinhall, S., Murphy, T., Bearn, D., Shaw, B., Sell, D., Devlin, B., and Murray, J. (1998) 'The Clinical Standards Advisory Group (CSAG) Cleft Lip and Palate Study'. *British Journal of Orthodontics* 25 (1), 21-30

Santelmann, L., Sussman, J., and Chapman, K. (1999) 'Perception of Middorsum Palatal Stops from the Speech of Three Children with Repaired Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 36 (3), 233-242

Scherer, N., Mathad, V., Cordero, K., Lien, K., Williams,J., Chapman.K., Liss, J. and Berisha, V. (2022) 'Prospective validation of AI-based algorithms for assessment of hypernasality and speech acceptability in children with cleft lip and palate'. in Mehendale, F. V. (ed.) *Book of Abstracts,* '14th International Congress of Cleft Lip and Palate and Related Craniofacial Anomalies'. held 11-15 July 2022 at Edinburgh International Conference Centre, Edinburgh, UK

Scherer, N. J., D'Antonio, L. L., and McGahey, H. (2008) 'Early Intervention for Speech Impairment in Children with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 45 (1), 18-31

Scherer, N. J. (1999) 'The Speech and Language Status of Toddlers with Cleft Lip and/Or Palate Following Early Vocabulary Intervention'. *American Journal of Speech-Language Pathology* 8 (1), 81-93

Schober, P., Boer, C., and Schwarte, L. A. (2018) 'Correlation Coefficients: Appropriate use and Interpretation'. *Anesthesia & Analgesia* 126 (5), 1763-1768

Schölin, J. S., Jonasson, Å, Axelsson, J., Havstam, C., Persson, C., Jönsson, R., and Mark, H. (2020) 'Surgical, Speech, and Hearing Outcomes at Five Years of Age in Internationally Adopted Children and Swedish-Born Children with Cleft Lip and/Or Palate'. *Journal of Plastic Surgery and Hand Surgery* 54 (1), 6-13

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., and Skoruppa, K. (2021) 'Using Nonword Repetition to Identify Developmental Language Disorder in

Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis.'. *Journal of Speech, Language, and Hearing Research* 64 (9), 3578-3593

Seifert, M., Davies, A., Harding, S., McLeod, S., and Wren, Y. (2021) 'Intelligibility in 3-Year-Olds with Cleft Lip and/Or Palate using the Intelligibility in Context Scale: Findings from the Cleft Collective Cohort Study'. *The Cleft Palate-Craniofacial Journal* 58 (9), 1178-1189

Sell, D. and Sweeney, T. (2020) 'Percent Consonant Correct as an Outcome Measure for Cleft Speech in an Intervention Study'. *Folia Phoniatrica Et Logopaedica* 72 (2), 143-151

Sell, D. and Harding-Bell, A. (2019) 'Speech and velopharyngeal dysfunction'. in *Case Studies in Cleft Palate Speech, Data analysis and principled intervention.* ed. Harding-Bell, A. Croydon: J&R Press, 111-156

Sell, D., Southby, L., Wren, Y., Wills, A. K., Hall, A., Mahmoud, O., Waylen, A., Sandy, J. R., and Ness, A. R. (2017) 'Centre-level Variation in Speech Outcome and Interventions, and Factors Associated with Poor Speech Outcomes in 5-year-old Children with Non-syndromic Unilateral Cleft Lip and Palate: The Cleft Care UK Study. Part 4'. *Orthodontics & Craniofacial Research* 20, 27-39

Sell, D. and Pereira, V. (2015) 'Speech: Perceptual and Functional Outcomes'.in *Comprehensive Cleft Care.* 2nd edn. ed. by Loose, J. E., and Kirschner. R. London: McGraw Hill, 689-713

Sell, D., Mildinhall, S., Albery, L., Wills, A. K., Sandy, J. R., and Ness, A. R. (2015) 'The Cleft Care UK Study. Part 4: Perceptual Speech Outcomes'. *Orthodontics & Craniofacial Research* 18 (S2), 36-46

Sell, D., John, A., Harding-Bell, A., Sweeney, T., Hegarty, F., and Freeman, J. (2009) 'Cleft Audit Protocol for Speech (CAPS-A): A Comprehensive Training Package for Speech Analysis'. *International Journal of Language & Communication Disorders* 44 (4), 529-548

Sell, D. (2005) 'Issues in Perceptual Speech Analysis in Cleft Palate and Related Disorders: A Review'. *International Journal of Language & Communication Disorders* 40 (2), 103-121

Sell, D., Grunwell, P., Mildinhall, S., Murphy, T., Cornish, T. A., Bearn, D., Shaw, W. C., Murray, J. J., Williams, A. C., and Sandy, J. R. (2001) 'Cleft Lip and Palate Care in the United Kingdom—the Clinical Standards Advisory Group (CSAG) Study. Part 3: Speech Outcomes'. *The Cleft Palate-Craniofacial Journal* 38 (1), 30-37

Sell, D., Harding, A., and Grunwell, P. (1999) 'GOS. SP. ASS.'98: An Assessment for Speech Disorders Associated with Cleft Palate and/Or Velopharyngeal Dysfunction (Revised)'. *International Journal of Language & Communication Disorders* 34 (1), 17-33

Sell, D. A. and Grunwell, P. (1990) 'Speech Results Following Late Palatal Surgery in Previously Unoperated Sri Lankan Adolescents with Cleft Palate'. *Cleft Palate Journal* 27 (2), 162-168

Sharma, A. and Cockerill, H. (2014) *From Birth to Five Years: Practical Developmental Examination*. London: Routledge

Shaw, W., Semb, G., Lohmander, A., Persson, C., Willadsen, E., Clayton-Smith, J., Trindade, I. K., Munro, K. J., Gamble, C., and Harman, N. (2019) 'Timing of Primary Surgery for Cleft Palate (TOPS): Protocol for a Randomised Trial of Palate Surgery at 6 Months Versus 12 Months of Age'. *BMJ Open* 9 (7), e029780

Shaw, W. C., Dahl, E., Asher-McDade, C., Brattström, V., Mars, M., McWilliam, J., Mølsted, K., Plint, D. A., Prahl-Andersen, B., and Roberts, C. (1992) 'A Six-Center International Study of Treatment Outcome in Patients with Clefts of the Lip and Palate: Part 5. General Discussion and Conclusions'. *The Cleft Palate-Craniofacial Journal* 29 (5), 413-418

Shriberg, L. D., McSweeny, J. L., Anderson, B. E., Campbell, T. F., Chial, M. R., Green, J. R., Hauner, K. K., Moore, C. A., Rusiewicz, H. L., and Wilson, D. L. (2005) 'Transitioning from Analog to Digital Audio Recording in Childhood Speech Sound Disorders'. *Clinical Linguistics & Phonetics* 19 (4), 335-359

Shriberg, L. D. and Lof, G. L. (1991) 'Reliability Studies in Broad and Narrow Phonetic Transcription'. *Clinical Linguistics & Phonetics* 5 (3), 225-279

Solot, C. B., Sell, D., Mayne, A., Baylis, A. L., Persson, C., Jackson, O., and McDonald-McGinn, D. M. (2019) 'Speech-Language Disorders in 22q11. 2 Deletion Syndrome: Best Practices for Diagnosis and Management'. *American Journal of Speech-Language Pathology* 28 (3), 984-999

Sommerlad, B. C., Fenn, C., Harland, K., Sell, D., Birch, M. J., Dave, R., Lees, M., and Barnett, A. (2004) 'Submucous Cleft Palate: A Grading System and Review of 40 Consecutive Submucous Cleft Palate Repairs'. *The Cleft Palate-Craniofacial Journal* 41 (2), 114-123

Speltz, M. L., Endriga, M. C., Hill, S., Maris, C. L., Jones, K., and Omnell, M. L. (2000) 'Cognitive and Psychomotor Development of Infants with Orofacial Clefts'. *Journal of Pediatric Psychology* 25 (3), 185-190

Stein, M. J., Zhang, Z., Fell, M., Mercer, N., and Malic, C. (2019) 'Determining Postoperative Outcomes After Cleft Palate Repair: A Systematic Review and Meta-Analysis'. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 72 (1), 85-91

Stevens, N. and Isles, D. (2011) *Phonological Screening Assessment (PSA).* London: Taylor & Francis

Stevens, S. S. (1975) 'Laws that Govern Behavior.(Book Reviews: Psychophysics. Introduction to its Perceptual, Neural, and Social Prospects)'. *Science* 188, 827-829

Stoel-Gammon. C. (2015) 'Assessment of the speech of an unintelligible child'. in *Children's Speech Sound Disorders.* ed. by Bowen, C. Oxford: Wiley Blackwell, 84-88

Sugden, E. and Cleland, J. (2021) 'Using Ultrasound Tongue Imaging to Support the Phonetic Transcription of Childhood Speech Sound Disorders'. *Clinical Linguistics & Phonetics*, 1-20

Swanson, J. W., Mitchell, B. T., Cohen, M., Solot, C., Jackson, O., Low, D., Bartlett, S. P., and Taylor, J. A. (2017) 'The Effect of Furlow Palatoplasty Timing on Speech Outcomes in Submucous Cleft Palate.'. *Annals of Plastic Surgery* 79 (2), 156

Sweeney, T., Hegarty, F., Powell, K., Deasy, L., Regan, M. O., and Sell, D. (2020) 'Randomized Controlled Trial Comparing Parent Led Therapist Supervised Articulation Therapy (PLAT) with Routine Intervention for Children with Speech Disorders Associated with Cleft Palate'. *International Journal of Language & Communication Disorders* 55 (5), 639-660

Sweeney, T. (2011) 'Nasality- Assessment and Intervention'. in *Cleft Palate Speech, Assessment and Intervention.* ed. by Howard, S. and Lohmander, A. Oxford: Wiley-Blackwell, 199-220

Sweeney, T. and Sell, D. (2008) 'Relationship between Perceptual Ratings of Nasality and Nasometry in Children/Adolescents with Cleft Palate and/Or Velopharyngeal Dysfunction'. *International Journal of Language & Communication Disorders* 43 (3), 265-282

Tardif, T., Gelman, S. A., and Xu, F. (1999) 'Putting the "noun Bias" in Context: A Comparison of English and Mandarin'. *Child Development* 70 (3), 620-635

Tillman, K. K., Hakelius, M., Höijer, J., Ramklint, M., Ekselius, L., Nowinski, D., and Papadopoulos, F. C. (2018) 'Increased Risk for Neurodevelopmental Disorders in Children with Orofacial Clefts'. *Journal of the American Academy of Child & Adolescent Psychiatry* 57 (11), 876-883

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Kastner, M., Levac, D., Ng, C., Sharpe, J. P., and Wilson, K. (2016) 'A Scoping Review on the Conduct and Reporting of Scoping Reviews'. *BMC Medical Research Methodology* 16 (1), 15

Vallino, L. D., Zuker, R., and Napoli, J. A. (2008) 'A Study of Speech, Language, Hearing, and Dentition in Children with Cleft Lip Only'. *The Cleft Palate-Craniofacial Journal* 45 (5), 485-494

van Eeden, S. and Stringer, H. (2020) 'Linguistic and Auditory Processing Skills in Non-Syndromic Children with Cleft Palate: A Scoping Review'. *Journal of Communication Disorders* 87, 106029

World Health Organisation. (2007) *International Classification of Functioning, Disability and Health. Children and Youth Version* (Geneva: WHO)

Whitehill, T., Gotzke, C.L. and Hodge, M. (2011) 'Speech Intelligibility'. in *Cleft Palate Speech, Assessment and Intervention.* ed. by Howard, S. and Lohmander, A. Oxford: Wiley-Blackwell, 293-304

Whitehill, T. L. (2002) 'Assessing Intelligibility in Speakers with Cleft Palate: A Critical Review of the Literature'. *The Cleft Palate-Craniofacial Journal* 39 (1), 50-58

Willadsen, E., Persson, C., Patrick, K., Lohmander, A., and Oller, D. K. (2020) 'Assessment of Prelinguistic Vocalizations in Real Time: A Comparison with Phonetic Transcription and Assessment of Inter-Coder-Reliability'. *Clinical Linguistics & Phonetics* 34 (7), 593-616

Willadsen, E., Boers, M., Schöps, A., Kisling-Møller, M., Nielsen, J. B., Jørgensen, L. D., Andersen, M., Bolund, S., and Andersen, H. S. (2018) 'Influence of Timing of Delayed Hard Palate Closure on Articulation Skills in 3-year-old Danish Children with Unilateral Cleft Lip and Palate'. *International Journal of Language & Communication Disorders* 53 (1), 130-143

Willadsen, E., Lohmander, A., Persson, C., Lundeborg, I., Alaluusua, S., Aukner, R., Bau, A., Boers, M., Bowden, M., and Davies, J. (2017) 'Scandcleft Randomised Trials of Primary Surgery for Unilateral Cleft Lip and Palate: 5. Speech Outcomes in 5-Year-Olds-Consonant Proficiency and Errors.'. *Journal of Plastic Surgery and Hand Surgery* 51 (1), 38

Willadsen, E. (2012) 'Influence of Timing of Hard Palate Repair in a Two-Stage Procedure on Early Speech Development in Danish Children with Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 49 (5), 574-595

Willadsen, E. and Poulsen, M. (2012) 'A Restricted Test of Single-Word Intelligibility in 3-Year-Old Children with and without Cleft Palate'. *The Cleft Palate-Craniofacial Journal* 49 (3), 6-16

Williams, C., Harding, S., and Wren, Y. (2021) 'An Exploratory Study of Speech and Language Therapy Intervention for Children Born with Cleft Palate±lip'. *The Cleft Palate-Craniofacial Journal* 58 (4), 455-469

Wren, Y., Humphries, K., Stock, N. M., Rumsey, N., Lewis, S., Davies, A., Bennett, R., and Sandy, J. (2018) 'Setting Up a Cohort Study in Speech and Language Therapy: Lessons from the UK Cleft Collective Speech and Language (CC-SL) Study'. *International Journal of Language & Communication Disorders* 53 (3), 421-430

Wren, Y., (2013) 'A survey of the assessment procedures at 24 and 36 months in UK cleft centres: The Cleft Collective Speech and Language Study' (unpublished survey).

Yamamoto, I. (2020) 'Tongue-Palate Contact Patterns for Japanese Speakers with and without Cleft Lip and Palate'. *International Journal of Speech-Language Pathology* 22 (1), 70-77

Yamashita, R. P., Borg, E., Granqvist, S., and Lohmander, A. (2018) 'Reliability of Hypernasality Rating: Comparison of 3 Different Methods for Perceptual Assessment'. *The Cleft Palate-Craniofacial Journal*, 1055665618767116

Yang, A. S., Richard, B. M., Wills, A. K., Mahmoud, O., Sandy, J. R., and Ness, A. R. (2020) 'Closer to the Truth on National Fistula Prevalence After Unilateral Complete Cleft Lip and Palate Repair? the Cleft Care UK Study'. *The Cleft Palate-Craniofacial Journal* 57 (1), 5-13

Yogendran, N., Lohamnder, L., Sell, D., Tsanas, A., Mamedov, T. and Mehendale, F. (2022) 'Machine Learning-based Classification of Cleft Speech: Cleft-Associated Resonance Imitation Study (CLARIS)'. in Mehendale, F. V. (ed.) *Book of Abstracts,* '14th International Congress of Cleft Lip and Palate and Related Craniofacial Anomalies'. held 11-15 July 2022 at Edinburgh International Conference Centre, Edinburgh, UK

Zajac, D. J., Vallino, L. D., Baylis, A. L., Adatorwovor, R., Preisser, J. S., and Vivaldi, D. (2021) 'Emergence of Prevocalic Stop Consonants in Children with Repaired Cleft Palate'. *Journal of Speech, Language, and Hearing Research* 64 (1), 30-39

Zanzi, M., Cherpillod, J., and Hohlfeld, J. (2002) 'Phonetic and Otological Results After Early Palate Closure in 18 Consecutive Children Presenting with Cleft Lip and Palate'. *International Journal of Pediatric Otorhinolaryngology* 66 (2), 131-137

Zraick, R. I. and Liss, J. M. (2000) 'A Comparison of Equal-Appearing Interval Scaling and Direct Magnitude Estimation of Nasal Voice Quality'. *Journal of Speech, Language, and Hearing Research* 43 (4), 979-988

Zraick, R. I., Liss, J. M., Dorman, M. F., Case, J. L., LaPointe, L. L., and Beals, S. P. (2000) 'Multidimensional Scaling of Nasal Voice Quality'. *Journal of Speech, Language, and Hearing Research* 43 (4), 989-996

Appendix D: Coventry University Ethical Approval Certificate for Phase 1: Scoping Review (P68435)

Certificate of Ethical Approval

Applicant:

Elizabeth Fitzpatrick

Project Title:

FULL/LONG TITLE OF THE STUDY
The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip

SHORT STUDY TITLE / ACRONYM
The EASO Study

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

23 February 2018

Project Reference Number: P68435

Appendix E: Coventry University Ethics Application Form for Phase 1: Scoping Review (P68435)

Full name: Elizabeth Fitzpatrick

Faculty/Subsidiary/Area: Faculty of Health and Life Sciences

School/Institute/Unit: School of Nursing, Midwifery and Health

Supervisor: Tanya Rihtman

Module nameD005RDC - Application of Research Methods

Project Summary

Project IDP68435

Project title: FULL/LONG TITLE OF THE STUDY The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip SHORT STUDY TITLE / ACRONYM The EASO Study

Module codeD005RDC - Application of Research Methods

Brief Project Summary

This ethics application refers to the first objective within a PhD study; subsequent objectives will require CU ethics and NHS IRAS approval. Following surgical repair children with cleft palate +/- cleft lip (CPL) remain an 'at risk' group for speech difficulties. Whilst there is a standardised UK protocol to assess cleft speech at age 5 years, there is no such protocol at age 3 years in the UK. It is vital that speech difficulties are identified in the pre-school years to minimise the impact of these difficulties on education outcomes and self-esteem. Identifying speech difficulties at age 3 may result in more timely therapy and surgical intervention. The PhD study aims to propose an assessment framework for CPL speech outcomes at age 3. An initial scoping review will inform subsequent objectives of the PhD study, specifically, the parameters of speech to be assessed, and the types of speech samples to be included in the speech assessment. This will be achieved by reviewing available literature and current monitoring practices for 3-year-old children with CPL in the UK (the assessments, speech samples, speech parameters, methods and rating scales currently in use).

Start and end dates01 Oct 2017 - 01 Sep 2020

Is this project externally funded?

Yes

Who is funding the project?Studentship via HLS

Has the funding been confirmed?

Yes

Use Professional Code of Ethical Practice?

Yes

Name of Professional Code of Ethical PracticeYes

Have you read the Code?

Yes

Project Detail

What are the aims and objectives of the project?

Overall PhD Study Aim: To propose an assessment framework to validly and reliably assess speech outcomes in three-year-old patients with CPL through the examination of different speech samples and rating methods and the acceptability and usability of the assessment to Speech and Language Therapists.

The objective which relates to this ethics application: To undertake a scoping and identification exercise, to inform the parameters of speech that should be assessed in three-year-old children with CPL, and the types of speech samples which are included in the speech assessment

This will be achieved by:

- reviewing the available literature

- reviewing current monitoring practices for 3-year-old children with CPL in the UK (assessments currently in use, speech samples currently in use, speech parameters currently in use, methods and rating scales currently in use).

Explain your research design

A scoping review will be undertaken, using the framework proposed by The Joanna Briggs Institute (2015) as a guide.

Reference:

Joanna Briggs Institute (2015) The Joanna Briggs Institute Reviewers Manual 2015: Methodology for JBI Scoping Reviews [online] available from [https://joannabriggs.org/assets/docs/sumari/Reviewers-Manual_Methodology-for-JBI-Scoping-Reviews_2015_v2.pdf] [20 February 2018]

Explain your research design and outline the principal method(s) you will use

The scoping review will be guided by the framework proposed by The Joanna Briggs Institute (2015).

Scoping review question: What are the parameters of speech that should be assessed in three-year-old children with CPL, and the types of speech samples that should be included in the speech assessment?

Inclusion Criteria- Population, Concept and Context (PCC)

Population: 3-year-old children with CPL, children with CPL, 3-year-old children without CPL

Concept: Speech assessments, and specifically the speech samples, methods of assessment, rating scales and speech parameters assessed.

Context: Literature and resources published within the past 20 years. Seminal works pre-dating this timeframe will also be included.

Types of sources:

• Database searches: relevant databases e.g. Medline, AMED, Embase using keyword searching. Articles available in English will be reviewed.

Keywords: Preschool, toddler, kindergarten, nursery, 3 years old, aged 3 years, 3 years of age, cleft palate, cleft lip and palate, articulation, cleft speech characteristics, cleft type characteristics, compensatory articulation, velopharyngeal dysfunction, velopharyngeal insufficiency, speech articulation, speech analysis, speech assessment, speech and language assessment.

• Citation tracking: the reference list of the identified articles will be reviewed to identify any additional studies.

• Information available from Cleft Centres: any information which has been previously been gathered about practices in assessment at different cleft centres and is available to the public or to the researcher will be utilised.

• Website searching to identify commercially available assessment products suitable to assess speech at age 3 years in the CPL population.

Charting the results: This may be refined in the review stage but will include the following:

Author, Year of publication, origin, study aim/objective, study population and sample, methodology and methods, speech assessment utilised, speech samples utilised, parameters of speech assessed, methods of assessment, rating methods used (e.g. VAS, ordinal).

Results: The PRISMA (2009) Flow Diagram will be used to outline the studies which were identified and included in the scoping review. Information gained during the review will be presented in a table formation (Appendix XX). The main results may be classified according to different areas under review e.g. speech samples, parameters assessed.

Reference:

PRISMA (2009) Flow Diagram [online] available from [http://www.prisma-statement.org/documents/PRISMA%202009%20flow%20diagram.pdf] [20 February 2018]

Are you proposing to use a validated scale or published research method / tool?

Yes

External Research Instrument

The scoping review will be guided by the scoping framework produced The Joanna Briggs Institute (2015). In addition, the PRISMA (2009) Flow Diagram will be used to outline article selection.

Data Analysis

1

Does the research seek to understand, identify, analyse and/or report on data/information on terrorism/terrorism policies?

No

2Does your research seek to understand, identify, analyse and/or report on information for other activities considered illegal in the UK and/or in the country you are researching in?

-

3Are you analysing Secondary Data?

Yes

Could an individual be identified from the data? e.g. identifiable datasets where the data has not been anonymised or there is risk of re-identifying an individual

N/A-

4Are you dealing with Primary Data involving people?

No

5Personal or Sensitive data

Are you dealing with personal data?

-N/A

Are you dealing with special category data (formerly known as sensitive data)?

No

6Is the project solely desk based secondary research?

Yes

Appendix F: Feedback from the Cleft and Craniofacial Studies Group

Feedback	Response/Changes
It would be useful for the views of parents and/or children to be built into the study so they have a voice.	The West Midlands Patient Voices Cleft Lip and Palate Association Group, comprised of teenage/adult service users and parents, reviewed all of the speech assessment materials used and provided feedback.
I don't know how attractive the study will be for SLTs to participate in- listening takes up a great deal of time, which may be free time of the recruited listeners,	Whilst the ideal recruitment target was 11 cleft SLTs, one from each NHS cleft team, consideration was given that not all teams may be able to participate. As such, the minimum sample was changed to 6 cleft SLTs which may include SLTs at the WMCLPS.
The sample size of 2-5 individuals in the control group seems small, particularly given the normal expected variation in speech and language skills at age 3-years.	The target sample size remained between 2-5 participants in the control group. However, further information was provided in the protocol to detail that the specific role of the control group is to contribute to specificity i.e. that these control group participants are not identified as having speech characteristics associated with CP±L. As such, any normal variation in the speech and language skills of 3-year-olds would not be a concern and would be anticipated at this age. The inclusion of the control group was to investigate if such normal variation in the speech and language skills of 3-year-olds was misidentified as cleft specific speech characteristics.
Would it be better to just recruit children with CP and exclude children with cleft lip? Repaired cleft lips will be immediately obvious and introduce assessment bias.	This was considered as the inclusion of children with cleft lip would mean that the listeners were not fully blinded as to whether the participant was in the cleft or control group. Ultimately the decision was made to include children with cleft lip for the following reasons: - A study objective is to propose an assessment framework for 3-year-olds with CP±L. Limiting the types of cleft types in the study could significantly limit the generalisability of the results and the usability of any proposed assessment framework.

	- Children with CP only had been previously recruited to the ToPS trial (Shaw et al. 2019), and participation in a second study may be an increased burden for these patients and families which in turn could limit recruitment. - Children had to be recruited in waiting list order with no prior speech 'screening'. Excluding children with UCLP and BCLP cleft types, who may have the most significant speech needs (CRANE, 2021) may have limited the range of speech outcomes in the study. This would have limited the validity of the study and may have impacted reliability outcomes. - Limiting the study to children with CP only would also have reduced the children available for recruitment by a third.
Would English as a second language impact speech development or ability to participate in the study or affect how well the speech assessment might work for children where English is not the first language at home?	- It was important to include children who speak English as a second language given evidence that over 20% of primary school children speak English as an additional language. - At the WMCLPS we have a significant proportion of patients who speak English as a second language, it was therefore important that their families were given the same opportunity to participate in the research study. - However, to meet the needs of children who come from a non-English speaking family, it was more appropriate to assess their speech in their first language with the aid of an interpreter. As such the decision was made to exclude children from non-English speaking families.

Appendix G: Coventry University Ethics Certificate for Phase 2 of the study (P66325)

Certificate of Ethical Approval

Applicant:

Elizabeth Fitzpatrick

Project Title:

FULL/LONG TITLE OF THE STUDY
The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip-

SHORT STUDY TITLE / ACRONYM
The EASO Study

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as High Risk

Date of approval:

18 June 2018

Project Reference Number: P66325

Appendix H: Coventry University Ethics Application Form for (P66325)

Full name: Elizabeth Fitzpatrick

Faculty/Subsidiary/Area: Faculty of Health and Life Sciences

School/Institute/Unit School of Nursing, Midwifery and Health

Supervisor: Tanya Rihtman

Module nameD005RDC - Application of Research Methods

Project Summary

Project IDP66325

Project title: FULL/LONG TITLE OF THE STUDY The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip- SHORT STUDY TITLE / ACRONYM The EASO Study

Module codeD005RDC - Application of Research Methods

Brief Project Summary

Following surgical repair children with cleft palate +/- cleft lip (CPL) remain an 'at risk' group for speech difficulties. Whilst there is a standardised UK protocol to assess cleft speech at age 5 years, there is no such protocol at age 3 years in the UK. It is vital that speech difficulties are identified in the pre-school years to minimise the impact of these difficulties on education outcomes and self-esteem. Identifying speech difficulties at age 3 may result in more timely therapy and surgical intervention. This study aims to propose an assessment framework to assess speech at age 3 in children with CPL. This will be achieved by investigating, in 3-year-old children with CPL, which speech samples, which methods of assessing velopharyngeal function for speech and which rating scales result in the most valid and reliable listener judgements. 3-year-old children with CPL and without (Control Group) will produce different speech samples which will be analysed by Speech and Language Therapists (SLTs) using different methods and rating scales. The acceptability of the different speech samples, methods and rating scales to SLTs will also inform the proposed assessment framework.

Names of Co-Investigators & their organisational affiliation (place of study/employer)

-

Is this project externally funded?

Yes

Who is funding the project?HLS PhD Studentship

Has the funding been confirmed?

Yes

Use Professional Code of Ethical Practice?

Yes

Name of Professional Code of Ethical PracticeYes

Have you read the Code?

Yes

Project Detail

What are the aims and objectives of the project?

AIM: To propose an assessment framework to validly and reliably assess speech outcomes in three-year-old patients with CPL through the examination of different speech samples, rating methods and scales, and the acceptability and usability of the assessment to SLTs.

See Appendix A for Study Objectives.

Explain your research design

The study aim aligns with a quantitative methodology (Pope & Mays, 1995). The aims and objectives are also congruent with a positivist research paradigm.

There is no single study design which encompasses the entire study. The study will be undertaken in various phases using different study designs relating to study objectives.

1) Scoping review: to inform the parameters of speech and speech samples which should be included in the speech assessment.

CU ethical approval for the Scoping review has already been gained. Project Number: P68435. Date of Approval: 23.02.18

2) Prospective parallel randomised cross-over group study: Three-year-old participants will be randomised in two groups to the order in which they produce speech samples.

3) Reliability and Validity Study: The speech samples will be analysed by a group of SLTs acting as listeners. Listener judgments will then be analysed to examine inter and intra-reliability. Content, construct validity and specificity will be examined.

4) Acceptability Study: SLTs will complete a questionnaire to examine how acceptable and usable the assessment tool was.

Explain your research design and outline the principal method(s) you will use

1.Location: SLT Department at Birmingham Children's Hospital

2.Participants

i-3-year-old children with CPL: to be recruited from West Midlands Cleft Lip and Palate Service (WMCLPS). Patient Information Leaflet (PIL) (Appendix XX) to be sent with the standard 3-year speech and language therapy appointment letter to the parent(s)/guardian(s) of children who meet inclusion criteria (section 4.4.1.3 of the attached protocol). The Chief Investigator (CI) will telephone the parent(s)/guardian(s) and use a pre-determined script (Appendix XX) to ascertain interest in participation in the study. If in agreement, informed consent will be gained at the start of the child's appointment (Appendix XX), otherwise, the child will be seen for the standard 3-year assessment and will not participate in the study.

ii-Control Group: 3-year-old children without CPL who meet the inclusion criteria (section 4.4.2.3 of the attached protocol) to be recruited as a control group. Staff members of the Therapies Department (OT, Physiotherapy and SLT) and WMCLPS at Birmingham Children's Hospital will act as gatekeepers (Appendix XX) and will pass the relevant PIL (Appendix XX) to acquaintances. Interested parent(s)/guardian(s) will contact CI if in agreement for their child to take part; CI will check the inclusion criteria. An assessment session will be arranged with the CI and informed consent will be gained at the start of the appointment (Appendix XX).

iii- Listener SLTs: SLTs who meet the inclusion criteria (section 4.4.3.3 of the attached protocol) will be recruited to act as expert listeners. The Cleft Clinical Excellence Network will act as gatekeepers (Appendix XX) and will pass the relevant PIL (Appendix XX) to Cleft SLTs who will be asked to contact the CI for further information. CI will check that the SLT meets the inclusion criteria. Informed consent will be gained via email return of the consent form (Appendix XX).

3.Procedure:

•During the assessment session, 3-year-old participants will complete different speech samples i.e. spontaneous speech, picture naming, sentence repetition (see section 4.5 of the attached protocol)which will be video recorded.

• The assessment session should not last longer than an hour and participants will be able to take breaks as required.

• The extent to which the participants fully complete the speech sample will be recorded by the CI (Appendix XX).

•There is no further direct involvement of the 3-year-old participants in the study.

•Listener SLTs will access the participants' recordings from their place of work using the NHS approved SharePoint External Data Exchange System and will carry out three listening sessions as per the listening process (Appendix XX). The SLTs will input their judgements using Qualtrics. Listening Session 1 will include judgements on the parameters: overall velopharyngeal (VP) function Hypernasality, Hyponasality, Audible Nasal Emission/Turbulence, consonant production and phonology (Henningsson et al. 2008), scored on an ordinal scale. VP function for speech will be scored ordinally using two methods (Hypernasality, Hyponasality, Audible Nasal Emission/Turbulence [1] and overall measure [2]). Session 2 will include the same judgements as session 1. In session 3, only hypernasality and the overall measure of VP for speech using a Visual Analogue Scale (VAS) will be scored.

•Data will be analysed as outlined in the attached protocol (section 4.7).

•The parameters of speech judged by the Cleft SLTs will be compared to the SLT who carried out the original assessment (Appendix XX).

•The Cleft SLTs will complete a questionnaire about the acceptability of the listening process and the methods used (Draft version Appendix XX).

•An assessment framework for three-year-old children with CPL will be proposed, utilising the information gained through the analysis of listener reliability, feedback from the SLTs and considering the speech samples completed most easily.

Are you proposing to use a validated scale or published research method / tool?

Yes

External Research Instrument

Universally Speaking (The Communication Trust 2015: 14-15) (Appendix XX), will be used as an informal screen to determine whether the Control Group participant meets the inclusion criteria.

To generate the speech samples, existing speech assessment materials will be used e.g. GOS.SP.ASS (Sell et al. 1999) and The Restricted Word List (Lohmander et al. 2009). These assessment materials and selected speech samples reflect those samples utilised most frequently in the literature and those used most frequently by Cleft SLTs in the UK (Wren ongoing). Where additional pictorial stimuli are required images utilised will be in the public domain or images produced specifically for the study.

The parameters of speech that will be assessed are based upon the findings of Objective 1 (protocol section 4.5). Ordinal scale descriptors are based upon the CAPS-A rating tool used at age 5 years (John et al. 2006) (Appendix XX) which is familiar to the Listener SLTs. The overall measure of velopharyngeal function for speech will be based upon that used by Lohmander et al. (2017).

Cleft SLTs will be asked to complete a questionnaire about the listening process and rating methods (draft, Appendix XX).

Data Analysis

1

Does the research seek to understand, identify, analyse and/or report on data/information on terrorism/terrorism policies?

No

2Does your research seek to understand, identify, analyse and/or report on information for other activities considered illegal in the UK and/or in the country you are researching in?

-

3Are you analysing Secondary Data?

No

4Are you dealing with Primary Data involving people?

Yes

5Personal or Sensitive data

Are you dealing with personal data?

-

Are you dealing with special category data (formerly known as sensitive data)?

Yes

Please specify what special category data you will be collecting

-

Will the Personal or Special Category data be shared with a third party?

No

Will the Personal or Sensitive data be shared outside of the European Economic Area ('EEA')?

No

6Is the project solely desk based secondary research?

No

7Will the data collection, recruitment materials or any other project documents be in any language other than English?

-

Areas of Study

1You have indicated the following are relevant to your study

Travel away from home campus

Biological samples, physical measurements/activities or substances/procedures administered to or taken from human participants

Interaction(s) with human participants

Biological materials including organisms, cell lines and samples (excluding humans)

Animals and their habitats, animal materials such as blood, tissue, or stool samples, primary cell cultures derived from animal tissues (excluding humans)

Hazardous substances

Machinery & equipment

Ionising Radiation

Flying unmanned aerial systems (e.g. drones)

Environmental samples

Appendix I: NHS Research Ethics Committee Approval

Health Research Authority

East Midlands - Nottingham 1 Research Ethics Committee

The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

15 January 2019

Miss Elizabeth (Beth) Fitzpatrick
Senior Specialist Speech and Language Therapist, PhD Student
Birmingham Women's and Children's NHS Trust/Coventry University
Speech and Language Therapy Department
Birmingham Children's Hospital
Steelhouse Lane, Birmingham
B4 6NH

Dear Miss Fitzpatrick

Study title:	**The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip**
REC reference:	**18/EM/0253**
Protocol number:	**N/A**
IRAS project ID:	**242296**

Thank you for your letter of 21 December 2018, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to make a request to postpone publication, please contact hra.studyregistration@nhs.net outlining the reasons for your request.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System, at www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database within 6 weeks of recruitment of the first participant (for medical device studies, within the timeline determined by the current registration and publication trees).

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS sites

The Committee has not yet completed any site-specific assessment (SSA) for the non-NHS research site(s) taking part in this study. The favourable opinion does not therefore apply to

any non-NHS site at present. We will write to you again as soon as an SSA application(s) has been reviewed. In the meantime no study procedures should be initiated at non-NHS sites.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Covering letter on headed paper [CoveringLetter]		02 May 2018
IRAS Application Form [IRAS_Form_30072018]		30 July 2018
Letter from sponsor [Sponsor Letter]		18 June 2018
Letters of invitation to participant [Checklist Q Invitation to SLT Participants via Gatekeeper]	3.0	26 April 2018
Other [Checklist L PIS for Control Group]	2.0	30 May 2018
Other [Checklist M The Communication Trust Universally Speaking]		27 March 2018
Other [Checklist S SLT Consent Form]	2.11	11 May 2018
Other [Sponsor Professional Indemnity Letter]		31 July 2017
Other [Coventry University Ethics Certificate]		18 June 2018
Other [Supervisor 3 CV]		18 June 2018
Other [CV for Supervisor 2 Jane Coad]		21 June 2018
Other [Checklist W CSG Feedback]	1.0	18 June 2018
Other [Checklist X CSG Changes]	1.0	28 June 2018
Other [Checklist C Participant Information Sheet Cleft Group]	3.4	26 October 2018
Other [Checklist E Cleft Group Consent Form]	2.3	26 October 2018
Other [Checklist K Email to gatekeepers for Control Group Participants]	2.0	09 October 2018
Other [Checklist N Consent Form Control Group]	2.3	26 October 2018
Other [Checklist R Participant Information Sheet SLT Group]	3.4	26 October 2018
Other [GCP 2018 Certificate]		12 September 2018
Other [Paediatric Consent Certificate]		12 September 2018
Other [Insurance Certificate 1]		15 August 2018
Other [Insurance Certificate 2]		01 August 2018
Other [Checklist B: Aims and Objectives]	1.0	18 June 2018
Other [Checklist D: Telephone script for Control Group]	2.0	26 March 2018
Other [Checklist G: PArticpants with CPL Identification Log]	1.1	26 October 2018
Other [Video/Audio Recording Request Form]		26 October 2018
Other [Checklist F: Screening and Enrolment Log CPL]	1.1	26 October 2018
Other [Assessment Recording Form]	1.1	26 October 2018
Other [Checklist J: Participant Video Recording Numbers]	1.1	26 October 2018
Other [Checklist O: Control Group Screening Enrolment LOg]	1.1	26 October 2018
Other [Checklist P: Control Group Identification Log]	1.3	26 October 2018

Other [Gatekeeper letter to SLTs]	3.1	26 October 2018
Other [Checklist T: SLT Questionnaire]	2.2	26 October 2018
Other [Checklist U: Universal Parameters Recording Form]	1.2	26 October 2018
Other [Checklist V: CPL Group Letter to Parents, Guardians]	1.2	26 October 2018
Other [Draft Flowchart of Listening Process]	2.1	26 October 2018
Other [Draft Flowchart of Control Group Involvement]	1.3	26 October 2018
Other [Draft Flowchart of CPL Group Involvement]	1.3	26 October 2018
Other [Draft Flowchart of SLT Participant Involvement]	1.3	26 October 2018
Other [Checklist A: CAPSA]		18 June 2018
Other [Checklist S: SLT Consent From]	2.3	26 October 2018
Other [Study Protocol]	3.3	23 October 2018
Other [Checklist Y: Letter to Consultant Surgeon]	1.1	26 October 2018
Other [Checklist N CG Consent 26 10 18]	3.4	26 October 2018
Other [Checklist S SLT Consent Form 26 20 18]	3.4	26 October 2018
Other [Study Protocol 23 10 18]	3.3	23 October 2018
Other [TRACK CHANGES Checklist C PIS CP Group 26 10 18]	3.4	26 October 2018
Other [Checklist L PIS Control Group 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist L PIS COntrol Group 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist R PIS SLT Version 3.4 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist E CPL Consent 26 10 18]	3.4	26 October 2018
Other [Checklist E CPL Consent 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist N CG Consnet 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist S SLT Consent Form 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist K Email to gatekeepers for Control Group Participants]	2.0	09 October 2018
Other [TRACK CHANGES Protocol 23 10 18]	3.3	23 October 2018
Other [HRA and REC Feedback 26 10 18]		26 October 2018
Referee's report or other scientific critique report [Amendments made following University Ethical review]	1.0	01 June 2018
Summary CV for Chief Investigator (CI) [CV for CI]	1.0	26 April 2018
Summary CV for student [Student, CI CV]		26 April 2018
Summary CV for supervisor (student research) [Tanya Rihtman Chief Supervisor CV]		19 June 2018

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:

http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

18/EM/0253	Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

This item has

This item has been removed due to third party copyright. The

Pp

Professor Cris Constantinescu Chair

Email:NRESCommittee.EastMidlands-Nottingham1@nhs.net

Enclosures:	"After ethical review – guidance for researchers"
Copy to:	*Professor Olivier Sparagano* *Ms Jaclyn Griffiths, Birmingham Women's and Children's NHS Trust/Coventry University*

Appendix J: Health Research Authority (HRA) Approval

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Miss Elizabeth Fitzpatrick

Senior Specialist Speech and Language Therapist, PhD

Student

Birmingham Women's and Children's NHS Trust/Coventry University

Speech and Language Therapy Department

Birmingham Children's Hospital

Steelhouse Lane, Birmingham

B4 6NH

Email: hra.approval@nhs.net

Research-permissions@wales.nhs.uk

15 January 2019

Dear Miss Fitzpatrick

HRA and Health and Care

Study title:	**The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip**
IRAS project ID:	**242296**
REC reference:	**18/EM/0253**
Sponsor	**Coventry University**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

How should I continue to work with participating NHS organisations in England and Wales? You should now provide a copy of this letter to all participating NHS organisations in

England and Wales, as well as any documentation that has been updated as a result of the assessment.

Following the arranging of capacity and capability, participating NHS organisations in England and Wales that are conduction all study activities (main site) should **formally confirm** their capacity and capability to undertake the study. How this will be confirmed is detailed in the "*summary of assessment*" section towards the end of this letter. You should then work with each organisation that has confirmed capacity and capability and provide clear instructions when research activities can commence.

Participating NHS organisations in England and Wales that are conducting the SLT rating of video recordings **will not** be required to formally confirm capacity and capability before you may commence

research activity at site. As such, you may commence the research at each organisation 35 days following sponsor provision to the site of the local information pack, so long as:

- You have contacted participating NHS organisations (see below for details)
- The NHS organisation has not provided a reason as to why they cannot participate
- The NHS organisation has not requested additional time to confirm.

You may start the research prior to the above deadline if the site positively confirms that the research may proceed.

If not already done so, you should now provide the local information pack for your study to your participating NHS organisations. A current list of R&D contacts is accessible at the NHS RD Forum website and these contacts MUST be used for this purpose. After entering your IRAS ID you will be able to access a password protected document (password: **Redhouse1**). The password is updated on a monthly basis so please obtain the relevant contact information as soon as possible; please do not hesitate to contact me should you encounter any issues.

Commencing research activities at any NHS organisation before providing them with the full local information pack and allowing them the agreed duration to opt-out, or to request additional time (unless you have received from their R&D department notification that you may commence), is a breach of the terms of HRA and HCRW Approval. Further information is provided in the "*summary of assessment*" section towards the end of this document.

It is important that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details of the research management function for each organisation can be accessed here.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?

HRA and HCRW Approval does not apply to NHS/HSC organisations within the devolved administrations of Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) has been sent to the coordinating centre of each participating nation. You should work with the relevant national coordinating functions to ensure any nation specific checks are complete, and with each site so that they are able to give management permission for the study to begin.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?

HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your nonNHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

I am a participating NHS organisation in England or Wales. What should I do once I receive this letter?

You should work with the applicant and sponsor to complete any outstanding arrangements so you are able to confirm capacity and capability in line with the information provided in this letter.

The sponsor contact for this application is as follows:

Name: Professor Olivier Sparagano

Tel: 02477659732

Email: iras-sponsor@coventry.ac.uk

Who should I contact for further information?

Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **242296**. Please quote this on all correspondence.

Yours sincerely

Kelly Rowe

Assessor

Email: hra.approval@nhs.net

Copy to: *Professor Olivier Sparagano, Coventry University, Sponsor contact*

Ms Jaclyn Griffiths, Birmingham Women's and Children's NHS Trust/Coventry University, Lead NHS R&D contact

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Covering letter on headed paper [Covering Letter]		02 May 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Public and Product Liability]		15 August 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Employer Liability]		01 August 2018
HRA Schedule of Events [Validated SOE: Other site]	1.0	20 August 2018
HRA Schedule of Events [Validated SOE: Main site]	1.0	20 August 2018
HRA Statement of Activities [Validated SOA: Other site]	1.0	10 September 2018
HRA Statement of Activities [Validated SOA: Main site]	1.0	10 September 2018
IRAS Application Form [IRAS_Form_30072018]		30 July 2018
Letter from sponsor [Sponsor Letter]		18 June 2018
Letters of invitation to participant [Checklist Q Invitation to SLT Participants via Gatekeeper]	3.0	26 April 2018
Other [Checklist L PIS for Control Group]	2.0	30 May 2018
Other [Checklist M The Communication Trust Universally Speaking]		27 March 2018

Other [Checklist S SLT Consent Form]	2.11	11 May 2018
Other [Sponsor Professional Indemnity Letter]		31 July 2017
Other [Coventry University Ethics Certificate]		18 June 2018
Other [Supervisor 3 CV]		18 June 2018
Other [CV for Supervisor 2 Jane Coad]		21 June 2018
Other [Checklist W CSG Feedback]	1.0	18 June 2018
Other [Checklist X CSG Changes]	1.0	28 June 2018
Other [Checklist C Participant Information Sheet Cleft Group]	3.4	26 October 2018
Other [Checklist E Cleft Group Consent Form]	2.3	26 October 2018
Other [Checklist K Email to gatekeepers for Control Group Participants]	2.0	09 October 2018
Other [Checklist N Consent Form Control Group]	2.3	26 October 2018
Other [Checklist R Participant Information Sheet SLT Group]	3.4	26 October 2018
Other [GCP 2018 Certificate]		12 September 2018
Other [Paediatric Consent Certificate]		12 September 2018
Other [Checklist B: Aims and Objectives]	1.0	18 June 2018
Other [Checklist D: Telephone script for Control Group]	2.0	26 March 2018
Other [Checklist G: Participants with CPL Identification Log]	1.1	26 October 2018
Other [Video/Audio Recording Request Form]		26 October 2018
Other [Checklist F: Screening and Enrolment Log CPL]	1.1	26 October 2018
Other [Assessment Recording Form]	1.1	26 October 2018
Other [Checklist J: Participant Video Recording Numbers]	1.1	26 October 2018
Other [Checklist O: Control Group Screening Enrolment LOg]	1.1	26 October 2018
Other [Checklist P: Control Group Identification Log]	1.3	26 October 2018
Other [Gatekeeper letter to SLTs]	3.1	26 October 2018
Other [Checklist T: SLT Questionnaire]	2.2	26 October 2018
Other [Checklist U: Universal Parameters Recording Form]	1.2	26 October 2018
Other [Checklist V: CPL Group Letter to Parents, Guardians]	1.2	26 October 2018
Other [Draft Flowchart of Listening Process]	2.1	26 October 2018
Other [Draft Flowchart of Control Group Involvement]	1.3	26 October 2018
Other [Draft Flowchart of CPL Group Involvement]	1.3	26 October 2018
Other [Draft Flowchart of SLT Participant Involvement]	1.3	26 October 2018
Other [Checklist A: CAPSA]		18 June 2018
Other [Checklist S: SLT Consent From]	2.3	26 October 2018
Other [Study Protocol]	3.3	23 October 2018
Other [Checklist Y: Letter to Consultant Surgeon]	1.1	26 October 2018
Other [HRA and REC Feedback]		26 October 2018
Other [Checklist N CG Consent 26 10 18]	3.4	26 October 2018
Other [Checklist S SLT Consent Form 26 20 18]	3.4	26 October 2018
Other [Study Protocol 23 10 18]	3.3	23 October 2018
Other [TRACK CHANGES Checklist C PIS CP Group 26 10 18]	3.4	26 October 2018

Other [Checklist L PIS Control Group 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist L PIS Control Group 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist R PIS SLT Version 3.4 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist E CPL Consent 26 10 18]	3.4	26 October 2018
Other [Checklist E CPL Consent 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist N CG Consent 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist S SLT Consent Form 26 10 18]	3.4	26 October 2018
Other [TRACK CHANGES Checklist K Email to gatekeepers for Control Group Participants]	2.0	09 October 2018
Other [TRACK CHANGES Protocol 23 10 18]	3.3	23 October 2018
Referee's report or other scientific critique report [Amendments made following University Ethical review]	1.0	01 June 2018
Summary CV for Chief Investigator (CI) [CV for CI]	1.0	26 April 2018
Summary CV for student [Student, CI CV]		26 April 2018
Summary CV for supervisor (student research) [Tanya Rihtman Chief Supervisor CV]		19 June 2018

Summary of assessment

The following information provides assurance to you, the sponsor and the NHS in England and Wales that the study, as assessed for HRA and HCRW Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England and Wales to assist in assessing, arranging and confirming capacity and capability.

Assessment criteria

Section	Assessment Criteria	Compliant with Standards	Comments

1.1	IRAS application completed correctly	Yes	The applicant has confirmed the following cleft centres may be approached to take part in the study; **Cleft Net East** Addenbrookes Hospital, Hills Road, Cambridge, CB2 0QQ **North Thames Cleft Service** Great Ormond Street Hospital for Children, Great Ormond Street , WC1N 3JH **Northern and Yorkshire Cleft Service** Leeds General Infirmary, Great George Street, Leeds View Map Royal Victoria Infirmary, Royal Victoria Infirmary, Newcastle-upon-Tyne **Northern Ireland Cleft Service** Royal Victoria Hospital, Falls Road, Belfast, Northern Ireland **Northwest England, Isle of Man, North Wales Cleft Service** Alder Hey Hospital, Liverpool Eaton Road, Liverpool , Merseyside, L12 2AP Royal Manchester Children's Hospital, Oxford road, Manchester **Scottish Cleft Service** Royal Hospital for Sick Children, 1345 Govan Road, Glasgow **South Thames Cleft Service** St Thomas Hospital, 1st Floor, South Wing, St Thomas' Hospital, Westminster Bridge Road, London, SE1 7EH

Section	Assessment Criteria	Compliant with Standards	Comments

			South Wales South West Managed Clinical Network Bristol Dental Hospital,South West Cleft Service, Bristol Dental Hospital, Lower Maudlin Street, Bristol Morriston Hospital, Swansea, SA6 6NL **Spires Cleft Service** Salisbury District Hospital, Salisbury, Wiltshire The John Radcliffe, Oxford, Headley Way, Headington, Oxford **Trent Cleft Service** Nottingham City Hospital, City Hospital Campus, Hucknall Road, Nottingham, NG5 1PB **West Midlands Cleft Service** Birmingham Children's Hospital, Steelhouse Lane, Birmingham, B4 6NH
2.1	Participant information/consent documents and consent process	Yes	No comments
3.1	Protocol assessment	Yes	No comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	A statement of activities has been submitted and the sponsor is not requesting and does not expect any other site agreement to be used. Although formal confirmation of capacity and capability is not expected of all or some organisations participating in this study, and such organisations would therefore be assumed to have confirmed their capacity and capability should they not respond to the contrary, we would ask that these organisations pro-actively engage with the sponsor in order to confirm at as early a date as possible.

Section	Assessment Criteria	Compliant with Standards	Comments
			Confirmation in such cases should be by email to the CI and Sponsor confirming participation based on the relevant Statement of Activities and information within this letter.
4.2	Insurance/indemnity arrangements assessed	Yes	Sponsor has confirmed the no fault indemnity will be covered by the medical malpractice policy they hold.
4.3	Financial arrangements assessed	Yes	No application for external funding has been made. The statement of activities confirms there are no funds available to sites from the sponsor.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	No comments
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England and Wales

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.
Participating NHS organisations will either be; • Conducting all study activities (main site) • conducting the SLT rating of video recordings only (other site) The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England and Wales in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. Where applicable, the local LCRN contact should also be copied into this correspondence. If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England and Wales which are not provided in IRAS, the HRA or HCRW websites, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net or HCRW at Research-permissions@wales.nhs.uk. We will work with these organisations to achieve a consistent approach to information provision.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and Wales, and the minimum expectations for education, training and experience that PIs should meet (where applicable).
A PI is expected at sites conducting all study activities; neither a PI nor local collaborator is expected at sites conducting the video rating. GCP training is not a generic training expectation, in line with the HRA/HCRW/MHRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken
It is anticipated that all study activities at participating sites will be conducted by local staff with an existing contractual relationship. No further HR good practice arrangements expected.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales to aid study set-up.
The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

Appendix K: The Combined Word List Picures

(from the next page)

The Combined Word List

The Combined Word List

The Combined Word List

The Combined Word List

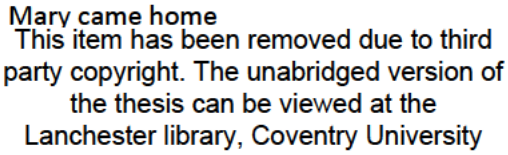
The Combined Word List

The Combined Word List

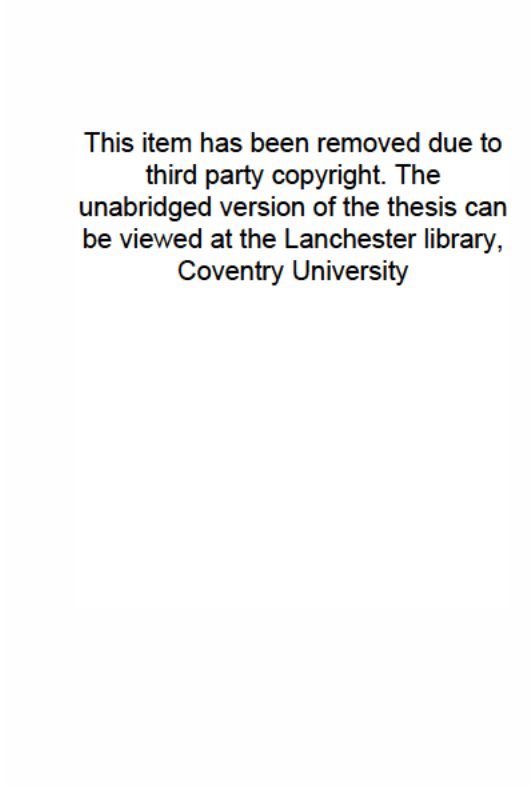
Appendix L: Speech Sample B Pictures

Sample B Sentences

1. Mary came home



2. Puppy has a paper



3. Bob is a baby

4. Phone fell off

5. Dave driving van

6. Neil is ten

7. I like the ball

8. Tim has a hat

9. This item has been

This item has been removed due to third party copyright. The unabridged version of the thesis can be viewed at the Lanchester library, Coventry University en

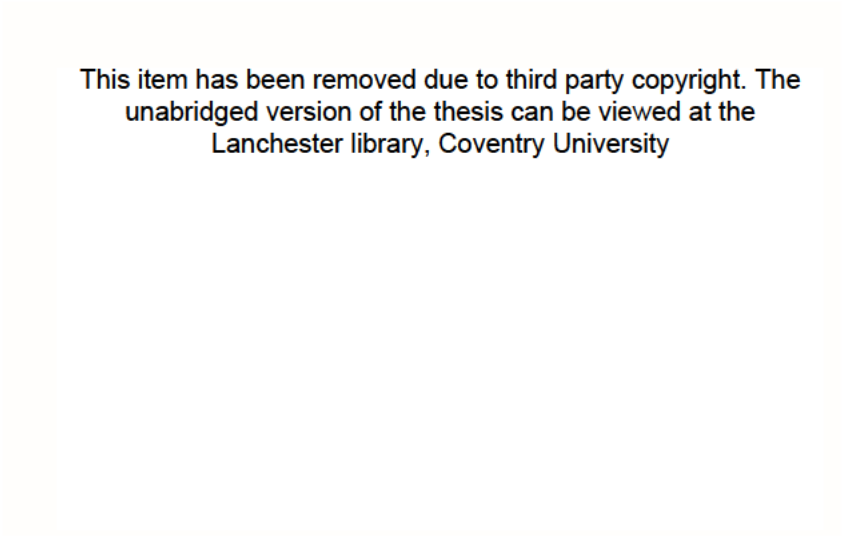
10. Girl washing her hands

This item has been removed due to third party copyright. The unabridged version of the thesis can be viewed at the Lanchester library, Coventry University

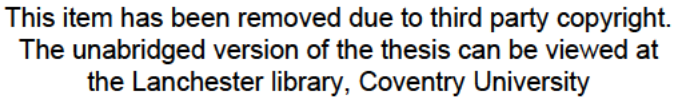
11. A sad face

This item has been removed due to third party copyright. The unabridged version of the thesis can be viewed at the Lanchester library, Coventry University

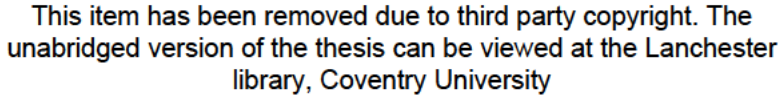
12. She is on the bus



13. Zebra lives at the zoo



14. Fish and chips



15. A snail shell

16. The chick is hatching

17. Slug in the salad

18. Karen is making a cake

19. The children juggle

20. Tiger in the jungle

21. Ring the bell

22. A spotty dog

23. Wear your welly

24. Wow, a yo-yo

Appendix M: CP±L Group Participant/Parent Information Leaflet (version 3.4)

The Early Assessment of Speech Outcomes in three-year-old children with Cleft Palate +/- Cleft Lip

(The EASO Study)

Research Team

Beth Fitzpatrick (Chief Investigator), Coventry University, Birmingham Women's and Children's NHS Trust

Supervisory Team at Coventry University: Dr Tanya Rihtman, Professor Jane Coad

External Supervisor: Dr Debbie Sell, Great Ormond Street Hospital

Your child is being invited to take part in a research study which aims to develop a speech assessment framework to assess speech outcomes in three-year-old children with cleft palate.

We would like to invite your child to take part in a research study.

- Please read the following information before you decide whether you want your child to take part. The information will tell you more about the study and what it will involve.
- Our team will talk through the information with you and answer any questions.
- You may also want to discuss the study with others to help you make your decision.
- Please take your time before deciding whether you would like your child to take part or not.
- You can decide that you want your child to stop taking part at any point up until the Speech and Language Therapists in the study have analysed all the participant recordings. At this point, the anonymised data will be collated and combined.
- If you decide to take part, we will give you a copy of this information to keep and ask you to sign a Consent Form.

The information is in two sections. Part 1 tells you about the background of the study, Part 2 tells you about what your child would need to do as part of the study and what will happen next.

Part 1: About the study

1) What is the purpose of the study?

The aim of the research is to develop an assessment framework to assess speech in three-year-old children with cleft palate. To develop the assessment framework we need to find out the following information in the study:

- Which speech samples (the type of speech produced e.g. single words in picture naming, repeating sentences, playing with and talking about toys) are most easily completed by three-year-olds with a cleft palate?
- How do different speech samples, and rating scales impact upon speech ratings made by Cleft Speech and Language Therapists?
- Which speech samples and rating scales do Cleft Speech and Language Therapists find easiest or the most appropriate to use?

We will then use the information we have collected to put forward a new assessment framework.

In the UK one of the key times that we see children for an assessment is when they are three years old. This is an important assessment time as we can pick up on any problems with their speech and arrange for them to have any extra help if they need it before they start school. However, at the moment there is no agreed format on how best to assess speech at this age. If we were able to develop an assessment framework this could be used at different cleft centres across the UK. This may allow us to compare how our patients are doing, measure the impact of any speech therapy, and track their progress between ages three and five.

2) Why is my child being asked to take part?

Your child is being asked to take part because they were born with a cleft palate +/- a cleft lip, because they are three years old, and because they are looked after by the West Midlands Cleft Lip and Palate Service (at Birmingham Children's Hospital).

3) Do we have to take part?

No. It is your decision whether you want your child to take part. If you do not want to take part you do not have to tell the team why.

4) Where is the study taking place?

The study is taking place in the Speech and Language Therapy Department at Birmingham Children's Hospital.

5) Who is organising and running the study?

The study is being run by Beth Fitzpatrick, a Senior Speech and Language Therapist in Cleft Palate at Birmingham Children's Hospital for the West Midlands Cleft Lip and Palate Service. Beth is completing the research study as part of a PhD at Coventry University under the supervision of a supervisory team (detailed above).

Speech and Language Therapists who work in Cleft Palate at Birmingham Children's Hospital are carrying out the speech assessments.

6) Who has checked the study?

The research study has been given ethical approval from:

- The NHS- East Midlands- Nottingham 1 Research Ethics Committee (242296)

- Coventry University- Research Ethics team at Coventry University (P66325)
- Birmingham Women's and Children's NHS Trust (18_BC_HNS_NO_134).

Part 2: What we are asking you and your child to do?

8) What will your child need to do if they take part?

- You and your child will be seen by a Cleft Speech and Language Therapist working at Birmingham Children's Hospital when they attend their usual Speech and Language Therapy appointment at age 3.
- Your child will complete different assessment activities e.g. naming pictures, repeating short phrases, playing with toys. This is similar to the usual assessment activities at age 3 but your child will be asked to complete more than one activity.
- As part of the study, all the assessments will be video recorded.
- We do not expect the study assessment to take longer than an hour. Usually, a three-year assessment takes 30-60 minutes. Your child can take breaks during the assessment if they need to.
- You are able to stop the assessment session at any time, and this will not affect your child's future treatment.
- After the appointment, you will receive a report from the Cleft Speech and Language Therapist as usual.
- The research team will write to your child's Cleft Consultant to let them know that your child took part in the study.

9) What will happen to the video recording?

- At the end of the session, your child's recording will be stored securely on the hospital computer drive used to store Cleft Patients' speech assessments.
- The recording will be checked to make sure that the sound and video are of high quality. We will not be able to use recordings which have poor sound or video quality.
- Cleft Speech and Language Therapists will watch the video and analyse your child's speech.
- Some of the Speech and Language Therapists will work in different cleft centres. This will tell us whether Speech and Language Therapists at different cleft centres rate speech differently.
- The Speech and Language Therapists rating the videos will be told if your child was aged was 3 years- 5 months or 3 years 6 months- 3 years 11 months at the assessment. This information will help them in the analysis. They will not be given any other information about your child.
- So that the Speech and Language Therapists can watch the video, your child's video will be uploaded onto a secure NHS approved SharePoint External Data Exchange System.
- Only Speech and Language Therapists taking part in the study will be able to watch your child's video and this will be securely password protected.
- Some of the videos will be used in practise ratings by the Speech and Language Therapists to give them feedback on how they rate speech.
- When all the Speech and Language Therapists have analysed your child's speech your child's video will be deleted from the NHS approved SharePoint External Data Exchange System.

10) What are the possible benefits or disadvantages of taking part?

- Your child will receive their speech assessment at 3-years of age as usual (this would be the case if they didn't participate in the study).

- You may be helping us to improve speech assessments for children affected by cleft palate in the future.
- Although the assessment should take no longer than an hour (excluding any breaks your child might want to take) the assessment may take slightly longer than the usual three-year assessment.

11) Will my child's participation be confidential?

- Yes. Information about your child will be strictly confidential. Information about your child will be given a unique number so only the research team will be able to identify your child. Your child will not be identifiable in the study.
- All Speech and Language Therapists who rate the videos are employed by the NHS and will follow the NHS Code of Practise for patient confidentiality.
- We will handle, use, and store data following the Data Protection Act 1998 and the General Data Protection Regulation 2016. **Please see the General Data Protection Regulation information at the end of this leaflet.**
- Data may be reviewed confidentially by members of the hospital and university to check the study is being carried out correctly.

12) What if I want to withdraw my child from the study?

- You can decide that you want your child to stop taking part at any point up until the Speech and Language Therapists in the study have rated all the videos. At this point, the anonymised data will be combined, and it will not be possible to take out your child's data.

13) What happens at the end of the study?

- We will make suggestions for cleft speech assessments at age 3.
- At the end of the study, your child's video recording will form part of their medical record and be stored securely alongside other cleft speech assessment recordings on a designated secure drive at Birmingham Children's Hospital.
- A copy of your child's documents will be securely archived (stored) for 25 years following procedures at Birmingham Women's and Children's NHS Trust.
- Only a fully anonymised dataset will be kept at the end of the study.
- At the end of the study, we would like to share our findings with other medical professionals for example in medical journals or presentations. Your child's data will be anonymised – this means that they will not be identifiable in any results or reports.
- If you agree we will send you a letter at the end of the study to let you know what we have found out. It will be around 2 years until the study is completed.

14) What if I have a problem?

For **questions, comments** or if you would like to **contact the team to say that you are interested in your child taking part** in the study then please contact:

Beth Fitzpatrick (Chief Supervisor)
Speech and Language Therapy
Birmingham Children's Hospital
Steelhouse Lane
Birmingham
B4 6NH

Email: beth.fitzpatrick@nhs.net
Telephone: 0121 3339382

If you have a **complaint** or any **concerns** about the study please contact:

Coventry University:
Professor Olivier Sparagano
Associate Pro-Vice-Chancellor (Research)
Coventry University
Alan Berry Building
Priory Street
Coventry
CV1 5FB
Email: iras-sponsor@coventry.ac.uk
Telephone: 02477659732

Birmingham Women's and Children's NHS Foundation Trust:

Patient Advice and Liaison Team:
0121 333 8961 or 0121 333 9391

If you would be interested in your child taking part in the study, please email Beth Fitzpatrick: beth.fitzpatrick@nhs.net or call 0121 333 9382

Please read the General Data Protection Regulation Information:

- *Birmingham Women's and Children's NHS Trust will collect information about your child for this research study from you and your child's medical records. Birmingham Women's and Children's NHS Trust will not provide any identifying information about your child to Coventry University. We will use this information to report on the overall group of participants i.e. the number of children with bilateral clefts, the number of females etc. and your child will not be identifiable.*
- *Birmingham Women's and Children's NHS Trust will keep your child's name, contact details and other information relating to your child confidential and will not pass this information to Coventry University. Birmingham Women's and Children's NHS Trust will use this information as needed, to contact you about the research study, and to oversee the quality of the study. Certain individuals from Coventry University and regulatory organisations may look at your medical and research records to check the accuracy of the research study. Coventry University will only receive information without any identifying information. The people who analyse the information will not be able to identify your child and will not be able to find out your child's name, contact details or other information relating to your child.*
- *Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.*
- *You can find out more about how we use your information by contacting Beth Fitzpatrick (see contact details above).*
- *Coventry University is the sponsor for this study based in the United Kingdom. We will be using information from you and your child's medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. All identifiable information about your child will be put*

into their medical record or destroyed at the end of the study. Coventry University will only retain a fully anonymised data set at the end of the study.

Appendix N: CP±L Group Letter to parents/guardians

Beth Fitzpatrick
Senior Cleft Speech and Language Therapist
Speech and Language Therapy Department
Birmingham Children's Hospital
Steelhouse Lane
Birmingham
B4 6NH
DATE:

Dear Parent(s)/Guardian(s)

A research study involving three-year-old children with cleft palate +/- cleft lip is currently taking place at the West Midlands Cleft Lip and Palate Service.

We would like to invite your child to take part in the study when they attend their upcoming appointment for their 3-year-old speech assessment.

Included with the appointment letter is a Participant Information Leaflet with more information about the research study and what it would involve if your child takes part.

In approximately 1 weeks' time the Researcher, Beth Fitzpatrick, who is a Senior Cleft Speech and Language Therapist at the West Midland's Cleft Lip and Palate Service will contact you by telephone to ascertain if you would be interested in your child taking part in the study.

There is no obligation for your child to take part. If you do not wish for your child to take part then please let us know. You do not need to give a reason. If you do not wish for your child to take part this will not affect their care in any way, they will be seen for their 3-year-old speech assessment as normal at the time and date of the appointment you have been sent, and they will receive further appointments from the team as normal.

If you are interested in your child taking part, then please let Beth know when she telephones or contact the team directly. If you would like your child to take part, then they will be seen for a study appointment at the time and date on the appointment letter you have been sent.

Yours sincerely,

Beth Fitzpatrick

Chief Investigator and Senior Cleft Speech and Language Therapist

Appendix O: Telephone Script- for use with parent/guardian of eligible patients with CP±L

1) Chief Investigator (CI) establishes that they are speaking to the parent(s)/guardian(s) of X.
   "Hello, am I speaking to the parent or guardian of X?"

   If the CI is speaking to the parent(s)/guardian(s) to continue to step 2, if the parent(s)/guardian(s) is unavailable, the CI will call back, repeating step 1.
2) CI introduces themselves, and the reason for the telephone call.
   "It's (Chief Investigator's Name) calling from Speech and Language Therapy at Birmingham Children's Hospital. Do you have time to talk about your child's upcoming appointment and some information we sent you with the appointment letter?"

3) If the parent(s)/guardian(s) do not have time to speak on the phone, the CI asks if they can call back at a convenient time. If the parent(s)/guardian(s) have time to speak then the CI will continue.
   "We recently sent your child an appointment for their three-year Speech and Language Therapy appointment and some information regarding a research study that is taking place. Have you received the appointment letter and this information?"

   If the parent(s)/guardian(s) have received the appointment letter and PIL, the CI will move on to step 5.

4) If the parent(s)/guardian(s) have not received the appointment letter containing the PIL then the CI will inform the parent(s)/guardian(s) of the appointment date, and what the PIL was about.

   "An appointment for your child has been scheduled on X. We also sent some written information about a research study that is taking place. The aim of the research is to develop an assessment framework to assess speech outcomes in three-year-old patients with cleft palate. We hope to find out information during the study that we can use to design the assessment. We want to find out which speech samples or the type of speech produced e.g. single words in picture naming, repeating sentences, playing with and talking about toys are most easily completed by three-year-olds with cleft palate and which of these speech samples when listened to by Cleft Speech and Language Therapists, results in the most consistent and repeatable speech ratings. We also want to find out which method of rating how well the palate is working during speech and what scale is best to use when children are aged three years. We will also gain information about whether the assessment investigates appropriate areas of cleft speech, what Cleft Speech and Language Therapists think of the assessment, and if they prefer listening to a particular speech sample.

   We sent the information because we wondered if this is something you would consider your child taking part in when they come for their three-year assessment. It is your decision if you would like your child to take part, and if you do not wish for them to take part we will carry out the usual speech assessment as planned."

   "Do you have any questions about the study?"
   The CI will answer any questions the parent(s)/guardian(s) have about the study.

   "Do you think the research is something you would like your child to take part in?"
   If the parent(s)/guardian(s) indicate that they would not like their child to participate the CI will confirm the time of the protocol appointment and that their child will be seen for the usual Speech and Language Therapy assessment at age three years.

If the parent(s)/guardian(s) indicate that they would like their child to participate, the CI will send the appointment letter and PIL to the parent(s)/guardian(s) again. In addition the CI will provide another copy of the PIL at the time of the appointment, and give the parent(s)/guardian(s) time to read through this and ask any questions, before asking them to indicate if they would like their child to participate, and taking formal consent for their child to participate in the research study.

5) If the parent(s)/guardian(s) have not read the PIL the CI will follow step 4. If the parent(s)/guardian(s) have read the PIL. The CI will ask if they have any questions about the study.

"Do you have any questions about the study?"
The CI will answer any questions the parent(s)/guardian(s) have about the study.

"It is your decision if you would like your child to take part, and if you do not wish for your child to take part we will carry out the usual speech assessment as planned. Do you think the research is something you would like your child to take part in?"
If the parent(s)/guardian(s) indicate that they would not like their child to participate the CI will confirm the time of the protocol appointment and that their child will be seen for the usual Speech and Language Therapy assessment at age three -years.

6) If the parent(s)/guardian(s) indicate that they would be willing for their child to participate the CI will ask the parent(s)/guardian(s) to consider if they would still like to participate in the study if their child is participating in any other research studies.

"Your child is able to participate in this study even if they are participating in another research study. You are not under any obligation to tell me if your child is participating in any other research studies, but please consider whether participating in this study is convenient for you and your child. It is your decision as to whether you are happy for your child to take part in more than one research study."

If the parent(s)/guardian(s) indicate that they would like their child to participate the CI will provide another copy of the PIL at the time of the appointment, and give the parent(s)/guardian(s) time to read through this and ask any questions, before asking them to indicate if they would like their child to participate, and taking formal consent for their child to participate in the research study.

Appendix P: Consent Form for CP±L Group (version 3.4)

The Early Assessment of Speech Outcomes in three-year-old children with Cleft Palate +/- Cleft Lip (The EASO Study)

	Please initial the box
1. I confirm that I have read and understood the participant information leaflet (version number 3.4) for the above study and that I have had the opportunity to ask questions and to have these questions answered satisfactorily	
2. I understand that my child's participation is voluntary and that I am able to withdraw my child from the study up until the Speech and Language Therapists have analysed all the recordings	
3. I understand that my child's assessment will be video recorded and that this video will be stored on a secure drive used to store speech recordings at Birmingham Children's Hospital	
4. I understand that my child's video will be viewed by Cleft Speech and Language Therapists taking part in the study, who may work in other NHS cleft centres	
5. I understand that the video of my child's assessment will be uploaded onto a secure NHS approved SharePoint External Data Exchange System and viewed by Cleft Speech and Language Therapists taking part in the study, who may work in other NHS cleft centres	
6. I give permission for the researcher to access my child's Speech and Language Therapy notes	
7. I understand that information about my child will be strictly confidential and my child will not be identifiable in any outputs from the research study e.g. presentations, written publications etc.	
8. I agree to my child's participation in the research study	
9. I agree to my child's Cleft Consultant being notified about their participation in the research study	
10. I confirm that I have Parental Responsibility and that I can give consent for my child	
11. I would like to be informed of the study findings at the end of the study	

Name of participant:……

Date of Birth of participant:……

Name of parent/legal guardian:……

Signature:…… Date……

Name of witness:……

Signature:…… Date……

Name of the researcher:……

Signature:…… Date……

Appendix Q: Email to gatekeepers for recruitment of Control Group

Dear Colleagues,

I am currently running a research study which aims to develop an assessment framework to assess speech outcomes in three-year-old children with cleft palate as part of PhD at Coventry University. The research study has the relevant ethical approvals:

- The NHS- East Midlands- Nottingham 1 Research Ethics Committee (242296)
- Coventry University- Research Ethics team at Coventry University (P66325)
- Birmingham Women's and Children's NHS Trust (18_BC_HNS_NO_134).

Although all UK cleft centres carry out a speech assessment at age three years, at the moment there is no agreed format on how best to assess speech at this age. If we were able to develop an assessment framework this could be used at different cleft centres across the UK. This may allow us to compare how our patients are doing and track their progress between ages three and five.

To develop the assessment framework we need to find out the following information in the study:

- Which speech samples (the type of speech produced e.g. single words in picture naming, repeating sentences, playing with and talking about toys) are most easily completed by three-year-olds with a cleft palate?
- Which speech samples result in the most consistent and repeatable speech ratings made by the Cleft Speech and Language Therapists?
- Which way of assessing how well a child's palate is working for their talking results in the most consistent and repeatable speech ratings made by the Cleft Speech and Language Therapists?
- Which rating scale results in the most consistent and repeatable speech ratings made by the Cleft Speech and Language Therapists?
- Does the assessment assess the areas of speech which should be investigated in three-year-old children with a cleft palate?
- Which speech samples do Speech and Language Therapists think are the most appropriate to listen to when analysing the speech of 3-year old children with CPL?

We will then use the information we have collected to put forward a new assessment framework.

In the study we will be recruiting three-year-old patients with cleft palate +/- cleft lip, assessing their speech using the different speech activities and recording the assessment. Cleft Speech and Language Therapists will then watch the recordings and analyse the children's speech. To check that children without a cleft palate are not scored as having the type of speech difficulties which are usually only seen in children with cleft palate, *we need to recruit three-year-old children without a cleft and who do not have any difficulties with their speech and language*. The children would be asked to attend an assessment session in the Speech and Language Therapy Department here at Birmingham Children's Hospital. Their parents/guardians will be sent a short report

following the assessment. Attached to this email is a Participant Information Sheet, with more details about the study.

I am emailing you to ask if you would consider talking about the study to any of your friends or family members who have three-year-old children. To ensure that the research study is carried out following the principals of ethical research, the three-year-old child cannot be your own and those who would be interested in participating in the study are asked to contact the Chief Investigator of their own volition.

If any of your friends or family are interested in their child's participation in the study they should contact me via email or phone: beth.fitzpatrick@bch.nhs.uk, 0121 3339382

Your help is much appreciated.

Yours faithfully,

Beth Fitzpatrick, Chief investigator

Supervisory Team at Coventry University: Dr Tanya Rihtman, Professor Jane Coad

External Supervisor: Dr Debbie Sell, Great Ormond Street Hospital

Appendix R: Control Group Participant/Parent Information Leaflet (version 3.4)

The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip

(The EASO Study)

Research Team

Beth Fitzpatrick (Chief Investigator), Coventry University, Birmingham Women's and Children's NHS Trust

Supervisory Team at Coventry University: Dr Tanya Rihtman, Professor Jane Coad

External Supervisor: Dr Debbie Sell, Great Ormond Street Hospital

We are inviting children *without* a cleft palate to take part in a research study which aims to develop a speech assessment for 3-year-old children with cleft palate.

We would like to invite your child to take part in a research study.

- Please read the following information before you decide whether you want your child to take part. The information will tell you more about the study and what it will involve.
- Our team will talk through the information with you and answer any questions.
- You may also want to discuss the study with others to help you make your decision.
- Please take your time before deciding whether you would like your child to take part or not.
- You can decide that you want your child to stop taking part at any point up until the Speech and Language Therapists in the study have analysed all the participant recordings. At this point, the data will be collated and combined, and it will not be possible to separate out your child's information.
- If you decide to take part, we will give you a copy of this information to keep and ask you to sign a Consent Form.

The information is in two sections. Part 1 tells you about the background of the study. Part 2 tells you about what your child would need to do in the study.

Part 1: About the study

1) What is the purpose of the study?

This aim of the study is to find out if there is a best way to assess speech in 3-year-old children with cleft palate. A cleft palate is an opening in the roof of the mouth which happens when a baby is growing in the womb. A cleft palate is repaired using surgery. Even after surgery, children with a cleft palate can have serious speech problems that other children do not usually have. This is sometimes called 'cleft speech.'

We assess speech when children are three years old so that we can help them as early as possible if need be. However, at the moment different Speech and Language Therapists and cleft centres assess cleft speech in different ways. We would like to make recommendations about how we all should assess speech in 3-year-old children with cleft palate. To do this we need to find out the following:

- What types of speech activities (the type of speech produced e.g. single words in picture naming, repeating sentences, playing with and talking about toys) are the easiest for most 3-year-olds with cleft palate to complete?
- How are ratings made by Cleft Speech and Language Therapists affected by the type of speech activity the child completes, and by different rating scales?
- Which speech activities and rating scales do Cleft Speech and Language Therapists find the easiest or the most suitable to use?

We will then use the information we have collected to make recommendations (when possible) about speech assessments at age 3.

2) Why is my child being asked to take part?

When we develop assessments for children who have certain types of difficulties, we need to make sure that we include children who do NOT have these difficulties. This allows us to be sure that our assessment picks up cleft speech difficulties for children who DO have a cleft, and does not pick up cleft speech difficulties for children who do NOT have a cleft.

Your child is being asked to take part because:

- They are three years old
- They were born *without* a cleft palate.

3) Do we have to take part?

No. It is your decision whether you want your child to take part. If you don't want your child to be in the study you don't have to contact us.

4) Where is the study taking place?

In the Speech and Language Therapy Department at Birmingham Children's Hospital.

Birmingham Children's Hospital is one of the NHS specialist centres that look after children born with cleft palates in the UK.

5) Who is organising and running the study?

The study is being run by Beth Fitzpatrick. Beth is a Senior Speech and Language Therapist in Cleft Palate at the West Midlands Cleft Lip and Palate Service (at Birmingham Children's Hospital). Beth is completing the research study as part of a PhD at Coventry University. She is being supervised by a team (see above).

Speech and Language Therapists who work in Cleft Palate at Birmingham Children's Hospital are carrying out the speech assessments.

6) Who has checked the study?

The research study has been given ethical approval from:

- The NHS- East Midlands- Nottingham 1 Research Ethics Committee (242296)
- Coventry University- Research Ethics team at Coventry University (P66325)
- Birmingham Women's and Children's NHS Trust (18_BC_HNS_NO_134).

Part 2: What we are asking you and your child to do?

8) What will your child need to do if they take part?

- You and your child will be invited to the Speech and Language Therapy Department at Birmingham Children's Hospital for an assessment. The research team will pay for your travel and parking.
- You and your child will be seen by a Cleft Speech and Language Therapist working at Birmingham Children's Hospital.
- Your child will complete different assessment activities e.g. naming pictures, repeating short phrases, playing with toys.
- The session will be video recorded.
- The study assessment should last about an hour (excluding breaks). Your child can take breaks during the assessment if they need to.
- After the appointment, you will receive a short report from the Cleft Speech and Language Therapist about your child's speech.
- We will ask you for the following information about your child: gender and child's date of birth. We will use this information to report on the overall group of children in the study e.g. how many were male, how many were aged between 3 years 0 months and 3 years 5 months etc.
- This information will be recorded on our database in an anonymised format e.g. using a number instead of writing male or female and your child will not be identifiable.

9) What will happen to the video recording?

- At the end of the session, your child's recording will be stored securely on the hospital computer drive used to store Cleft Patients' speech assessments.
- The recording will be checked to make sure that the sound and video are of high quality. We will not be able to use recordings which have poor sound or video quality.
- Cleft Speech and Language Therapists will watch the video and analyse your child's speech.
- Some of the Speech and Language Therapists will work in different cleft centres. This will tell us whether Speech and Language Therapists at different cleft centres rate speech differently.
- The Speech and Language Therapists rating the videos will be told if your child was aged was 3 years- 5 months or 3 years 6 months- 3 years 11 months at the assessment. This information will help them in the analysis. They will not be given any other information about your child.
- So that the Speech and Language Therapists can watch the video, your child's video will be uploaded onto a secure NHS approved SharePoint External Data Exchange System.

- Only Speech and Language Therapists taking part in the study will be able to watch your child's video and this will be securely password protected.
- Some of the videos will be used in practise ratings by the Speech and Language Therapists to give them feedback on how they rate speech.
- When all of the Speech and Language Therapists have analysed your child's speech your child's video will be deleted from the NHS approved SharePoint External Data Exchange System.

10) What are the possible benefits and disadvantages of taking part?

- We don't expect any disadvantages with your child's participation in the study.
- You will need to travel to Birmingham Children's Hospital and to stay for the assessment session; this will take up some of your time.
- Your child will receive a speech assessment which they may not have had otherwise.
- You will be sent a short report about your child's assessment after the appointment.
- In the unlikely event that the Speech and Language Therapist has any concerns about your child's speech, we will ask you if we can refer your child to community Speech and Language Therapy services.

11) Will my child's participation be confidential?

- Yes. Information about your child will be strictly confidential. Information about your child will be given a unique number so only the research team will be able to identify your child. Your child will not be identifiable in the study.
- All Speech and Language Therapists who rate the videos are employed by the NHS and will follow the NHS Code of Practise for patient confidentiality.
- We will handle, use, and store data following the Data Protection Act 1998 and the General Data Protection Regulation 2016. **Please see the General Data Protection Regulation information at the end of this leaflet.**
- Data may be reviewed confidentially by members of the hospital and university to check the study is being carried out correctly.

12) What if I want to withdraw my child from the study?

- You can decide that you want your child to stop taking part at any point up until the Speech and Language Therapists in the study have rated all the videos. At this point, the anonymised data will be combined, and it will not be possible to take out your child's data

13) What happens at the end of the study?

- We will make suggestions for cleft speech assessments at age 3.
- A copy of your child's documents will be securely archived (stored) for 25 years following procedures at Birmingham Women's and Children's NHS Trust.
- Only a fully anonymised dataset will be kept at the end of the study.
- At the end of the study, we would like to share our findings with other medical professionals for example in medical journals or presentations. Your child's data will be anonymised – this means that they will not be identifiable in any results or reports.
- If you agree we will send you a letter at the end of the study to let you know what we have found out. It will be around 2 years until the study is completed.

14) What do I do if I would like my child to take part?

Please contact **Beth Fitzpatrick (Chief Investigator)**

Email: **beth.fitzpatrick@nhs.net**

Telephone: **0121 3339382**

- Beth will check that your child meets the criteria to be included in the study.
- The study aims to recruit between 2 and 5 children without a cleft palate. It may not be possible to include every child whose parents would like them to take part. This will be determined on a first come first served basis.

15) What do I do if I have a problem?

For **questions, comments** or if you would like to **contact the team to say that you are interested in your child taking part** in the study then please contact:

Beth Fitzpatrick (Chief Supervisor)
Speech and Language Therapy
Birmingham Children's Hospital
Steelhouse Lane
Birmingham
B4 6NH
Email: beth.fitzpatrick@nhs.net
Telephone: 0121 3339382

If you have a **complaint** or any **concerns** about the study please contact:

Coventry University:
Professor Olivier Sparagano
Associate Pro-Vice-Chancellor (Research)
Coventry University
Alan Berry Building
Priory Street
Coventry
CV1 5FB
Email: iras-sponsor@coventry.ac.uk Telephone: 02477659732

Birmingham Women's and Children's NHS Foundation Trust:

Patient Advice and Liaison Team:
0121 333 8961 or 0121 333 9391

If you would be interested in your child taking part in the study, please email Beth Fitzpatrick: beth.fitzpatrick@nhs.net or call 0121 333 9382

Please read the General Data Protection Regulation Information:

- *Birmingham Women's and Children's NHS Trust will keep your child's name, contact details and other information relating to your child confidential and will not pass this information to Coventry University. Birmingham Women's and Children's NHS Trust will use this information as needed, to contact you about the research study, and to oversee the quality of the study.*

Certain individuals from Coventry University and regulatory organisations may look at your medical and research records to check the accuracy of the research study. Coventry University will only receive information without any identifying information. The people who analyse the information will not be able to identify your child and will not be able to find out your child's name, contact details or other information relating to your child.

- *Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw your child from the study, we will keep the information about your child that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.*
- *You can find out more about how we use your information by contacting Beth Fitzpatrick (see contact details above).*
- *Coventry University is the sponsor for this study based in the United Kingdom. We will be using information from you and your child's medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your child's information and using it properly. All identifiable information about your child will be stored for 25 years at the end of the study* years *following procedures at Birmingham Women's and Children's NHS Trust*. *Coventry University will only retain a fully anonymised data set at the end of the study.*

Appendix S: Consent form for Control Group (version 3.4)

NHS Birmingham Women's and Children's NHS Foundation Trust

Coventry University

The Early Assessment of Speech Outcomes in three-year-old children with Cleft Palate +/- Cleft Lip (The EASO Study)

Please initial the box

1. I confirm that I have read and understood the participant information leaflet (version number 3.4) for the above study and had the opportunity to ask questions and to have these questions answered satisfactorily	
2. I understand that my child's participation is voluntary and that I am able to withdraw my child from the study up until the Speech and Language Therapists in the study have analysed all the participant recordings. At this point, the anonymised data will be collated and combined.	
3. I understand that my child's assessment will be video recorded and that this video will be stored on a secure drive used to store confidential recordings at Birmingham Children's Hospital during the study	
4. I understand that my child's video will be viewed by Cleft Speech and Language Therapists taking part in the study, who may work in cleft centres other than the West Midlands Cleft Centre at Birmingham Children's Hospital	
5. I understand that the video of my child's assessment will be uploaded onto a secure NHS approved External SharePoint Data Exchange System and viewed by Cleft Speech and Language Therapists taking part in the study, who may work in other NHS cleft centres	
6. I understand that information about my child will be strictly confidential and my child will not be identifiable in any outputs from the research study e.g. presentations, written publications etc.	
7. I agree to my child's participation in the research study	
8. I confirm that I have Parental Responsibility and that I can give consent for my child	
9. I would like to be informed of the study findings at the end of the study	

Name of participant:……

Date of Birth of participant:……

Name of parent/legal guardian:……

Signature:…… Date……

Name of witness:……

Signature:…… Date……..

Name of the
researcher:……

Signature:…… Date……

Appendix T: Email to gatekeepers for recruitment of SLT Listeners

Dear Speech and Language Therapists,

Please find attached a Participant Information Leaflet which provides more information regarding a research study which is taking place at the West Midlands Cleft Lip and Palate Service. The study aim is to propose an assessment framework to validly and reliably assess speech outcomes in 3-year-old patients with CPL. This will be achieved through the examination of different speech samples, methods of rating speech characteristics associated with velopharyngeal function and different rating scales on the validity and reliability of listener judgements. The research study has the relevant ethical approvals from The NHS- East Midlands- Nottingham 1 Research Ethics Committee (242296), Coventry University- Research Ethics team at Coventry University (P66325), Birmingham Women's and Children's NHS Trust (18_BC_HNS_NO_134).

To determine which speech samples and rating methods result in the most reliable listener ratings we are recruiting Cleft Speech and Language Therapists, with CAPS-A training, to the study to act as expert listeners. This will involve listening to and making judgements about the speech of 3-year-old participants on three occasions four weeks apart. It is anticipated that the listening will take the most of two days in total. You will be able to access the recordings at your own NHS Cleft Centre via the NHS External Data Exchange system which you will be given access to.

We would like to recruit Cleft SLTs working in the NHS with designated sessions working with cleft patients who have completed CAPS-A training and carried out consensus listening in the audit of speech outcomes at age 5 years.

We will acknowledge all SLTs involved in the listening in any subsequent publications arising from the study. We will also give you a small gift as an expression of thanks.

Please forward this email to any of your colleagues who you think may be interested.

After reading the Participation Information Leaflet, if you have any questions or would like to be involved please contact beth.fitzpatrick@bch.nhs.uk, 0121 3339382. Please contact Beth within two weeks of receiving this email if you would like to participate in the study.

Yours faithfully,

Beth Fitzpatrick, Chief investigator

Appendix U: SLT Listeners- Participant Information Leaflet (version 3.4)

The Early Assessment of Speech Outcomes in three-year-old children with Cleft Palate +/- Cleft Lip
(The EASO Study)

Research Team

Beth Fitzpatrick (Chief Investigator), Coventry University, Birmingham Women's and Children's NHS Trust

Supervisory Team at Coventry University: Dr Tanya Rihtman, Professor Jane Coad

External Supervisor: Dr Debbie Sell, Great Ormond Street Hospital

You are being invited to take part in a research study investigating the impact of different speech samples and rating methods on listener reliability ratings in three-year-old children with Cleft Palate +/- Cleft Lip (CPL). These findings will be used to propose an assessment framework to assess speech outcomes in three-year-old children with CPL.

The information is in two sections. Part 1 outlines the background of the study, Part 2 outlines your involvement if you agree to participate in the study and what happens next.

Part 1: About the study

1) What is the aim of the study?

To propose an assessment framework to validly and reliably assess speech outcomes in three-year-old children with CPL, through the examination of the impact of different speech samples and rating methods on the validity and reliability of listener judgements, and the extent to which three-year-old children are able to fully complete different speech samples.

This will be achieved through the completion of the following objectives:

- To complete a scoping and identification exercise, to inform the parameters of speech that should be assessed and the types of speech samples included in the speech assessment.

- To determine the extent to which three-year-old participants with CPL can complete different target speech samples.

- To determine the impact of different speech samples on the validity and reliability of listener judgements.

- To ascertain the impact of different rating methods and scales on the reliability of judgements made by listeners of the speech characteristics associated with velopharyngeal function e.g. using separate scales to measure hypernasality, hyponasality, nasal emission, nasal turbulence, or using an overall scale to measure speech characteristics associated with velopharyngeal function, and the use of ordinal and visual analogue scales.

- To gain further information regarding the specificity of the speech assessment by using the assessment with three-year-old children without CPL and any known speech difficulties (acting as a control group) and examining Cleft Speech and Language Therapist's (SLTs) listener judgements.

- To measure the acceptability and usability of the speech assessment and rating methods to the Cleft SLTs who act as listeners through a questionnaire.

2) What are the potential benefits of the study?

The regular audit of speech outcomes at age five years has been identified as a factor which has contributed to the improvement in speech outcomes in this population (Sell et al. 2015). The use of an assessment framework, which could be developed to audit speech outcomes at age three years, may facilitate the improvement of speech outcomes through the earlier identification of children at risk of a poor speech outcome and early intervention.

A protocol assessment framework, which assesses speech outcomes at age three years in the most valid and reliable way may be useful to other cleft centres. Cleft centres in the UK may wish to use the protocol assessment framework for the following reasons:

- To compare speech outcomes at three and five years.
- To measure the effectiveness of therapy intervention.
- To measure the effectiveness of surgical intervention at an earlier stage (which may be particularly helpful for centres working with a new cleft surgeon).
- To validly and reliably report speech outcomes at age three years in research.
- To identify children who may require a different care pathway i.e. more/less regular SLT monitoring.

3) What is the study procedure?

- Three-year-old children with CPL as well as three-year-old control group children will complete different speech samples.
- Participants with CPL will be recruited from the West Midlands Cleft Lip and Palate Service, and the assessment will be carried out at Birmingham Children's Hospital. The speech assessment will be video recorded.
- Cleft SLTs are being invited to participate in the rating of the speech samples using an assessment tool on three occasions four weeks apart. We are asking SLTs to analyse the recordings on separate occasions to calculate intra-rater reliability and compare rating scales.
- In a practise session, Cleft SLTs will be able to complete ratings for two different speech samples using different rating methods and scales. To support listener calibration the Cleft SLTs will be able to compare their ratings to other listeners (each Cleft SLT will only be able to identify their own listening results).

- Cleft SLTs will be able to access the recordings via an NHS approved SharePoint External Data Exchange System.
- Cleft SLTs will analyse different speech parameters which will include using two different methods to rate the speech characteristics associated with velopharyngeal function.
- In addition, both an ordinal scale and visual analogue scale will be used to compare the reliability of these two rating scales.
- Inter-rater and intra-rater reliability will be calculated based upon the ratings made by the Cleft SLTs to determine the reliability of ratings using different assessment methods, scales and based on different speech samples.
- Cleft SLTs will be asked to complete a questionnaire about the listening process and the rating methods they used.

4) Who is organising and running the study?

The study is being run by Beth Fitzpatrick, a Senior Speech and Language Therapist in Cleft Palate. Beth works at Birmingham Children's Hospital for the West Midlands Cleft Lip and Palate Service. Beth is completing the research study as part of a PhD at Coventry University under the supervision of a supervisory team at Coventry University (detailed above).

Speech and Language Therapists who work in Cleft Palate at Birmingham Children's Hospital are carrying out the speech assessments.

5) Who has checked the study?

The research study has been given ethical approval from:

- The NHS- East Midlands- Nottingham 1 Research Ethics Committee (242296)
- Coventry University- Research Ethics team at Coventry University (P66325)
- Birmingham Women's and Children's NHS Trust (18_BC_HNS_NO_134).

Part 2: Your involvement

You are being invited to take part in the study because you are a Cleft SLT working in an NHS cleft centre and are experienced in making judgements about cleft speech using CAPS-A.

To test the reliability of the audit rating tool we will need Cleft SLTs to analyse the participants' recordings and make listener judgements.

Listener judgements will be carried out on three occasions in your own locality viewing videos uploaded on the NHS approved SharePoint External Data Exchange System. Listener ratings will be input into an online form. You will be given a two-week period to complete the first two listening sessions, and approximately one week to complete the shorter final listening session. There will be a four-week period between listening sessions. The overall listening time may be approximately two days. If you plan to carry out the listening during your normal working hours it is advised that you discuss your participation with your Line Manager in advance.

Listener judgements will be measured to determine listener inter-rater and intra-rater reliability. You will be informed of your personal intra-rater reliability score by the Chief Investigator, and you will have the opportunity to discuss the results if you wish.

You will then be asked to complete a questionnaire about the listening process and the rating methods that were used. The questionnaire will be completed online.

Participation in the study is entirely voluntary; you can withdraw from the study at any point until the data analysis has been completed, without giving a reason for doing so. Please be assured that the **information you provide will remain strictly confidential and anonymous**. The results will be reported so that no individual will be identifiable from any publication presenting the results of listening.

At the end of the study, we would like to share our findings with other medical professionals in medical journals. We would also like to present the findings of the study at medical conferences.

Any participant identifiable information will be securely archived for 25 years in line with the policy at Birmingham Women's and Children's NHS Trust. Only a fully anonymised data set will be retained at the end of the study.

We will handle, use, and store data following the Data Protection Act 1998 and the General Data Protection Regulation 2016. **Please see the General Data Protection Regulation information at the end of this leaflet.**

With your consent, we will acknowledge your involvement in the listening in any subsequent publications arising from the study.

We will also give you a £20 Amazon voucher as a token of thanks for completing the listening.

If you decide to take part, we will give you a copy of this information to keep and ask you to sign a Consent Form and return this to the Chief Investigator via email.

For **questions, comments** or if you would like to **contact the team to say that you are interested in taking part** in the study then please contact:

Beth Fitzpatrick (Chief Supervisor)
Speech and Language Therapy
Birmingham Children's Hospital
Steelhouse Lane
Birmingham
B4 6NH
Email: beth.fitzpatrick@nhs.net
Telephone: 0121 3339382

If you have a **complaint** or any **concerns** about the study please contact:

Coventry University:
Professor Olivier Sparagano
Associate Pro-Vice-Chancellor (Research)
Coventry University
Alan Berry Building
Priory Street
Coventry
CV1 5FB
Email: iras-sponsor@coventry.ac.uk
Telephone: 02477659732

Birmingham Women's and Children's NHS Foundation Trust:

Patient Advice and Liaison Team:
0121 333 8961 or 0121 333 9391

If you would be interested in taking part in the study, please email

Beth Fitzpatrick: beth.fitzpatrick@nhs.net, 0121 3339382

Reference:
Sell, D., Mildinhall, S., Albery, L., Wills, A. K., Sandy, J. R., and Ness, A. R. (2015) 'The Cleft Care UK Study. Part 4: Perceptual Speech Outcomes'. *Orthodontics & Craniofacial Research* 18 Suppl 2, 36-46

Please read the General Data Protection Regulation Information:

Coventry University is the sponsor for this study based in the United Kingdom. We will be using information from you in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. All identifiable information about you will be stored for 25 years at the end of the study years *following procedures at Birmingham Women's and Children's NHS Trust. Coventry University will only retain a fully anonymised data set at the end of the study.*

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information by contacting Beth Fitzpatrick (see details above).

Birmingham Women's and Children's NHS Trust will keep your name, and contact details including your cleft team confidential and will not pass this information to Coventry University. Birmingham Women's and Children's NHS Trust will use this information as needed, to contact you about the research study, and to oversee the quality of the study. Certain individuals from Coventry University and regulatory organisations may look at your research records to check the accuracy of the research study. Coventry University will only receive information without any identifying information. The people who analyse the information will not be able to identify you and will not be able to find out your name or contact details.

When you agree to take part in a research study, the information may be provided to researchers running other research studies in this organisation and in other organisations. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research.

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you or to affect your care. It will not be used to make decisions about future services available to you, such as insurance.

Appendix V: Consent Form for SLT Listeners (version 3.4)

The Early Assessment of Speech Outcomes in 3-year-old children with Cleft Palate +/- Cleft Lip (The EASO Study)

	Please initial the box
1) I confirm that I have read and understood the participant information sheet (version number 3.4) for the above study and had the opportunity to ask questions and to have these questions answered satisfactorily	
2) I understand that my participation is voluntary and that I am free to withdraw from the study up until the data analysis takes place	
3) I understand that my information will be treated confidentially and that I will not be identifiable in the research study (unless I give specific consent in question 8 to be acknowledged in any subsequent publications)	
4) I agree to treat all video data confidentially in accordance with the NHS Code of Confidentiality	
5) I understand what my involvement is in the research study	
6) I understand how long the listening may take	
7) I agree to participate in the research study	
8) I would like my contribution to the project to be acknowledged by name in any publications arising from the study	
9) I would like to be informed of the study findings at the end of the study	

Name of participant:……

Signature:…… Date……

Name of the researcher:……

Signature:…… Date……

Appendix W: Constrained Randomisation Example, Week 1

Seed: 147429734191541

Block sizes: 4

Actual List Length: 6

Block identifier, block size, sequence within block, treatment

- **1, 4, 1, Group A**
- **1, 4, 2, Group A**
- **1, 4, 3, Group B**
- **1, 4, 4, Group B**
- **1, 4, 5, Group B**
- **1, 4, 6, Group A**

Appendix X: Assessment Recording Form

(Version 1.1)

KEY:

FC: speech sample fully completed (fully completed, or >90% completed)

PC: speech sample partially completed (started but not completed; >10-90% completed)

NC: speech sample not completed (< 10% completed)

	Sample A		**Sample B**
Participant Number	**Spontaneous Speech**	**The Combined Word List**	**Sentence Repetition**

Appendix Y: Speech Parameters Recording Form

Participant Number	**Control Group (CG) or Participant with CP±L (CPL)**	**Parameter:**		**Parameter:**		**Parameter:**		**Parameter:**		**Parameter:**	
		Adapted CAPS-A	Clinical Assessment	Adapted CAPS-A	Clinical Assessment	Adapted CAPS-A	Clinical Assessment	Adapted CAPS-A	Clinical Assessment	Adapted CAPS-A	Clinical Assessment

Appendix Z: Example Feedback from Listener SLT Practice Listening Session

Please note that I'm not providing a 'right answer' or giving my own judgement on the listening. All I am doing is comparing your judgement to others. Looking at agreement I am taking the overall majority and highlighting any instances in which your judgement differed significantly from the majority.

General Comments

Overall, the practice listening results were very encouraging, particularly with two quite complex participants and speech profiles.

Overall VP Function Rating (categorical): there was good agreement overall particularly as this is a new scale. Please be aware that the category for Incompetent is likely to be wider than that of marginally incompetent. That is, mild but consistent symptoms of VPD would go into incompetent, whilst inconsistent signs of VPD would be rated as marginally competent. Please go with your first impression, and try not to listen for problems.

Competent score should be given if there is no evidence of VPD, or if there are just a couple of signs i.e. you hear a mild symptom once or twice and then not again over the whole sample.

Hypernasality and Passive CSCs: Nasalised consonants (i.e. red) would be associated with a severe score on hypernasality (evident on vowels and voiced consonants). Please double check that your passive ratings and your hypernasality score are in harmony, particularly for the consonant sounds.

VAS Scores: There was good agreement between individual listener's score for hypernasality and VP function ordinal scales and those on VAS.

Phonology Ratings: this varied quite significantly, which is fine because this was the practice! For clarification… For the first participant, 100% put backing to velar. This should then be recorded as disordered phonology. The categories are not mutually exclusive, so you could potentially use: age appropriate, delayed, and disordered. This would indicate that for this participant they had some age appropriate phonological processes, some delayed processes, and a disordered process(es).

Specific Comments

Listener Number X:

Thank you for completing the practice listening. Of course, listening is very subjective, and it is expected that there will be some variability in the results. I have just highlighted a couple of areas where your rating varied from the majority of other listeners. You may wish to have a quick re-listen to the file and bear this in mind. There is no suggestion that your listening is 'wrong' merely that for the study, it would really help if you could consider calibrating your listening in any areas where your judgement was different from the majority.

Example 1:

- Your hypernasality rating varied from the majority, as did audible NAE, so you may wish to re-listen to this again

Example 2:

- Your hypernasality rating varied from the majority who rated this as being present.
- Rating for pharyngeal articulation varied from the majority who rated this as absent.

Appendix AA: Copy of Questions sent to Listener SLTs via the questionnaire on Qualtrics

Thank you for completing the listening sessions. We would be grateful if you could take the time to answer the following questions relating to your listening experience and the assessment tool. This will provide us with information that may help us to improve upon the assessment tool and the listening experience. There is the opportunity for you to add any additional comments either after a specific question or at the end of the section. All comments are welcomed.

Section A: About the speech samples

Sample A = spontaneous speech sample + single word naming

Sample B = short sentence repetition

Which sample did you find it easier to analyse?

- Sample A
- Sample B
- Both
- Neither

Please add any comments as to why you found one sample easier to analyse or not

Which sample would you prefer to use when assessing children at age 3-years with cleft palate?

- Sample A
- Sample B
- Both
- Neither

Please add any comments relating to your preferred speech sample

Which sample do you think the 3-year-old children engaged with the most?

- Sample A
- Sample B
- Both
- Neither

Please add any comments as to why you think the 3-year-old children engaged with a speech sample or not

Which sample do you think the 3-year-old children most fully completed?

- Sample A
- Sample B
- Both

- Neither

Please add any comments relating to completion of the speech samples

Which speech sample most closely matches the speech sample you use in clinical practice when assessing speech in 3-year-old children with cleft palate?

- Sample A
- Sample B
- Both
- Neither

Please add any comments relating to how the speech samples in the study match those you use in clinical practice.

Which speech sample do you think you would be most likely to use in your future clinical practice if available?

- Sample A
- Sample B
- Both
- Neither

Please add any additional comments relating to the speech sample you would use in your future clinical practice.

Section B: About the method and scales used when rating velopharyngeal function for speech

Method 1= composite areas e.g. hypernasality, hyponasality, nasal turbulence, nasal emission

Method 2= overall judgement of speech characteristics associated with velopharyngeal function

Thinking about Method 1, how closely does this match the methods you use to assess the speech characteristics associated with velopharyngeal function when you carry out speech assessments with 3-year-old children with cleft palate in your clinical practice? Please rate on a scale of 0-10, with 0 being *does not match at all*, and 10 being matches *extremely well*.

Does not match at all										Matches extremely well
0	1	2	3	4	5	6	7	8	9	10

Please comment on how Method 1 matches or does not match the methods you use in clinical practice to assess speech in 3-year-old children with cleft palate

Thinking about Method 1, how acceptable to you is this method to assess the speech characteristics associated with velopharyngeal function in 3-year-old

children with cleft palate? Please rate on a scale of 0-10, with 0 being *not at all acceptable*, and 10 being *extremely acceptable*.

Not at all acceptable — Extremely acceptable

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please comment on what made Method 1 acceptable or not

Thinking about Method 1, how easy was it to form judgements using this method? Please rate on a scale of 0-10, with 0 being *not at all easy* and 10 being *extremely easy*.

Not at all easy — Extremely easy

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please comment on what made Method 1 easy or not

Thinking about Method 2, how closely does this match the methods you use to assess the speech characteristics associated with velopharyngeal function when you carry out speech assessments with 3-year-old children with cleft palate in your clinical practice? Please rate on a scale of 0-10, with 0 being *does not match at all*, and 10 being matches *extremely well*.

Does not match at all — Matches extremely well

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please comment on how Method 2 matches or does not match the methods you use in clinical practice to assess speech in 3-year-old children with cleft palate

Thinking about Method 2, how acceptable to you is this method to assess the speech characteristics associated with velopharyngeal function in 3-year-old children with cleft palate? Please rate on a scale of 0-10, with 0 being *not at all acceptable*, and 10 being *extremely acceptable*.

Not at all acceptable — Extremely acceptable

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please comment on what made Method 2 acceptable or not

Thinking about Method 2, how easy or was it to form judgements using this method? Please rate on a scale of 0-10, with 0 being *not at all easy* and 10 being *extremely easy*.

Not at all easy — Extremely easy

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please comment on what made Method 2 easy or not

Thinking about both methods, which method did you prefer in the listening?

- Method 1
- Method 2
- Both
- Neither

Please comment as to what made you prefer a particular method or not

Thinking about the scales you used to rate the participant's speech, which scale did you prefer to use?

- Visual analogue scale
- Nominal scale
- Both
- Neither

Describe the level of your previous experience using Visual Analogue Scales to rate children's speech. Please rate on a scale of 0-10, with 0 being *none*, and 10 being *very experienced*.

None Very experienced

0	1	2	3	4	5	6	7	8	9	10

Describe the level of your previous experience using nominal scales to rate children's speech. Please rate on a scale of 0-10, with 0 being none, and 10 being very experienced.

None Very experienced

0	1	2	3	4	5	6	7	8	9	10

Please add any comments about your scale preference here

Section C: About the listening process

Sample A = spontaneous speech sample + single word naming

Sample B = short sentence repetition

Thinking about Sample A, please describe the amount of time it took you to listen to and analyse this speech sample. Please rate on a scale of 0-10, with 0 being far too long and 10 being far too quick.

Far too long Far too quick

0	1	2	3	4	5	6	7	8	9	10

Thinking about Sample B, please describe the amount of time it took you to listen to and analyse this speech sample. Please rate on a scale of 0-10, with 0 being far too long and 10 being far too quick.

Far too long Far too quick

0	1	2	3	4	5	6	7	8	9	10

Please add any comments about the length of each speech sample and the time it took for analysis

Please describe your thoughts about the adapted CAPS-A used in the study Please tick all that apply.

- Complicated
- Unclear
- Too lengthy
- Clear
- Concise
- Easy to use
- Usable in clinical practice
- Usable in audit
- Not usable in clinical practice
- Not usable in audit
- Age appropriate to use in the analysis of 3-year-old's speech
- Not age appropriate to use in the analysis of 3-year-old's speech
- Too simple for use in the analysis of 3-year-old's speech
- Too complex for use in the analysis of 3-year-old's speech

Please add any additional comments about the adapted CAPS-A

Section D: About the content of the tool

Thinking about the parameters of speech that you analysed using the tool when making judgements about speech, how appropriate were these for the assessment of speech in 3-year-old children with cleft palate? Please rate on a scale of 0-10, with 0 being not at all appropriate and 10 being extremely appropriate.

Not at all appropriate — Extremely appropriate

0 1 2 3 4 5 6 7 8 9 10

Please add any comments regarding why the parameters were appropriate or not

In your clinical practice which parameters of speech do you assess in children with cleft palate at age 3-years? Tick all that apply

- Hypernasality
- Hyponasality
- Nasal Emission
- Nasal Turbulence
- Overall judgement of nasal airflow characteristics

- ☐ Overall judgement of velopharyngeal function for speech
- ☐ Cleft speech characteristics
- ☐ Developmental speech sound processes
- ☐ Phonology
- ☐ Voice
- ☐ Intelligibility
- ☐ Phonetic Inventory
- ☐ Other

Please write any other parameters which you think should be assessed in children with cleft palate at age 3-years

Sample E: About the need for an assessment framework at age-3-years

Please rate how important it is to you that there is a valid and reliable framework to assess speech at age-3-years in the cleft palate population. Please rate on a scale of 0-10, with 0 being not at all important and 10 being extremely important.

Not at all important										Extremely important
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please add any additional comments about the importance or not of using a valid and reliable framework to assess speech at age 3-years in the cleft palate population

Please add any comments about the contribution of a valid and reliable framework to assess speech at age 3-years in the cleft population to your clinical practice

Please rate how likely you would be to use a valid and reliable assessment framework to measure speech outcomes at age three-years in the cleft palate population. Please rate on a scale of 0-10, with 0 being not at all likely and 10 being extremely likely.

Not at all likely										Extremely likely
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please add any additional comments relating to how likely you would be to use a valid and reliable assessment framework to measure speech outcomes at age three-years in the cleft palate population

Please add any additional comments you wish to make here

Thank you for taking the time to complete this survey, and for all your help in completing the listening.

Appendix BB: Listener SLT Free Text Questionnaire Responses

Question 2: Please add any comments as to why you found one sample easier to analyse or not.

- Greater sense of resonance from phrase level sample
- I found it slightly easier but mainly because it was quicker which made it less tiring. It took a while to get used to which sounds I was listening for but after a few children this eased. The length of the spontaneous speech made sample A harder to focus on. Especially without the SLT edited out.
- I'm used to listening to the GOS.SP.ASS sentences for CAPS-A listening so I'm not sure if this has influenced the way I listen and what I find easier to listen to?

Question 4: Please add any comments relating to your preferred speech sample.

- Unless the child is a confident communicator, the spontaneous speech sample can be frustrating/time consuming to assess (B)
- I think you gather relevant info from both samples. In terms of consensus listening I would rather use sample B. (B)
- It seemed to be quicker for clinical purposes and you can also listen to resonance too (B)
- I'd be happy to use either (Both)

Question 6: Please add any comments as to why you think the 3-year-old children engaged with a speech sample or not.

- Pictures for both seemed equally engaging; perhaps lower demand from single word naming, but I didn't notice huge difference. (Both)
- Single word naming and sentence repetition engagement seemed quite equal to me, i think the children were less engaged with the spontaneous speech sample (B)
- I was surprised at how much the children participated in both samples (Both)
- There were a lot of pictures for posting on the single word sample so I think it might have been easier for them to comment on the short sentences (B)

Question 8: Please add any comments relating to completion of the speech samples.

- Not hugely different, single word perhaps slightly easier? (A)
- I think it depended on the individual child. Some found it hard to engage in spontaneous speech but the single words were ok. The sentences appeared ok for most to copy (Both)
- wonder if single words were easier for some children with reduced language skills?? However, it felt like there were a lot of pictures for the single naming task, so it may be difficult to hold their attention (A)
- if child has delayed language skills then the single word picture naming would be more appropriate (A)

Question 10: Please add any comments relating to how the speech samples in the study match those you use in clinical practice.

- These were more child focused, using vocabulary likely to be more familiar to 3 year olds. (B)
- I don't generally use a spontaneous speech sample although I think its actually a really useful thing to carry out. I chose single words OR sentence rep dependent on the level of ability and cooperation of the child (Both)
- This varies depending on the language skills of the child. I will try to hear some words initially to get the child warmed up and to listen for oral pressure sounds and get an idea of their speech patterns. However, if they are able to repeat sentences, I will try to obtain a short sentence repetition task (Both)

Question 12: Please add any additional comments relating to the speech sample you would use in your future clinical practice.

- I would feel more comfortable using phrase level. Spontaneous speech would be part of wider assessment/judgment (B)
- Probably sample B as it seemed quicker to analyse speech and resonance and may be more effective for clinical practice. Depending on the purpose - it can be useful to hear some single words to get an idea of developmental speech patterns and CSCs but you can pick this up from the short sentences and interestingly the speech of 3 year olds can be very inconsistent (B)

Question 14: Please comment on how Method 1 matches or does not match the methods you use in clinical practice to assess speech in 3-year-old children with cleft palate.

- Similar principles (10)
- I tend to assess hypernasality separately from emission and turbulence (9)
- Sometimes the method we use depends on the child, if it is not appropriate we use a method similar to method 2 (8)

Question 16: Please comment on what made Method 1 acceptable or not.

- clear descriptions and parameters (8)
- Its how I am used to doing it but I think it could be too specific for rating a 3 year old and not necessary to be so detailed (5)
- providing the sample was good enough (9)

Question 18: Please comment on what made Method 1 easy or not.

- Ease depended on quality of the sample (7)
- Too much to rate with limited speech samples and inconsistency is tricky too (6)
- Similar to what I use already (9)

Question 20: Please comment on how Method 2 matches or does not match the methods you use in clinical practice to assess speech in 3-year-old children with cleft palate.

- Frames it differently but matches well (10)
- I do make an overall judgement whether I think there is VPI or a risk of this. I don't tend to 'rate' this (4)
- If possible, we try to obtain some additional information about articulation and cleft patterns and more detail in relation to resonance and NACs. Method 2 is the

impression we want to achieve by the end of our assessment, whether the child's VP function is competent or not. In order to pass on information to local SLTs, it is good to have more detail so you can discuss therapy ideas with them going forward. It can be difficult to make decisions about VP competency if the child is not using any oral pressure consonants or if they are quite inconsistent in productions which makes this scale more difficult to use. By teasing out more details in relation to our usual method looking at articulation patterns, resonance and NACs, it can inform management decisions. Children may be incompetent on this scale, but after diagnostic therapy, they may be able to achieve oral pressure sounds. I would describe this scale as the summary or overall impression we reach by the end of our assessment. It lacks details in relation to articulation and stimulability. It is useful to know if a child is hypernasal or if they have nasal emission on anterior oral pressure consonants as the latter may be a result of a fistula, etc. (5)

Question 22: Please comment on what made Method 2 acceptable or not.

- Links symptoms to palate function more directly, including rating of inconsistent/consistent features. (10)
- This method feels a little rudimentary in nature. It gives an overall impression, but lacks the finer details. I tend to use this scale in my own head when listening to the 2 year olds we see in clinic to get an impression about whether their VP mechanism appears to be competent or not. But many 3 year olds can sit and do quite a detailed speech assessment which allows us to make future decisions regarding therapy or further investigations of palate function. (4)
- I think at 3 years (particularly the 3 - 3;6 age group) it is more useful and valid to comment on the overall judgement considering all factors rather than separating them out. (8)

Question 24: Please comment on what made Method 2 easy or not.

- Think it would become easier with more familiarity (8)
- In my own head, I think I'm still using Method 1 and listening out for CSCs, resonance, NACs before completing this rating. Once you have listened to the sample, you can make an overall impression, but it can be difficult to rate due to inconsistencies in the productions of 3 year olds. (6)
- Less detailed, (7)

Question 26: Please comment as to what made you prefer a particular method or not.

- I think I am used to listening using Method 1 and thinking about resonance, NACs, CSCs, etc. It gives more detail. Method 2 feels like a summary or impression following your assessment. (method 1)
- The link to palate function (2)
- More precise and useful diagnostically (1)

Question 30: Please add any comments about your scale preference here.

- I would need more experience to feel confident about VAS and more consensus practice.

- I think I am so used to using nominal scales, this feels more familiar. I have no experience of using VAS and may need some calibration!!!
- My preference definitely relates to my experience and therefore confidence is using the scales.

Question 33: Please add any comments about the length of each speech sample and the time it took for analysis

- Varied depending on the child re spontaneous sample
- Sample A seemed to take a long time as the children often said very little spontaneous speech, but you had to listen to all of this section. The single word speech sample felt quite long and sometimes it was difficult to interpret a sound and then I would play it back and hear something else!! Again, I am not sure if I am more used to listening to sentences for CAPS-A 5 year audits, but I found the sentences in Sample B much quicker to analyse.

Question 41: Please add any additional comments about the importance or not of using a valid and reliable framework to assess speech at age 3-years in the cleft palate population.

- 3 is a clinically useful time point - allowing time for SLT intervention ahead of school start.
- I am sure this study will reveal just how co-operative our 3 year olds are and how they can sit and do a detailed speech assessment. It would be useful to have a valid and reliable speech framework at 3 years as this could be compared to their CAPS-A at 5 years. It can inform management earlier on and this may be a better age for assessing children's speech and carrying out audit? It would be interesting to see comparisons between their speech at 3 and 5 years. It may inform which children need more intervention and what other factors influence progress with speech, etc.

Question 42: Please add any comments about the contribution of a valid and reliable framework to assess speech at age 3-years in the cleft population to your clinical practice.

- Few children can complete the GOS.SP.ASS at 3, using the picture stimuli, so assessment can be very variable between SLTs and centres. It is a much needed framework.
- It would be invaluable to have a valid and reliable tool for 3 year olds to help with management decisions and comparisons across 3 year olds. It would inform therapy plans and recommendations and help make decisions about whether diagnostic therapy is needed or further investigations.

Question 44: Please add any additional comments relating to how likely you would be to use a valid and reliable assessment framework to measure speech outcomes at age three-years in the cleft palate population.

- Our Centre would certainly use.

- We have looked at 5 year outcomes for quite a long time now and it would be interesting to look at the speech of children with cleft palate at a younger age and see if we can detect any further patterns and how outcomes change over time with/without intervention, etc.

Question 45: Please add any additional comments you wish to make here.

- Really interesting piece of research! It would be beneficial to formalise an assessment framework and have a valid and reliable tool for assessing 3 year olds to allow for comparisons across centres and to look at trends over time and to see if some CSCs resolve on their own or need more intervention, etc.

Appendix CC: Summary of Inter-Rater Reliability results for studies used for comparison

(grouped by CAPS-A, studies assessing 3-year olds, and those using VAS)

Study	Age of participants	No Listeners	Scale	Statistics	Hypernasality	Hyponasality	NAE	VPC-Rate	Anterior Oral CSCs	Posterior Oral CSCs	Non Oral CSCs	Passive CSCs	Phonology / Non-cleft
Ogata et al. (2022)	4-7	6	CAPS-A Ordinal	ICC Single measures	.67	.08	Combined measure: NE .45	NA	NA	NA	.80	.76	.38
Bruneel et al. (2020)	3-10 (mean 6.5 years)	Study 1: 2 Study 2: 4	CAPS-A Ordinal	ICC Single measures, % agreement	Study 1: .75, 55% Study 2: .69, 56.7%	Study 1:.49, 90% Study 2: .85, 83.3%	NE Study 1: .51, 60% NT Study 1: .90, 85% NE Study 2: .36, 63.3% NT Study 2: .58, 60%	NA	Study 1: .10, 40% Study 2: .10, 40%	Study 1: .70, 70% Study 2: .56, 55%	Study 1: .73, 75% Study 2: .73, 63.3%	Study 1: .64, 60% Study 2: .45, 43.3%	NA NA
Baillie & Sell (2020)	4.6-7.1 years (mean 5.2 years)	2	CAPS-A Ordinal	Weighted Kappa, % agreement	.70-.72, 91.8-98/4%	NA	NE: -.03, 95.9% NT: .61, 92.5-96.8%	NA	Palatal: .33, 86.6% Lateral: .76, 98.1%	Backing to velar/uvular: .77, 98.1%	NA	NA	NA
Chapman et al. (2016)	5-7 years	Study 1: 9 Study 2: 6	CAPS-A Ordinal	Kappa, Weighted Kappa	Study 1 Inter: wKappa: .70-.82 Study 2 Inter: wKappa: .70-71	Study 1 Inter: wKappa: .25-.50 Study 2 Inter: wKappa: .39-.67	NT Study 1 Inter: wKappa: .51-.68 NE Study 1 Inter:	NA	Study 1 Inter: wKappa: .34-.51 Study 2 Inter:	Study 1 Inter: wKappa: .62-.69 Study 2 Inter:	Study 1 Inter: wKappa: .65-.73 Study 2 Inter:	Study 1 Inter: wKappa: .66-.75 Study 2 Inter:	Study 1 Inter: Kappa: .16-.28 Study 2 N/A

							wKappa: .43-.72 Study 2 Combined Inter: wKappa: .53-.71		wKappa: .38-.45	wKappa: .00-.06	wKappa: .60-.78	wKappa: .55-.72	
Sell et al. (2015)	5 years (mean 5.5 years)	2	CAPS-A Ordinal	Weighted Kappa, % agreement	.60, 92%	.67, 95%	NE: .46, 84% NT: .69, 90%	NA	.51, 81%	.54, 90%	.36, 88%	.60, 95%	.30, 62%
Sell et al. (2009)	5-7 years	36	CAPS-A Ordinal	ICC Average measures	.72	.72	NE:.58 NT: .69		.22	.70	.69	.81	.35
John et al. (2006)	5-10 years	7	CAPS-A Ordinal	ICC Single measures	Study 1: .51 Study 2: .88	Study 1: .71 Study 2: .64	NE Study 1: .16 NE Study 2: .64 NT Study 1: .18 NT Study 2: .67	NA	Study 1: .47 Study 2: .84	Study 1: .60 Study 2: .34	Study 1: .91 Study 2: .85	Study 1: .74 Study 2: .85	Study 1: .39 Study 2: .69
Persson et al. (2022)	3	3	VPC-Rate ordinal					80% same language 90-87% across languages					
Jørgensen, L. D. and Willadsen, E. (2020) Willadsen et al. (2018)	2.83-3.23 years	4	PCC-A	Point by point transcription agreement	No inter-rater as consensus listening used								

Chacon et al. (2017)	Group 1: 2;10-3;11 years Group 2: 4;10-5;09 years	Inter-rater: 4 Intra-rater: 1		Point by point transcription agreement					92.9% for articulation
Larsson et al. (2017)	35-43 months	Inter: 2 Intra rater: 1	PCC-A, % consonants produced with audible nasal air leakage	Point by point percentage agreement			Combined: 72-100%		Total agreement: 33-73% Place of articulation: 58-91% Manner of articulation: 71-100%
Lohmander et al. (2017a)	5 years	12-17	Scandcleft Ordinal scale, transcription	Pont by point agreement, Weighted kappa	Good: 50% Moderate: 40% Fair: 10%				84-89%
Lohmander et al. (2017c)	5 years	10	Ordinal scale, composite scale	Weighted kappa				2 rates of same language: .55 2 rates speaking different language: .33	
Raud Westberg et al. (2017)	3 years	2	PCC-A, Ordinal scales	Point by point percentage agreement	80%	88%	Combined: 88%	Perceived VP function: 80%	50-70% (mean 90%)
Klintö et al. (2016) Klintö et al. (2015b) Klintö et al. (2014a)	3 years	2	PCC-A	Point by point percentage agreement					UCLP Group: 62-80%

Klintö et al. (2014b)													
Klintö et al. (2011)	5 years	2	PCC (variations of PCC)	Point by point percentage agreement					Word naming: 47-95.2%, median 89.7% Sentence Repetition: 69.0-98.9%, median 86.0% Bus Story: 52.9-94.2%, median 79.5% Conversational speech: 50.3-93.9% median 86.2%				
Lohmander et al. (2009)	5 years	5 from 5 different language backgrounds	Point by point agreement for consonants, Ordinal scale for other variables assessed	% agreement	Only intra-rater reliability reported								
Lohmander & Persson (2008)	3-7 years	3	PCC Ordinal scales for hypernasality and nasal air leakage	Point by point agreement, Spearman's rho	39%		Combined: 85%		rho= .91-.97				
Persson et al. (2006)	3-10 years	3	Ordinal scale Transcription for articulation	Percentage agreement Articulation: point by point agreement	55-44%	N/A	Combined: 57-59%	VP Impairment: 45-58%	N/A	Retracted oral articulation: 83-84%	Glottal articulation: 91-93%	Weak pressure consonants: 67-75%	N/A
Bettens et al. (2018)	4-15	4	CAPS-A Ordinal Scale VAS	ICC Average: Inter Single: Intra	Ordinal: .82 VAS: .87		Combined Ordinal: .71 VAS: .74						
Castick et al. (2017)	NA	5	CAPS-A Ordinal Scale VAS	ICC	Ordinal: .844 VAS: .821	Ordinal: .636 VAS: .314	NE Ordinal: .820 VAS: .755 NT Ordinal: .822						

							VAS: .763						
Bayliss et al. (2015)	5-6 years	5	CAPS-AM Ordinal Scale VAS	ICC Weighted Kappa	Ordinal: ICC: .776 Kappa: .436 VAS: ICC: .982 Kappa: .534		Combined Ordinal: ICC: .626 Kappa: .289 VAS: ICC: .969 Kappa: .716						

Appendix DD: Summary of Intra-Rater Reliability results used for comparison

(grouped by CAPS-A, studies assessing 3-year olds, VPC-Rate, and those using VAS)

Study	Age of participants	No Listeners	Scale	Statistics	Hypernasality	Hyponasality	NAE	VPC-Rate	Anterior Oral CSCs	Posterior Oral CSCs	Non Oral CSCs	Passive CSCs	Phonology / Non-cleft
Ogata et al. (2022)	4-7	6	CAPS-A Ordinal	ICC Single measures	.83-.96	0.0-1.0	.24-97	NA	NA	NA	.84-99	.68-.99	.40-1.00
Bruneel et al. (2020)	3-10 (mean 6.5 years)	Study 1: 2 Study 2: 4	CAPS-A Ordinal	ICC Single measures, % agreement	Study 1: .83-.87, 60-70% Study 2: .52-.83, 50-70%	Study 1: 1.00, 100% Study 2: .80-1.00, 90-100%	NE Study 1: .82-.83, 70-80% NT Study 1: .77-.84, 70-80% NE Study 2: .65-.82, 60-80% NT Study 2: .66-1.00, 50-100%	NA	Study 1: .62-.89, 70-90% Study 2: .53-.93, 70-90%	Study 1: .92-1.00, 70-90% Study 2: .45-.75, 70%	Study 1: .84-1.00, 80-100% Study 2: .53-.94, 60-90%	Study 1: .74-1.00, 70-100% Study 2: .70-.83, 43.3%	NA NA
Baillie & Sell (2020)	4.6-7.1 years (mean 5.2 years)	2	CAPS-A Ordinal	Weighted Kappa, % agreement	.38, 91.8%	NA	NE: .10-.46, 96.8-97.6% NT: .70-.87, 92.5-96.8%	NA	Palatal: .55-.79, 90.7-98.1% Lateral: .49-.93, .90-.94%	Backing to velar/uvular: .90-.94, 98.4-99.6%	NA	NA	NA
Chapman et al. (2016)	5-7 years	Study 1: 9 Study 2: 6	CAPS-A Ordinal	Kappa, Weighted Kappa	Study 1 wKappa: .84 Study 2 wKappa: .85	Study 1 wKappa: .70 Study 2 wKappa: .62	NT Study 1 wKappa: .77 NE	NA	Study 1 wKappa: .60 Study 2 wKappa: .81	Study 1 wKappa: .81 Study 2 wKappa: .46	Study 1 wKappa: .84	Study 1 wKappa: .81 Study 2 wKappa: .78	Study 1 Kappa: .62 Study 2: N/A

							Study 1 wKappa: .70 Study 2 NT wKappa: .75				Study 2 wKappa: .85		
Sell et al. (2009)	5-7 years	36	CAPS-A Ordinal	% agreement	**90-100% agreement:** 88.9	**90-100% agreement:** 91.7%	**90-100% agreement:** NE: 5.8% NT: 66.7%	NA	**90-100% agreement:** 36.1%	**90-100% agreement:** 80.6%	**90-100% agreement:** 63.9%	**90-100% agreement:** 77.8%	**90-100% agreement:** 13.9%
Sell et al. (2015)	5 years (mean 5.5 years)	2	CAPS-A Ordinal	Weighted Kappa, % agreement	.69-1.00, 93-100%	.46-1.00, 88-100%	NE: .85-.87, 96% NT: .73-.83, 90-94	NA	.66-.89, 90-98%	.84-.89, 98%	.71-.77, 96-99%	-	.58-.81, 82-92%
John et al. (2006)	5-10 years	7	CAPS-A Ordinal	ICC Single measures	Study 1: .43 Study 2: .62	Study 1: .21 Study 2: .73	NE Study 1: .51 NE Study 2: .69 NT Study 1: .41 NT Study 2: .61	NA	Study 1: .57 Study 2: .64	Study 1: .37 Study 2: .53	Study 1: .68 Study 2: .83	Study 1: .54 Study 2: .80	Study 1: .84 Study 2: .76
Persson et al. (2022)	3	3	VPC-Rate ordinal					94% same language					
Jørgensen, L. D. and Willadsen, E. (2020) Willadsen et al. (2018)	2.83-3.23 years	4	PCC-A	Point by point transcription agreement	No inter-rater as consensus listening used				88.2-97.5%				
Chacon et al. (2017)	Group 1: 2;10-3;11 years	Inter-rater: 4		Point by point transcriptio					96.5% for articulation				

	Group 2: 4;10-5;09 years	Intra-rater: 1		n agreement					
Larsson et al. (2017)	35-43 months	Inter: 2 Intra rater: 1	PCC-A, % consonants produced with audible nasal air leakage	Point by point percentage agreement			Combined: 72-100%		Total agreement: 57-90% Place of articulation: 71-100% Manner of articulation: 72-100%
Lohmander et al. (2017a) Willadsen et al. (2017)	5 years	12-17	Scandcleft Ordinal scale, transcription	Pont by point agreement, Weighted kappa	Very good: 42% Good: 54% Moderate: 4%				82-96%
Raud Westberg et al. (2017)	3 years	2	PCC-A, Ordinal scales	Point by point percentage agreement	NA	NA	NA	NA	86-89%
Klintö et al. (2016) Klintö et al. (2015b) Klintö et al. (2014a) Klintö et al. (2014b)	3 years	2	PCC-A	Point by point percentage agreement					UCLP Group: 79-92%
Klintö et al. (2011)	5 years	2	PCC (variations of PCC)	Point by point percentage agreement					Word naming: 91.0-98.8%, median 97.5% Sentence Repetition: 93.5-100%, median 94.9% Bus Story: 70.6-99.4%, median 88.3% Conversational speech: 85.7-97.3% median 93.7%
Lohmander et al. (2009)	5 years	5 from 5 different language backgrounds	Point by point agreement for consonants, Ordinal scale for	% agreement	100% on connected speech Single words; 25%-100%	NA	NA	100%	Phonetic transcriptions: 27-88%

			other variables assessed										
Lohmander & Persson (2008)	3-7 years	3	PCC Ordinal scales for hypernasality and nasal air leakage	Point by point agreement, Spearman's rho	60-70%		Combined: 60-70%		rho= .97-.98				
Persson et al. (2006)	3-10 years	3	Ordinal scale Transcription for articulation	Percentage agreement Articulation: point by point agreement	92-58%	N/A	Combined: 72-88%	VP Impairment: 77-96%	N/A	Retracted oral articulation: 88-100%	Glottal articulation: 96-100%	Weak pressure consonants: 81-100%	N/A
Bettens et al. (2018)	4-15	4	CAPS-A Ordinal Scale VAS	ICC Average: Inter Single: Intra	Ordinal: .63-.95 VAS: .42-.93		Combined Ordinal: .64-1.00 VAS: .92-.96						
Castick et al. (2017)		5	CAPS-A Ordinal Scale VAS	ICCs Spearman's rho	Ordinal: .806 VAS: .856	Ordinal: .621 VAS: .796	NE Ordinal: .835 VAS: .875 NT Ordinal: .851 VAS: .833						
Bayliss et al. (2015)	5-6 years	5	CAPS-AM Ordinal Scale VAS	ICC Weighted Kappa	Ordinal: ICC: .396 Kappa: .471 VAS: ICC: .429 Kappa: .499		Combined Ordinal: ICC: .467 Kappa: .289 VAS: ICC: .564 Kappa: .740						