Title: Development of an Agreed Labelling System and Protocol for the Diagnosis of Speech Sound Disorder Subtypes in the United Kingdom

Cleland, J[1]., Burr, S.A.[2], Harding, S.A.[2,3], Stringer, H.[4], Wren, Y.[2,5,6]

Corresponding Author:

Dr Joanne Cleland

[1]Psychological Sciences and Health, Graham Hills Building, University of Strathclyde, Glasgow, Scotland

Email: joanne.cleland@strath.ac.uk

ORCiD: 0000-0002-0660-1646

Dr Sam Burr

[2]Bristol Speech and Language Therapy Research Unit, North Bristol NHS Trust, Southmead Hospital, Bristol, BS10 5NB

[3]North Bristol NHS Trust, Southmead Hospital, Bristol, BS10 5NB

Email: sam.burr@nbt.nhs.uk

ORCiD: 0000-0002-4177-2459

Dr Sam Harding

[2]Bristol Speech and Language Therapy Research Unit, North Bristol NHS Trust, Southmead Hospital, Bristol, BS10 5NB

[3]Research and Innovation, North Bristol NHS Trust, Southmead Hospital, Bristol, BS10 5NB

Email: sharding.jb@gmail.com

ORCiD: 0000-0002-5870-2094

Dr Helen Stringer

[4]King George VI Building, Newcastle University, Queen Victoria Road, Newcastle upon Tyne, NE1 7RU

Email: helen.stringer@ncl.ac.uk

ORCiD: 0000-0002-7470-2166

Dr Yvonne Wren

[2]Bristol Speech and Language Therapy Research Unit, North Bristol NHS Trust, Southmead Hospital, Bristol, BS10 5NB

[5] Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, Llandaff Campus, Western Avenue, Cardiff, Wales CF5 2YB

[6] Bristol Dental School, University of Bristol, Lower Maudlin St, Bristol, UK, BS1 2LY Email: Email: Yvonne.Wren@Bristol.ac.uk

ORCiD: 0000-0002-1575-453X

Contact: Joanne Cleland, joanne.cleland@strath.ac.uk

Abstract

Purpose: This study aimed to achieve a consensus on a diagnostic protocol, classification system, and subtype definitions for the differential diagnosis of speech sound disorder of unknown origin in the United Kingdom.

Method: A mixed methods participatory design was used. Five services from the UK provided all paperwork, including guidelines and care pathways, related to Speech Sound Disorder for content analysis. Two participatory workshops with six speech and language therapists from these five services were used to discuss and agree: 1. A classification system, 2. Subtype labels and definitions, and 3. A diagnostic protocol for speech sound disorder suitable for use in the UK context.

Result: Participants agreed that the Differential Diagnosis System (Dodd, 2014) was suitable for use in the UK. This system comprises five speech sound disorder subtypes. Participants suggested minor changes to the definitions of these subtypes to make them more suitable for implementation in a clinical context. A minimum diagnostic protocol, with additional assessment for children with more complex or severe speech sound disorder, was agreed.

Conclusion: A consensus diagnostic protocol, classification system, and subtype names and definitions was reached and is broadly in line with Dodd (2014). Future work will implement this in the national health service in the UK.

Key Words: Speech Sound Disorder, diagnosis, assessment, participatory design.

Introduction

Speech Sound Disorder (SSD) is an umbrella term for any difficulty acquiring the sounds of the ambient language. SSDs are many and varied, ranging from difficulty with only one or two speech sounds, to severely unintelligible speech. SSD is a theory-neutral term, likely comprising subtypes that arise from different aetiologies and with different behavioural manifestations requiring different treatment approaches. The clinical management of SSDs involves two main clinical reasoning processes on the part of the Speech and Language Therapist (SLT) (Diepeveen et al., 2020). First, a diagnostic process designed to determine whether a child presents with an SSD, and if so what subtype; and secondly therapeutic reasoning to decide which intervention, if any, to choose. These processes are linked. Most interventions are designed by their originators to treat specific subtypes of SSD (Wren et al., 2018). For example, (conventional) minimal pair intervention was originally designed for children who display a loss of phonemic contrast in their speech, usually known as a phonological SSD subtype (McLeod & Baker, 2017). However, there is a lack of consensus in the literature on labelling and indeed on clear descriptors for any labels. While some SSD are attributable to medical or genetic conditions, e.g. cleft lip +/- palate, cerebral palsy, Down syndrome, the majority, to date, remain of unknown origin (Shriberg et al., 2010). These are the most commonly referred to SLT services (Broomfield & Dodd, 2004) and are considered here. In their textbook on SSD, McLeod and Baker (2017), list 41 different terms used in the literature to describe SSDs of unknown origin. They classify these as overarching terms (e.g., SSD) and further subclassification into phonology (e.g. phonological disorder) and motor speech (e.g. articulation impairment) terms. However, even within the two subcategories of phonology or motor, it is not possible to know if terms are used synonymously by clinicians. Studies asking clinicians which labels they use typically do not also ask them to give information about how they operationalise these labels or which labels they consider synonymous. For example, a recent mixed-methods study (Diepeveen et al., 2020) recorded 35 different terms used by SLTs in the Netherlands to describe SSDs, but the SLTs were not asked to give precise definitions for the terms they use, making it difficult to determine whether or not terms like “phonological impairment” are

synonymous with “phonological disorder” or “phonological delay”. This lack of consensus on labels is problematic for clinicians, researchers, and indeed parents and carers. There is therefore a need to agree labels that have consistent definitions that clinicians understand and can implement in their own clinical practice.

Although there are many different terms used to describe SSDs, there are fewer classification systems. Waring and Knight (2013) provided a review and critical evaluation of three commonly used paediatric-specific SSD classification systems (Waring and Knight, 2013): The Speech Disorder Classification System (SDCS) (Shriberg et al., 2010); the Differential Diagnosis System (Dodd, 2014) and the Psycholinguistic Framework (Stackhouse & Wells, 1993). Of these, the Differential Diagnosis System (often referred to as Dodd’s system) and the SDCS are most widely used internationally (Terband et al., 2019). Dodd’s system is based on a psycholinguistic model of speech production and incorporates the subtype labels: phonological delay; consistent atypical phonological disorder; inconsistent phonological disorder; articulation disorder; and Childhood Apraxia of Speech (CAS- sometimes known as Developmental Verbal Dyspraxia or DVD). In contrast, the SDCS is an aetiology-based system including the terms: speech delay-genetic; speech delay-otitis media with effusion; speech delay- developmental psychosocial involvement; speech errors-/s/; speech errors- /r/; motor speech disorders- dysarthria; motor-speech disorders-CAS; and “speech motor delay”. Notably absent from Dodd’s system is the motor speech disorder label “dysarthria”. This is because her system focuses on SSD of unknown origin and in most cases of dysarthria a cause is known, for example cerebral palsy.

Use of additional terms outside these two classification systems is likely due to idiosyncrasies in naming conventions adopted by individual SLTs or the teams they work in (Diepeveen et al., 2020). For example, an SLT may prefer to use the term “phonological impairment” to “phonological disorder” because they prefer the word “impairment”. In contrast, it is likely that choice of which classification system to use differs geographically and is influenced by training programmes in each country. Within the UK, the professional body, the Royal College of Speech and Language Therapists, sets curriculum

guidelines for higher education institutions (RCSLT, 2021). Although these guidelines specify that subtyping of SSD should be on the curriculum, no particular classification system is mandated. However, anecdotal evidence suggests that many SLTs in the UK use Dodd's (2014) classification system and the terms used within the curriculum guidelines are more closely aligned with these than those in the SDCS (Shriberg et al., 2010).

SSD assessment procedures

SSD classification system selection is important because it can influence a clinician's choice of procedure for assessment and subsequent choice of intervention. For example, assessment based on the SDCS might look for acoustic or genetic markers to support diagnosis (Shriberg & Wren, 2019). The process of differential diagnosis using Dodd's system, might involve the use of Dodd's assessment designed specifically for this purpose: the Diagnostic Evaluation of Articulation and Phonology (Dodd et al., 2002). This involves an initial screening process, the results of which guide the SLT to select specific subtests to diagnose specific subtypes of SSD (Dodd, 2014). However, this assessment is designed for and standardised on English speakers, therefore SLTs assessing speakers of other languages will need to use either standardised tests developed for their own language, or self-designed measures while incorporating the principles of Dodd's system. Alongside these approaches, SSD can also be assessed within the International Classification of Functioning, Disability and Health (ICF), which considers the wider context of a child's day-to-day experience of SSD, including their activity, participation and impairment (McLeod & McCormack, 2007). Moreover, there has been an increasing focus on assessing the impact of SSD on the child using patient reported outcome measures including measures of quality of life (Cohen, 2020).

A thorough SSD assessment is therefore multifaceted. Macrae (2016) suggests that assessment should include standardised single-word testing, additional single-word testing designed to look at each child's difficulties in-depth, a connected speech sample, stimulability testing, and an assessment of inconsistency. Similarly, the Child Speech Disorder Research Network suggest that

speech samples include single-word testing of at least 100 words, connected speech sampling, and, again, an additional wordlist designed to look at a child's difficulties in detail. Surveys of SLTs suggest that some, but not all, of these are included by clinicians in their diagnostic toolbox. For example, in a survey of 333 SLTs in the USA, Skahan et al. (2007) reported that clinicians used standardised single-word tests; stimulability testing; and an estimate of intelligibility. In Australia, 231 SLTs surveyed by McLeod and Baker (2014) also reported using single-word tests, connected speech samples, stimulability testing and an estimate of intelligibility.

Diepeveen et al. (2020) suggest that SLTs main motives for deciding what to include during diagnostic testing are time and ease of use. Likewise, familiarity with a system and its subtype labels is also likely to make it more useable. Although an international consensus on classification and a diagnostic protocol is desirable (Waring & Knight, 2013), this study focuses on establishing a classification system and diagnostic protocol acceptable and feasible for SLTs working within the publicly funded, free at point of access, National Health Service (NHS) in the UK. We limit our study to the UK for several reasons. Firstly, time available to undertake assessment and ease of use/familiarity will differ between healthcare systems. Secondly, specific assessments are likely to be standardised on specific populations and with speakers of specific languages; and lastly, the UK NHS service is unique in that there is potential to collect large scale data from every patient in the UK that accesses the service, enabling us to, in the future, answer questions about treatment effectiveness in children with SSDs once we have agreed on subtype definitions.

To ensure a classification system and diagnostic protocol is fit for purpose within this context, it must be co-designed with the clinicians and services who are going to use it. We therefore employed a participatory design, focused on involving end users of the SSD subtype labels and diagnostic protocols (Roper & Skeat, 2022) which in this case are SLTs, with the recognition that any labels used should be sensitive to the needs of parents and carers and any diagnostic protocol should be acceptable to both parents and children.

Aims

This study aimed to achieve a consensus on which classification system for differential diagnosis is most appropriate for use in the UK. Specifically, we had the following objectives:

1. To agree on an SSD classification system which reflects current clinical practice in the UK, supports the clinical decision-making process, and is appropriate for large scale adoption.
2. To agree definitions for each subtype of SSD which are clinically relevant for the UK context.
3. To agree a diagnostic protocol for categorising children with SSD into these subtypes which is feasible for the publicly funded NHS in the UK.

Method

Participatory Design

Participatory designs democratise the development process by involving the end users of the product or system (Roper & Skeat, 2022). This contrasts with the traditional approach of experts, usually researchers, designing a classification system based on their own theoretical position; or researchers suggesting diagnostic protocols without considering feasibility within a clinical context. This more traditional approach can lead to problems implementing research because clinicians have different challenges and priorities to researchers (Douglas et al., 2023). Participatory designs can lead to quicker implementation in clinical practice. This study used mixed methods. Firstly, a content analysis of any local paperwork/guidelines from different SLT teams within the UK on how they classify and diagnose SSDs, and secondly two participatory workshops with SLTs representatives from these teams and a parent of a child with an SSD.

Ethics

Ethical approval for this study was obtained from the study Sponsor, [redacted for peer review]. Full approval for the study was obtained from the NHS Health Research Authority (22/HRA/1962) prior to recruitment.

Participants

SLT teams were recruited from NHS providers (i.e., services) who expressed an interest in participating in a larger project designed to establish the most effective care pathways for children with SSD. Five NHS services in England (n=4) and Scotland (n=1) took part in providing local service relevant paperwork, and six individual SLTs from these five NHS services consented to participate in the workshops. The participants came from services that employed between 11 and 115 full-time or part-time SLTs. SLTs were selected to represent their service by their managers because they had specific responsibility or expertise in SSD. One service put forward two SLTs to take part in the workshops as both SLTs had a particular remit for SSD but worked in different teams. A parent of a children with severe SSD, and member of the project patient public involvement group, joined the calls as an expert by experience to support the study.

Content Analysis of Local Paperwork and Guidelines

Prior to the workshops, participants were asked to submit via email any diagnostic criteria, care pathway information or clinical decision-making resources used by SLTs working with children with SSD. These documents were analysed to identify:

1. Which, if any, classification systems were teams using?
2. How, if at all, were the subtypes of SSD described and defined?
3. Which assessments (published or locally designed) were the teams using and how, if at all, did they lead to diagnosis of specific subtypes of SSD?

Content was tabulated and similarities and differences identified to be presented for discussion at the participatory workshops.

Participatory Workshops

Two three-hour workshops were held using the online meeting platform Microsoft Teams. The workshops were held one week apart to allow participants to reflect on the discussions. The

workshops were chaired by two members of the study team [redacted for peer review], who are academic SLTs who specialise in SSD. Field notes were taken during the workshops, and they were recorded and transcribed using the speech to text function in MS teams. The second author quality checked the automatic transcription and corrected errors by reviewing the meeting recordings.

Workshop 1 Structure

The results of the content analysis were shown to participants. They were encouraged to give opinions about how any potential subtype definitions would work for their service. Once subtypes were agreed, the content of diagnostic pathways was discussed, with a focus first on screening and then on differential diagnosis. A feasible diagnostic protocol was agreed with a focus on what types of speech samples or assessments were necessary for differential diagnosis.

Workshop 2 Structure

The agreed subtype definitions and then diagnostic protocol were presented to participants for comment and refinement.

Post Workshop Data Checking

A report of the workshops was sent to participants six weeks after the final workshop for checking and confirmation of accuracy.

Results

Content Analysis of Local Paperwork and Guidelines

Data received from the five NHS sites varied from brief summary guideline documents to service-level policies for SSD populations and documents to support the clinical decision-making process for SSD diagnosis and intervention selection (Table 1). The number of documents received from each SLT service varied across the five sites (range 1-11).

Differences were observed in the content of the documentation provided by the NHS services (table 1). Inclusion of a clinical decision-making tool to support SSD diagnosis and/or intervention selection

was the most common element in the data, evidenced by four of the five participating NHS services. Clinical guidance on assessment, diagnosis and intervention for children with SSD, definitions of SSD subtypes and information on appropriate assessment selection were, in each case, provided by only one of the five participating services.

NHS Site	No. files	SSD Guidance	References	Pathway flowchart	Decision making tool	SSD subtype definition	SSD subtype diagnostic criteria	Assessment tool information	Assessment selection
A	4	✓	✓	✓	✓		✓		✓
B	11		✓	✓	✓	✓	✓	✓	
C	1			✓	✓				
D	6				✓				
E	2							✓	

Table 1: Content Analysis

Aim 1: To agree on an SSD classification system which reflects current clinical practice in the UK, supports the clinical decision-making process, and is appropriate for large scale adoption.

The content analysis and workshop 1 revealed that all services were using the SSD classification proposed by Dodd (2014). It was agreed as the classification system that is appropriate for use in the UK and likely to be familiar to most SLTs practicing in the UK.

Aim 2: To agree definitions for each subtype of SSD which are clinically relevant for the UK

Following the consensus to use the Differential Diagnosis System (Dodd, 2014), some amendments to the definitions of the sub-types were proposed by participants and agreed. These amendments were

necessary to fit with how the clinicians operationalise these definitions in clinical practice. For those children who may not fit neatly in to one subtype, the group agreed to prioritise an initial diagnosis that would allow treatment planning with consideration that a child's presentation may change over time. Participants agreed that giving children a mixed SSD diagnosis could be overly complicated as well as confusing for parents. This was agreed by the parent representative.

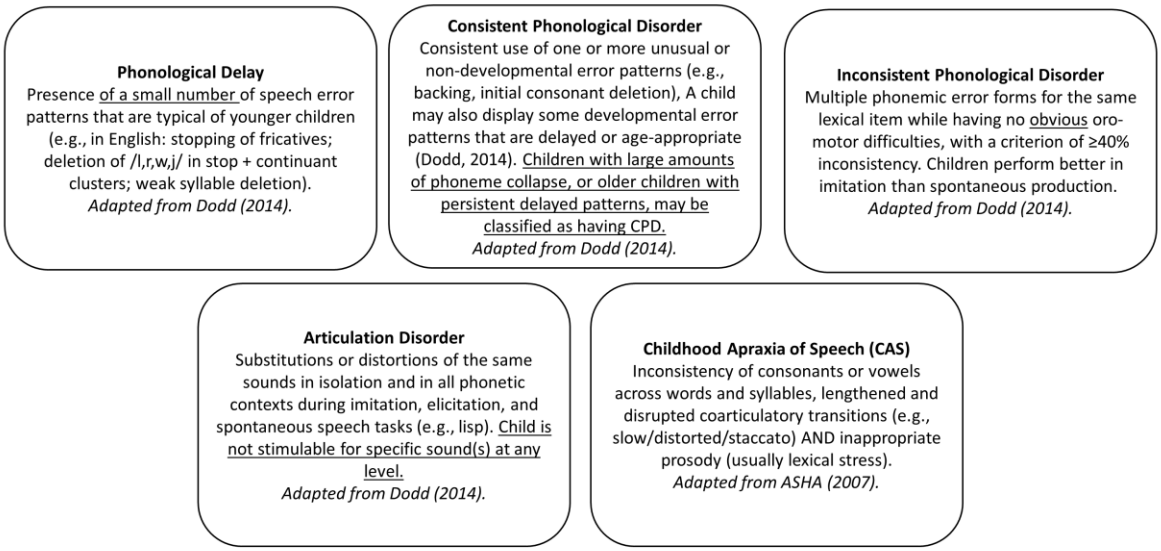

Figure 1: Agreed SSD Subtype definitions with changes to the wording in Dodd (2014) underlined.

Articulation Disorder

There was discussion surrounding the idea that an articulation disorder should involve a motoric or phonetic difficulty. Although the original definition suggests that any difficulty producing specific sounds includes within "imitation and elicitation", the group agreed that clarification should be added

that this means children are not stimulable for these specific sounds at any level (e.g., C, CV/VC, single word).

Phonological Delay

Participants discussed severity and persistence of a phonological delay, and at what point, if any, a delay should be considered a disorder and indeed whether "phonological delay" and "phonological disorder" are distinct subtypes. The group decided to keep delay and disorder as distinct subtypes, with the caveat that a minor amendment was added to specify that the number of delayed speech error patterns observed should be small. The group discussed whether it would be possible or appropriate to specify a number of typical error patterns which constitutes "small". While the group did agree that it was likely to be only two to three, the decision was made not to specify a number because of the need to consider the impact of multiple factors such as the combination of patterns, the child's age, and the overall impact on intelligibility. The group agreed that children with system wide contrast collapse were likely better served by a phonological disorder diagnosis.

Consistent Phonological Disorder (CPD)

The amended definition specifies that children diagnosed with CPD may have a large amount of phoneme collapse or may be older, school-aged children with persistent delayed error patterns.

Inconsistent Phonological Disorder (IPD)

There was discussion around whether children with this diagnosis could present with mild or sub-clinical oro-motor difficulties and it was agreed to add "no obvious" to the definition to add clarity. It was agreed that, for clarity, the 40% inconsistency criterion should be amended to 40% or more using the '≥' symbol.

Childhood Apraxia of Speech (CAS)

Dodd's original definitions use the term "developmental verbal dyspraxia". The study team introduced, in line with the international consensus, the term "childhood apraxia of speech" and

discussion was had around the adoption of this term. Participants agreed with this terminology change. Given the change in terminology and acknowledging the increase in the research literature, and particularly new treatment approaches on this disorder over the last decade, participants agreed to use the American Speech-Hearing Association (ASHA) criteria for the definition of CAS instead of Dodd's definition.

Aim 3: To agree a diagnostic protocol for categorising children with SSD into these subtypes which is feasible for the publicly funded NHS in the UK.

Only two services provided assessment/protocol information for how to differentially diagnose children with SSD (Content Analysis). Participants agreed that the first step in diagnosis is confirming the presence/absence of an SSD and that this screening may happen prior to direct contact with an SLT, i.e., via a telephone helpline or referral from, for example, a teacher. The development of a screening protocol for SSD was discussed and debated by workshop participants. It was agreed that a measure of intelligibility [e.g., using the Intelligibility in Context Scales, ICS (McLeod et al., 2012)] is useful to evaluate the impact and severity of a child's speech difficulties on their communication with the people around them. Participants agreed that, while some therapists may complete the ICS with parents or families as part of assessment, many services use the ICS as part of the referral process and documentation from education settings. In both cases, it was agreed that the ICS provides useful screening information about a child's speech that informs prioritisation of care and assessment. Participants also agreed that the screening protocol should include informal conversation with the child to provide a subjective impression of speech, together with a connected speech screen to provide a more objective impression of speech skills. Participants' clinical perspectives aligned with the growing body of evidence that collecting connected speech data is important to better understand where the level of breakdown occurs for different children with SSD:

"*at the simplest level, also I think assessing connected speech is so important for those children who are referred who are unintelligible. But when you assess it (at) single word level they're fine. And*

actually, it's almost like, 'is there a breakdown at this level? Yes / No. Is that because of XYZ?', rather than actually transcribing the data."

Participants discussed the types of speech sample data that would need to be collected to make a diagnosis for each and any of the subtypes. It was agreed that different levels of data would be required, but that there should be a common 'core' speech sample collected for all children. This core sample includes single words, consonant (C), consonant-vowel (CV) and vowel-consonant (VC) stimulability, connected speech and parent/carer/teacher ratings of the child's intelligibility. The inclusion of consistency/inconsistency was discussed, but it was agreed that this would not be required for all children and would instead be part of a more in-depth, or 'drill down' sample required for children where diagnostic uncertaintainty. The need to allow for the impact that different levels of clinical experience may have on the speed of the diagnostic process was discussed regarding the protocol, and it was acknowledged that assessment and diagnosis may take more than one clinical session.

"it's about the speed at which you think, and for a lot of staff, it'll take them a bit of thinking to realise inconsistency as somewhere they should go and therefore having it in a sort of second session or whatever is OK. But some of us might decide really quickly we're going to do that inconsistency assessment, but it doesn't mean that everybody has the experience to do that. So I think it is fine in the drill down"

Although the group acknowledged the benefits to research of having larger datasets for each child, the need to minimise burden of assessment for children, families and clinicians was agreed by participants.

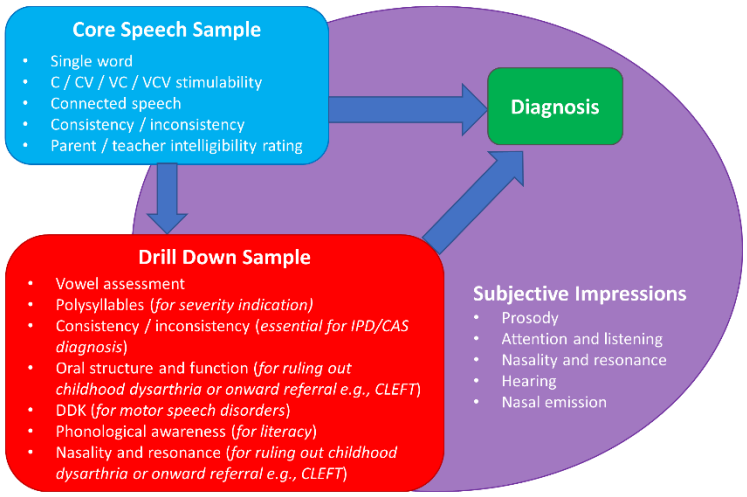

Figure 2: Diagram showing the minimum diagnostic protocol. The "Core Speech Sample" (blue) should be completed for every child with a suspected SSD. The "Drill down sample" (red) is for children with more complex SSD, or where onward referral might be necessary (shown in brackets). The "subjective impressions" (purple) details the other areas the SLT should assess from interaction and conversation with the child.

Stimulus Tools and Assessments

Single word assessments

Two services gave suggested published speech assessments such as the Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al., 2002); the ICS (McLeod et al., 2012); the South Tyneside Assessment of Phonology (Armstrong & Ainley, 1993); the CLEAR Phonology Screen (Kerryjane & Spilsby, 2006) and the Nuffield Dyspraxia Programme assessment (Williams & Stephens, 2004) in their guidelines. Of these the South Tyneside Assessment of Phonology, and the CLEAR are single-word speech assessments. The Nuffield Dyspraxia programme is primarily a single word assessment, but also contains phrases. The DEAP is also mainly a single-word assessment, but incorporates stimulability testing; inconsistency assessment; a limited assessment of connected speech; and a short oro-motor assessment. No service stipulated that a particular assessment must

be used. The participatory workshop discussed the assessments and stimulus tools with regard to clinical experience, evidence base and UK standardisation, using the APEASE (Acceptability, Practicability, Effectiveness, Affordability, Side-effects, and Equity) criteria (Michie et al., 2014) to guide discussion. The issue of cost of assessment materials in particular was discussed at length and it was agreed that, because of funding differences across services, some flexibility would have to be built in in order to make implementation in clinical services feasible.

In terms of the single word assessment tools, participants agreed that the DEAP (Dodd et al., 2002) was a preferred method of assessment. This was because the DEAP is standardised on a UK population and offers subtests designed to differentially diagnose within Dodd's framework (Dodd, 2014). However, it was agreed that this could not be universally specified as a required stimulus tool because of the financial implications of making the assessment available to all services and staff. Nevertheless, participants agreed to include the DEAP (Dodd et al., 2002) as a suggested tool for single word assessment. The STAP (Armstrong & Ainley, 1993) was discounted due to participants feeling it is outdated. Some participants felt that using the CLEAR (Kerryjane & Spilsby, 2006) was a false economy because, although the initial outlay for the tool was lower than the DEAP, the cost of clinical time to analyse and interpret the results negated any significant cost saving over time, compared with the DEAP which offers a framework for identifying phonological processes.

"Let's think about how much time you have to spend analysing and OK face to face with the child, …it might appear quicker, it might be cheaper to buy the assessment. But in terms of how hard you then have to work to do the analysis on the CLEAR, you have to spend a lot more time organising how you're going to pull together that data. And you know when you've got something like the DEAP and, it's phonological processes. It's a very quick run through that. So it appears cheap the CLEAR because it's cheap to buy the assessment."

The single word section of the Nuffield Dyspraxia Programme (Williams & Stephens, 2004) was agreed to be a tool preferred as part of the drill-down criteria if CAS was suspected and it was agreed it is not designed for children with phonological disorders.

Assessing stimulability

Participants acknowledged that stimulability would often be assessed informally by SLTs, with some specific tools being used to assess specific sounds that had not been elicited during single word or connected speech assessment and which required closer examination. Participants agreed that the Stimulability Assessment (Williams et al., 2010) offers a broad range of contexts for examining specific sounds, and as a free resource it may be more appealing and accessible to NHS services, but that other tools, such as the DEAP articulation assessment stimulability section (Dodd et al., 2002), could also be used. The group agreed stimulability of every consonant should not be assessed, rather SLTs should focus on specific sounds absent from the child's inventory following phonological analysis (e.g., using the DEAP).

Connected speech stimulus tools

While it was agreed that connected speech is important, participants discussed what SLTs would do with the connected speech sample after it was collected and how much analysis was required to make a diagnosis. Challenges with the DEAP connected speech assessment (which comprises three composite pictures for children to describe) were discussed, including the limited potential for eliciting natural, connected speech. The group agreed that an informal, tick-box type approach to provide a sense of the characteristics of a child's speech skills listed under 'subjective impressions' (Figure 2, purple) would be appropriate to provide enough initial information. Participants agreed that using an expressive language assessment to collect connected speech would be a useful way to simultaneously assess for any concomitant language difficulties. It was agreed that a structured format for eliciting connected speech was preferable to an informal conversation because, for children with lower intelligibility, it provides a context and targets to help the SLTs identify the child's intended target

words. On this basis, the Renfrew Action Picture Test [RAPT; (Renfrew, 2016)] is a suggested tool for the protocol.

Intelligibility rating stimulus tools

Participants agreed that the ICS (McLeod et al., 2012) is appropriate to recommend as a parent/carer reported measure of impact because it is quick to administer, free, and widely available in a variety of languages.

Inconsistency stimulus tools

The group discussed the use of the DEAP inconsistency assessment (Dodd et al., 2002) and acknowledged the challenge of asking children to repeat 25 items three times. The use of the "core efficacy monitoring assessment" was discussed as a free, quick and simple tool to use to screen for inconsistency (Dodd et al., 2006). Using this as a tool to generate a definitive diagnosis was seen as challenging because there is no evidence to support this. Consideration was given to the suggestion that administering the DEAP diagnostically to screen for consistency and using the cut-off of ≥50% inconsistency to trigger further full assessment using the DEAP inconsistency subtest. This was agreed on the basis that it would be achievable with all children from a time perspective, compared with the full DEAP inconsistency assessment (Figure 3).

Analysing Speech Data

For intervention planning, particularly target selection, SLTs need to be mindful of how they analyse assessment data. Participants discussed the use of the freely available Phonetic and Phonological Systems Analysis (PPSA) (Bates & Watson, 2012), which is in use in some services. While some participants felt that therapists conducted this type of analysis without the specific use of the published tool itself, it was agreed that the tool did support accurate phonological analysis which is necessary for subsequent target selection for interventions. Participants discussed and agreed that intelligibility ratings obtained from the ICS (McLeod et al., 2012) are important for determining severity and measuring change.

Summary of Recommended Tools

Table two details the suggested tools participants agreed could be used to achieve the core sample (Figure 1, blue). These tools additionally provide enough opportunity to complete the suggested "subjective impressions" (Figure 1, purple) SLTs should aim to collect.

Speech Area	Suggested Tools for Core Speech Sample
Initial screen	DEAP screen and the Intelligibility in Context Scale
Single Word Naming	DEAP phonology subtest or toddler version
Stimulability	Stimulability Assessment (Powell & Miccio, 1996) or DEAP stimulability for consonants and vowels absent from the phonetic inventory
Intelligibility	Informal clinician rating based on connected speech from, for example, the Renfrew Action Picture Test and the Intelligibility in Context Scale
(In)consistency	DEAP screen (repeated twice) then the DEAP inconsistency assessment if indicated

Table 2: Suggested stimulus tools and assessments for the Core Speech Sample. DEAP= Diagnostic Evaluation of Articulation and Phonology

Data Checking

A draft report, including the summary figures, was sent to participants for checking after the workshop. One participant reported that children might meet the criteria for phonological delay or consistent phonological disorder but be minimally stimulable and therefore require an articulatory intervention approach. We suggest that these children do indeed receive a diagnosis of a phonological subtype of disorder, although some initial stimulability intervention may be required.

Three participants reported that there was ambiguity surrounding whether the DEAP inconsistency (Dodd et al., 2002) assessment was part of the core (figure 1 blue) or drill down (figure 1 red) assessment. A summary table (table 2) clarifies that the DEAP screen includes an opportunity to sample the 10 words twice and this should form part of the core assessment. Children who are more than 50% inconsistent should receive further inconsistency assessment as part of the "drill down" assessment and in line with the DEAP manual instructions (Dodd et al., 2002).

One participant suggested that a specific vowel assessment might be a useful addition to the "drill down assessment". This was added. A different participant suggested that core stimulability should also include /VCV/ stimulability. This was also added.

Discussion

This study aimed to establish a consensus on classifying SSD subtypes and an agreed diagnostic protocol for arriving at these subtypes for SLTs working in the publicly funded national health service in the UK. We used participatory methods to arrive at agreed subtypes, with workable definitions, and a diagnostic protocol with suggested stimulus tools/assessments that would be feasible for clinicians. Although an international consensus on SSD subtype labels and descriptions is desirable (Waring & Knight, 2013), differences in the way SLT services are designed and funded, as well as the way SLTs are educated in different countries makes this challenging, even within English speaking countries. Previous work has suggested that two main classification systems are in use internationally: the SDCS (Shriberg et al., 2010) and Dodd's classification system (Dodd, 2014). Our content analysis of documents from services confirmed the anecdotal evidence that Dodd's system is preferred in the UK. The choice of the Dodd (2014) classification system is likely due to the author having developed much of her work in the UK, therefore influencing university teaching on SSDs. Further, the availability of an assessment tool, the DEAP (Dodd et al., 2002) which maps directly to these subtypes and is standardised on an English-speaking UK population, makes this system an obvious choice for the UK context. It is also worth noting that although the UK professional body, the RCSLT, do not specify a particular classification system must be taught, the subtypes of SSD they suggest in their curriculum guidelines are broadly in line with the terms used by Dodd (2014).

SSD Terms and Definitions

Although the classification system was unanimously agreed, the group decided to change some of the subtype names and subsequently the definitions. Firstly, CAS was chosen as the preferred term over Developmental Verbal Dyspraxia to reflect a growing international consensus that this term

be used (Broomfield et al., 2022). Although the group opted to keep separate labels for phonological delay and consistent phonological disorder, there was considerable discussion over whether these should be subsumed into one label. Indeed, McLeod and Baker (2017) suggest one category of "phonological impairment" to cover both phonological delay and disorder. This is because the intervention approaches for both are often the same, and there is not necessarily a difference in prognosis: children with only delayed phonological patterns do not necessarily resolve their speech quicker that those with disordered patterns (To et al., 2022). Moreover, the choice of intervention should be based more on the number of errors than the nature of these per se (Storkel, 2022). To our knowledge there is no evidence in the current literature to support or specify a given maximum number of processes as a criterion for diagnosis of phonological delay, although the phrase 'small number' is used to suggest SSD treatment selection (Storkel, 2022). Despite this, the group thought it useful to maintain both sub-types and this therefore remains consistent with Dodd (2014).

In terms of changes to the definitions, some were minor clarifications, for example clarifying that children with inconsistent phonological disorder should present with "*more than* or *equal to*" 40% inconsistency; and that "*sounds in isolation…during imitation*" refers to stimulability in the definition of articulation disorder. However, the group felt that it was important that the definition of phonological disorder be widened to include children with large amounts of phoneme collapse. A bigger change was made to the definition of CAS. Rather than remain consistent with the original definition given by Dodd (2014), they opted to adopt the definition given by the American Speech-Language-Hearing Association (2007). There has been an increase the amount of treatment studies for CAS over the last decade and most of these use the ASHA definition, making this a useful change.

Diagnostic Protocol

Researchers suggest that a thorough diagnostic protocol for SSD of unknown origin should include single-word testing, additional single-word testing, a connected speech sample, stimulability testing, and an assessment of inconsistency (Macrae, 2016). Our participants suggested including all of these

aspects in their core diagnostic protocol, with the exception of additional single word testing and the caveat that inconsistency should only be screened for all children and probed in depth for children who show evidence of a potential inconsistent phonological disorder. This contrasts with previous survey studies which suggest that some of these aspects are often missing in SLTs' batteries of assessment (Diepeveen et al., 2020; McLeod & Baker, 2014; Skahan et al., 2007). Our SLTs did agree that time and costs, especially costs of assessment, is often a factor in choosing assessments (Diepeveen et al., 2020) and for this reason, there was a reluctance to suggest specific commercial assessments for collecting data. Indeed, Fabiano-Smith (2019) suggests that commercial standardised assessments may be less accurate for differential diagnosis of SSD than criterion-referenced measures, and it is certainly the case that a thorough analysis of a spontaneous speech sample can be diagnostically powerful (Bates & Titterington, 2021). However, using an assessment such as the DEAP is likely to be much less time consuming, and therefore potentially cost saving in terms of staff time, than analysing a connected speech sample. The DEAP (Dodd et al., 2002) was therefore suggested as a commercial assessment that could be recommended as a tool for gathering information about single-word speech production, stimulability, and inconsistency. This is consistent with the choice of Dodd's classification system as this test is specifically designed for differential diagnosis for the subtypes chosen by participants.

For children who do not speak English, the SLT will need to use equivalent tests developed for the target language, if available, or a spontaneous speech sample with assistance from an interpreter. There are still limitations even if these approaches are put in place. Many language-specific assessments have been developed for children growing up in monolingual environments (McLeod & Verdon, 2014) and interpreters would need specialist training to be able to identify specific errors in the speech of children with SSD. Evidence from cross-linguistic studies is helpful, however, and tells us that most children have acquired the sound system of the languages they are exposed to by age 5, with clear patterns regarding which consonants are acquired early and late in development according to manner and place characteristics (McLeod & Crowe, 2018).

The participants also highlighted the importance of collecting a parent/carer measure of a child's perceived speech intelligibility. Participants suggested that the Intelligibility in Context Scale (McLeod et al., 2012) was both a useful method for screening for SSD (McLeod, 2020) and for obtaining parents'/carers' views on their child's intelligibility. Patient (or in this case parent) Reported Outcome Measures (PROMS) are relatively under-used in speech and language therapy (Cohen & Hula, 2020) and the inclusion of this measure highlights the importance of gathering the views of parents when assessing their child's speech. This assessment as also favoured because it is available in a large number of languages, is free to use, and quick to complete.

Limitations

This study is necessarily limited by its focus on the UK context. However, we argue that a similar process could be undertaken in other settings using as a starting point the subtype labels agreed here. Choice of assessments will be constrained by what is available in any particular country in the correct language/s and with relevant norms. A further limitation is our inclusion of only five services in the UK, represented by only six specialist SLTs. Although these services represent a variety of urban and rural locations, they covered only two (England and Scotland) of the four nations and the more experienced SLTs may have been biased to designing a diagnostic protocol that was feasible for more experienced SLTs. Those with fewer years of experience may find that the protocol is more time consuming or more difficult to implement. A future study seeks to trial the implementation of these subtype labels and the diagnostic protocol with a large number of children with suspected SSD.

Lastly, we must acknowledge the focus here on SSD of unknown origin. Children with SSDs of known origin, for example SSD associated with cleft palate +/- lip or Down syndrome, will need a different approach. Likewise, children with suspected severe motor speech disorders, for example concomitant CAS and childhood dysarthria, will need a more in-depth assessment of their speech strengths and weakness to allow treatment planning. Moreover, children with co-occurring

neurodiversities such as autism or developmental language disorder will need careful consideration during the diagnostic process.

Conclusions

This study sought to determine a common classification system and diagnostic protocol for childhood SSD of unknown origin using participatory methods. SLTs practicing in the UK agreed that Dodd's classification system (Dodd, 2014) was fit for purpose in the UK with some minor amendments to the descriptions of the subtypes and replacing "Developmental Verbal Dyspraxia" with "Childhood Apraxia of Speech" and adopting the ASHA definition of this SSD subtype. In terms of a diagnostic protocol, the participants agreed that a feasible protocol should include a minimum assessment required for children with more straightforward or mild SSD, with optional additional assessment for more complex cases. In conclusion, consensus was reached on a classification system and diagnostic protocol.

Acknowledgments

The authors wish to thank the Speech and Language Therapists and parent representative who participated in the workshops: [redacted for peer review]. This research was funded by the [redacted for peer review]. For the purpose of Open Access, the author has applied a CC licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Declaration of interest statement

The authors report no competing interests.

References

American Speech-Language-Hearing Association. (2007). Childhood apraxia of speech.

Armstrong, S., & Ainley, M. (1993). *South Tyneside assessment of phonology*. STASS. https://books.google.co.uk/books?id=QzKpAAAACAAJ

Bates, S., & Titterington, J. (2021). Good practice guidelines for the analysis of child speech.

Bates, S., & Watson, J. (2012). Phonetic and Phonological Systems Analysis (PPSA). In: Retrieved from.

Broomfield, J., Cleland, J., & Williams, P. (2022). What's in a name? Dr Jan Broomfield, Dr Joanne Cleland and Dr Pam Williams make their case for adopting the term 'childhood apraxia of speech'. *Bulletin: The Official Magazine Of The Royal College Of Speech & Language Therapists*(833), 1-2. https://strathprints.strath.ac.uk/83091/

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: caseload characteristics. *International Journal of Language & Communication Disorders*, *39*(3), 303-324. https://doi.org/https://doi.org/10.1080/13682820310001625589

Child Speech Disorder Research Network. *Good practice guidelines for the transcription of children's speech in clinical practice and research.* Retrieved 27/02/2023 from https://www.nbt.nhs.uk/sites/default/files/BSLTRU_Good%20practice%20guidelines_Transcription_2Ed_2017.pdf

Cohen, M. L., & Hula, W. D. (2020). Patient-Reported Outcomes and Evidence-Based Practice in Speech-Language Pathology. *American Journal of Speech-Language Pathology*, *29*(1), 357-370. https://doi.org/doi:10.1044/2019_AJSLP-19-00076

Diepeveen, S., Haaften, L. v., Terband, H., Swart, B. d., & Maassen, B. (2020). Clinical Reasoning for Speech Sound Disorders: Diagnosis and Intervention in Speech-Language Pathologists' Daily Practice. *American Journal of Speech-Language Pathology*, *29*(3), 1529-1549. https://doi.org/doi:10.1044/2020_AJSLP-19-00040

Dodd, B. (2014). Differential Diagnosis of Pediatric Speech Sound Disorder. *Current Developmental Disorders Reports*, *1*(3), 189-196. https://doi.org/10.1007/s40474-014-0017-3

Dodd, B., Holm, A., Crosbie, S., & McIntosh, B. (2006). A core vocabulary approach for management of inconsistentspeech disorder. *Advances in Speech Language Pathology*, *8*(3), 220-230. https://doi.org/10.1080/14417040600738177

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

Douglas, N., Hinckley, J., Grandbois, K., Schliep, M., Wonkka, A., Oshita, J., & Feuerstein, J. (2023). How a Power Differential Between Clinicians and Researchers Contributes to the Research-to-Practice Gap. *American Journal of Speech-Language Pathology*. https://doi.org/doi:10.1044/2022_AJSLP-22-00207

Fabiano-Smith, L. (2019). Standardized Tests and the Diagnosis of Speech Sound Disorders. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 58-66. https://doi.org/doi:10.1044/2018_PERS-SIG1-2018-0018

Kerryjane, M., & Spilsby, K. (2006). CLEAR: Phonology Screening Assessment. In: Clear Resources.

Macrae, T. (2016). Comprehensive Assessment of Speech Sound Production in Preschool Children. *Perspectives of the ASHA Special Interest Groups*, *1*(1), 39-56. https://doi.org/doi:10.1044/persp1.SIG1.39

McLeod, S. (2020). Intelligibility in Context Scale: cross-linguistic use, validity, and reliability. *Speech, Language and Hearing*, *23*(1), 9-16. https://doi.org/10.1080/2050571X.2020.1718837

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics & Phonetics*, *28*(7-8), 508-531. https://doi.org/10.3109/02699206.2014.926994

McLeod, S., & Baker, E. (2017). *Children's speech: An evidence-based approach to assessment and intervention*.

McLeod, S., & Crowe, K. (2018). Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review. *American Journal of Speech-Language Pathology*, *27*(4), 1546-1571. https://doi.org/doi:10.1044/2018_AJSLP-17-0100

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The Intelligibility in Context Scale: Validity and Reliability of a Subjective Rating Measure. *Journal of Speech, Language, and Hearing Research*, *55*(2), 648-656. https://doi.org/doi:10.1044/1092-4388(2011/10-0130)

McLeod, S., & McCormack, J. (2007). Application of the ICF and ICF-Children and Youth in Children with Speech Impairment. *Semin Speech Lang*, *28*(04), 254-264.

McLeod, S., & Verdon, S. (2014). A Review of 30 Speech Assessments in 19 Languages Other Than English. *American Journal of Speech-Language Pathology*, *23*(4), 708-723. https://doi.org/doi:10.1044/2014_AJSLP-13-0066

Michie, S., Atkins, L., & West, R. (2014). The behaviour change wheel. *A guide to designing interventions. 1st ed. Great Britain: Silverback Publishing*, 1003-1010.

Powell, T. W., & Miccio, A. W. (1996). Stimulability: A useful clinical tool. *Journal of Communication Disorders*, *29*(4), 237-253. https://doi.org/https://doi.org/10.1016/0021-9924(96)00012-3

RCSLT. (2021). *Curriculum Guidance* Retrieved 15/11/22 from https://www.rcslt.org/wp-content/uploads/2020/08/RCSLT-Curriculum-Guidance-March2021.pdf

Renfrew, C. E. (2016). *The Renfrew Language Scales: Action Picture Test*. Speechmark.

Roper, A., & Skeat, J. (2022). Innovation through participatory design: Collaborative qualitative methods in the development of speech-language pathology technology. *International Journal of Speech-Language Pathology*, *24*(5), 527-532. https://doi.org/10.1080/17549507.2022.2050943

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilkens, C. M., & Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics*, *24*(10), 795-824. https://doi.org/10.3109/02699206.2010.503006

Shriberg, L. D., & Wren, Y. E. (2019). A frequent acoustic sign of speech motor delay (SMD). *Clinical Linguistics & Phonetics*, *33*(8), 757-771. https://doi.org/10.1080/02699206.2019.1595734

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-Language Pathologists' Assessment Practices for Children With Suspected Speech Sound Disorders: Results of a National Survey. *American Journal of Speech-Language Pathology*, *16*(3), 246-259. https://doi.org/doi:10.1044/1058-0360(2007/029)

Stackhouse, J., & Wells, B. (1993). Psycholinguistic assessment of developmental speech disorders. *European Journal of Disorders of Communication*, *28*(4), 331-348. https://doi.org/10.3109/13682829309041469

Storkel, H. L. (2022). Minimal, Maximal, or Multiple: Which Contrastive Intervention Approach to Use With Children With Speech Sound Disorders? *Language, Speech, and Hearing Services in Schools*, *53*(3), 632-645. https://doi.org/doi:10.1044/2021_LSHSS-21-00105

Terband, H., Maassen, B., & Maas, E. (2019). A Psycholinguistic Framework for Diagnosis and Treatment Planning of Developmental Speech Disorders. *Folia Phoniatrica et Logopaedica*, *71*(5-6), 216-227. https://doi.org/10.1159/000499426

To, C. K. S., McLeod, S., Sam, K. L., & Law, T. (2022). Predicting Which Children Will Normalize Without Intervention for Speech Sound Disorders. *Journal of Speech, Language, and Hearing Research*, *65*(5), 1724-1741. https://doi.org/doi:10.1044/2022_JSLHR-21-00444

Waring, R., & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language & Communication Disorders*, *48*(1), 25-40. https://doi.org/https://doi.org/10.1111/j.1460-6984.2012.00195.x

Williams, A. L., McLeod, S., & McCauley, R. J. (2010). *Interventions for speech sound disorders in children*. Paul H. Brookes Publishing.

Williams, P., & Stephens, H. (2004). *Nuffield centre dyspraxia programme*. Miracle Factory.

Wren, Y., Harding, S., Goldbart, J., & Roulstone, S. (2018). A systematic review and classification of interventions for speech-sound disorder in preschool children. *International Journal of Language & Communication Disorders*, *53*(3), 446-467. https://doi.org/https://doi.org/10.1111/1460-6984.12371